2.31 $p(x) = N(x \mid \mu_x, \Sigma_x)$ $\quad E[y] = A\mu + b$

$p(z) = N(z \mid \mu_z, \Sigma_z)$ $\quad \mathrm{cov}[y] = L^{-1} + A\Lambda^{-1}\Lambda^T$

$y = x + z$

$$p(y) = p(x) \cdot p(y \mid x) = N(y \mid A\mu + b, L^{-1} + A\Lambda^{-1}A^T)$$

$$= N(x \mid \mu, \Lambda^{-1}) \cdot N(y \mid \{\Lambda + A^T L A\}^{-1}\{A^T L(y-b) + \Lambda\mu\}, \{\Lambda + A^T L A\}^{-1})$$

2.32 $p(x,y) =$

$p(x) = N(x \mid \mu, \Lambda^{-1})$

$p(y \mid x) = N(y \mid Ax + b, L^{-1})$

$= \dfrac{p(y,x)}{p(x)}$

① Examine quadratic exponent

$$p(x,y) = p(y,x) = p(x) \cdot p(y \mid x)$$

$$= N(x \mid \mu, \Lambda^{-1}) \cdot N(y \mid Ax + b, L^{-1})$$

$$= e^{-\frac{1}{2}(x-\mu)^T \Lambda^{-1}(x-\mu)} \cdot e^{-\frac{1}{2}(y - Ax - b)^T L^{-1}(y - Ax - b)}$$

② Complete the square

$$= -\frac{1}{2}(x-\mu)^T \Lambda^{-1}(x-\mu) - \frac{1}{2}(y - Ax - b)^T L^{-1}(y - Ax - b)$$

$$= -\frac{1}{2}\left[y^T L^{-1} y - y^T L^{-1} Ax - y^T L^{-1} b - (Ax)^T L^{-1} y + (Ax)^T L^{-1} Ax + (Ax)^T L^{-1} b - b^T L^{-1} y - b^T L^{-1} Ax + b^T L^{-1} b\right]$$

$$+ \frac{1}{2}(x-\mu)^T \Lambda^{-1}(x-\mu)$$

$$= -\frac{1}{2}\left[y^T L^{-1} y - y^T L^{-1} Ax - y^T L^{-1} b - (Ax)^T L^{-1} y - b^T L^{-1} y\right]$$

$$- \frac{1}{2}\left[-(Ax)^T L^{-1} Ax + (Ax)^T L^{-1} b - b^T L^{-1} Ax + b^T L^{-1} b\right]$$

$$- \frac{1}{2}(x-\mu)^T \Lambda^{-1}(x-\mu)$$

$$= -\frac{1}{2}\left[(x-\mu)^T \Lambda^{-1}(x-\mu) - y^T L^{-1} Ax - (Ax)^T L^{-1} Ax + (Ax)^T L^{-1} b - b^T L^{-1} Ax\right]$$

$$- \frac{1}{2}\left[y^T L^{-1} y - y^T L^{-1} b - b^T L^{-1} y - y^T L^{-1} Ax + b^T L^{-1} b\right]$$

$$= -\frac{1}{2}\left[x^T \Lambda^{-1} x - \mu^T \Lambda^{-1} x - x^T \Lambda^{-1} \mu + \mu^T \Lambda^{-1} \mu - y^T L^{-1} Ax - (Ax)^T L^{-1}(Ax) + (Ax)^T L^{-1} b - b^T L^{-1} Ax\right]$$

$$- \frac{1}{2}\left[(y - b)^T L^{-1}(y - b)\right]$$

2.34 $\ln p(x|\mu,\Sigma) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x-\mu)$

$$\frac{\partial}{\partial x}(A^{-1}) = -A^{-1}\frac{\partial}{\partial x}A^{-1}$$

$$\frac{\partial}{\partial A}Tr(A) = I$$

$$\frac{\partial}{\partial A}\ln|A| = (A^{-1})^T$$

$$\frac{\partial}{\partial\Sigma}\ln p(x|\mu,\Sigma) = p(x|\mu,\Sigma)^{(-1)^T} = \frac{\partial}{\partial\Sigma}\left[-\cancel{\frac{ND}{2}\ln(2\pi)} - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

$$= -\frac{N}{2}(\Sigma^{-1})^T + \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x-\mu)\frac{\partial}{\partial x}\Sigma^{-1}$$

$$\frac{N}{\cancel{2}}\cancel{\Sigma^{-1}} = \frac{1}{\cancel{2}}\sum_{n=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x-\mu)\cancel{\frac{\partial}{\partial x}\Sigma^{-1}}$$

$$\Sigma = \frac{1}{N}\sum(x_n-\mu)^T(x_n-\mu)$$

2.35 $E[x]=\mu$ Prove $E[xx^T] = \mu\mu^T + \Sigma$

$$E[xx^T] = \int N(x|\mu,\Sigma)\,xx^T dx =$$

$$\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}\int e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}\cdot x\cdot x^T dx\ ;\ z = x-\mu,\ x = z+\mu$$

$$= \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}\int e^{-\frac{1}{2}z^T\Sigma^{-1}z}\cdot(z+\mu)^T(z+\mu)\,dz$$

$$= \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}\int e^{-\frac{1}{2}z^T\Sigma^{-1}z}\ z^Tz\,dz\ ;\ z = \sum_{j=1}^{D}u_j^T z u_j$$

$$= \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}\int e^{-\frac{1}{2}\sum_{k=1}^{D}\frac{(u_j^T z u_j)^2}{\lambda_k}}\ dy + 1$$

$$= \frac{1}{\cancel{(2\pi)^{D/2}|\Sigma|^{1/2}}}\sum_{i=1}^{D}\sum_{j=1}^{D}u_i u_j^T\int e^{-\frac{1}{2}\sum_{k=1}^{D}\frac{(u_j^T z u_j)^2}{\lambda_k}}\ y_i y_j\,dy + 1$$

$$= \sum_{i=1}^{D}\sum_{j=1}^{D}u_i u_j^T\lambda_i + [\ ] = \sum[x^T x]\Sigma + \mu^T\mu = E[x^Tx]$$

2.36 $\sigma_{ML}^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2 -$

$\mu_{ML}^{(N)} = \mu_{ML}^{(N-1)} + \frac{1}{N}(\vec{x}_N - \vec{\mu}_{ML}^{(N-1)})$

$$= \frac{1}{N}(x_N-\mu)^2 + \frac{1}{N}\sum_{i=1}^{N-1}(x_n-\mu)^2$$

$$= \frac{1}{N}(x_N-\mu)^2 + \frac{N-1}{N}(x_N-\mu_{ML}^{(N-1)})^2$$

$$= \frac{1}{N}\left[(x_n-\mu)^2 + (x_n-\mu_{ML})^2 - \frac{(x_n-\mu_{ML}^{(N-1)})^2}{N}\right]$$

$$= \frac{1}{N}\left((x_n-\mu)^2 - \sigma_{N-1}^2\right) - \sigma_{(N-1)}^2$$

2.37 $\cancel{E[xx] - E[x]^2 - E[x^2]}$

$$\mathrm{cov}[x,y] = E[(x-\mu_1)(y-\mu_2)]$$

$$= E[xy] - E[x]E[y]$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i$$

$$\boxed{= \frac{1}{N}x_N y_N - \left[\frac{1}{N}x_N + \frac{N-1}{N}\mu_x^{(N-1)}\right]\left[\frac{1}{N}y_N + \frac{N-1}{N}\mu_y^{(N-1)}\right]}$$

$$= \frac{1}{N}x_n y_n - \mu_x\mu_y - 2\frac{N-1}{N}\mu_x^{(N-1)}\mu_y^{(N-1)} + \frac{N-1}{N}\mu_x^{(N-1)}\mu_y^{(N-1)}$$

$$=$$

Tridag.h

$$\begin{bmatrix} b_0 & c_0 & 0 & \cdots & & \\ a_1 & b_1 & c_1 & \cdots & & \\ & & \ddots & & & \\ & & \cdots & a_{N-2} & b_{N-2} & c_{N-2} \\ & & \cdots & 0 & a_{N-1} & b_{N-1} \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ \cdots \\ u_{N-2} \\ u_{N-1} \end{bmatrix} = \begin{bmatrix} r_0 \\ r_1 \\ \cdots \\ r_{N-2} \\ r_{N-1} \end{bmatrix}$$

```
tridiag(a, b, c, r, u)
    j, n = a.size()
    Doub = bet;
    Vec = gam(n)
    if (b[0] == 0.0) throw("Error");
    u[0]=r[0]/bet=b[0];
    for(j=1;j++){
      gam[j] = c[j-1]/bet;
      bet  = b[j] - a[j]*gam[j]
      if (bet==0) throw ("Error");
      u[j]=(r[j]-a[j]*u[j-1])/bet;
    }
    for (j=(n-2);j--)
      u[j] -= gam[j+1]*u[j+1]
```

c[0]/bet

1	2	3

bet = b[1] - a[1]·(c[1]/b[0])

u[1]=(r[1]-a[1]*u[0])/bet

c[0]/bet c[1] / b[1]-a[1]·c[1]/bet

1	2	3

bet = b[2] - a[2]·(c[2] / b[1]-a[1]·c[1]/b[0])

u[2]=(r[2]-a[2]*u[1])/bet

~~2.35 $E[x]=\mu$~~

~~Prove $E[xx^T]=\mu\mu^T+\Sigma$~~

2.39. $\mu_N = \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{ML}$ $\quad \frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}$ $\quad \sigma_N^2 = \frac{\sigma_0^2\sigma^2}{\sigma^2+N\sigma_0^2}$

$$P(\mu|X) = N(\mu|\mu_N;\sigma_N^2)$$

$$= \frac{1}{(2\pi\sigma_N^2)^{N/2}} e^{-\frac{\sigma^2+N\sigma_0^2}{2\sigma_0^2\sigma^2}\left(\mu-\left[\frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0+\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{ML}\right]\right)^2}$$

$$= N(\mu|\mu_0;\sigma_0^2)\,N(\mu|\mu_1,\sigma_1^2)\cdots N(\mu|\mu_{N-1},\sigma_{N-1}^2)\,N(\mu|\mu_N,\sigma_N^2)$$

2.40

2.38 Completing the square of μ_N $\frac{1}{\sigma_N^2}$

2.37 ML of covariance multivariate Gaussian distribution

$$\mu_{ML}^{(N)} = \frac{1}{N}\sum_{n=1}^{N} x_n = \frac{1}{N}x_n + \frac{1}{N}\sum_{n=1}^{N-1} x_n = \frac{1}{N}x_n + \frac{N-1}{N}\mu_{ML}^{(N-1)}$$

$$= \mu_{ML}^{(N-1)} + \frac{1}{N}\left(x_n - \mu_{ML}^{(N-1)}\right)$$

from $\Sigma_{ML} = \frac{1}{N}\sum(x_n-\mu_{ML})(x_n-\mu_{ML})^T$

1.38. $\mu_N = \frac{\sigma^2}{N\sigma_0^2 + \sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2}\mu_{ML}$

$$\frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}$$

$$p(\mu|X) = N(\mu|\mu_N, \sigma_N^2)$$

$$p(\mu|X) = p(X|\mu)p(\mu) = N(x_n|\mu, \sigma^2) \cdot N(\mu|\mu_0, \sigma_0^2)$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_n-\mu)^2} \frac{1}{(2\pi\sigma_0^2)^{N/2}} e^{\frac{1}{2\sigma_0^2}\sum_{i=1}^{N}(\mu-\mu_0)^2}$$

$$= \frac{1}{2\pi\sigma^N\sigma_0^N} e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{N}(x_n-\mu)^2 + \frac{1}{2\sigma_0^2}\sum_{i=1}^{N}(\mu-\mu_0)^2}$$

$$= \frac{1}{2\pi\sigma^N\sigma_0^N} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_n - 2x_n\mu + \mu)^2 - \frac{1}{2\sigma_0^2}\sum_{i=1}^{N}(}$$

$$= \frac{1}{2\pi\sigma^N\sigma_0^N} e$$

$$\mathrm{Cov}[xy] = E[(x-\mu_x)(y-\mu_y)]$$

$$= \frac{1}{N}\sum_{i=1}^{N}[(x_i-\mu_x)(y_i-\mu_y)]$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \frac{1}{N}\sum_{i=1}^{N} x_i \cdot \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$= \frac{1}{N} x_N y_N + \frac{1}{N}\sum_{i=1}^{N-1} x_i y_i - \frac{1}{N}\left[x_N + \sum_{i=1}^{N-1} x_i\right] \cdot \frac{1}{N}\left[y_N + \sum_{i=1}^{N-1} y_i\right]$$

$$= \mu_x^N \mu_y^N + \mu_x^{N-1}\mu_y^{N-1} - \left[\mu_x^N + \frac{1}{N}\sum_{i=1}^{N-1} x_i\right]$$

$$= E[xy] - E[x]E[y]$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \frac{1}{N}\sum_{i=1}^{N} x_i \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$= \cancel{\mu_N \mu_N + \mu_x^{N-1}\mu_y^{N-1} - \left[\mu_{x_1} + \frac{1}{N}\sum_{i=1}^{N-1} x_i\right]\left[\mu_y^N + \frac{1}{N}\sum_{i=1}^{N-1} y_i\right]}$$

$$\boxed{\mathrm{Cov}[xy]} = E[(x-\mu_x)(y-\mu_y)]$$

$$= E[xy] - E[x]E[y]$$

$$= \sum_{i=1}^{N} x_i y_i - n\mu_x\mu_y$$

$$= \mu_x\mu_y + \sum_{i=1}^{N-1} x_i y_i - n\mu_x\mu_y$$

$$= \frac{1}{2}\left[x^T \Lambda^{-1} x - \mu^T \Lambda^{-1} x - x^T \Lambda^{-1} \mu + \mu^T \Lambda^{-1} \mu - \cancel{y^T L^{-1} A x} - \cancel{(Ax)^T L^{-1} (Ax)} + (Ax)^T L^{-1} b - \cancel{y^T L^{-1} A x} - \cancel{b^T L^{-1} A x}\right] - \frac{1}{2}\left[(y-b)^T L^{-1} (y-b)\right]$$

$$= \frac{1}{2}\left[x^T (\Lambda^{-1} - L^{-1} A) x + (\mu^T \Lambda^{-1} - L^{-1} A (y^T + b^T)) x\right.$$

$$= -\frac{1}{2}(x-m)^T(\Lambda + A^T L A)(x-m) + \frac{1}{2} m^T (\Lambda + A^T L A) m + \text{const}$$

$$m = (\Lambda + A^T L A)^{-1}[\Lambda\mu + A^T L (y-b)]$$

$$\int_{-\infty}^{\infty} e$$

$$= 2 \cdot \frac{1}{2} \sqrt{\frac{\pi}{\Lambda + A^T L A}} \cdot e^{\frac{1}{2} m^T (\Lambda + A^T L A) m} + \int_{-\infty}^{\infty} e^{\cdots}$$

$$= \frac{1}{2} y^T \{L - LA(\Lambda + A^T L A)^{-1} A^T L\} y + y^T \{(L - LA(\Lambda + A^T L A)^{-1} A^T L\} b + LA(\Lambda + A^T L A)^{-1} \Lambda \mu$$

Lkr Fxııs

2.33

2.34 $\ln p(X|\mu,\Sigma) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T \Sigma^{-1}(x_n-\mu)$

$$\frac{\partial}{\partial x}(A^{-1}) = -A^{-1}\frac{\partial A}{\partial x}A^{-1}$$

$$\frac{\partial}{\partial A}\operatorname{Tr}(A) = I$$

$$\frac{\partial}{\partial A}\ln|A| = (A^{-1})^T$$

$$\frac{\partial}{\partial x}\ln p(x|\mu,\Sigma) = \frac{\partial}{\partial x}\left[-\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)\right]$$

$$\max: 0 = [p(x|\mu,\Sigma)]^{-1\,T} = \sum_{n=1}^{N}(x-\mu)^T \Sigma^{-1}(x_n-\mu)$$

$$\Sigma_{ML} = \sum_{n=1}^{N}(x-\mu)^T(x_n-\mu)$$

2.40 $N(X|\mu,\Sigma)$, $X=\{x_1,\cdots,x_n\}$

$p(\mu)=N(\mu|\mu_0,\Sigma)$

Posterior $\propto$ liklihood $\cdot$ prior

$$p(\mu|X)\propto p(\mu)\prod_{N=1}^{N} p(x|\mu,\Sigma)$$

$$\propto N(\mu|\mu_0,\Sigma)\prod_{n=1}^{N} p(x|\mu,\Sigma) \propto \frac{1}{(2\pi\sigma_0)^{1/2}} e^{\frac{-(\mu-\mu_0)^T\Sigma_0^{-1}(\mu-\mu_0)}{2}} \cdot \frac{1}{(2\pi\sigma)^{1/2}} e^{-\frac{1}{2}\sum_{i=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)}$$

$$\propto \frac{1}{(2\pi)(\sigma_0\sigma)^{1/2}} e^{-\frac{1}{2}(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)-\frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)}$$

$$= -\frac{1}{2}(\mu-\mu_0)^T\Sigma_0^{-1}(\mu-\mu_0)-\frac{1}{2}\sum_{i=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)$$

$$= -\frac{1}{2}\left[\mu^T\Sigma_0^{-1}\mu-\mu^T\Sigma_0^{-1}\mu_0-\mu_0^T\Sigma_0^{-1}\mu+\mu_0^T\Sigma^{-1}\mu_0\right]$$

$$-\frac{1}{2}\sum_{i=1}^{N}\left[x_n^T\Sigma^{-1}x_n-x_n\Sigma^{-1}\mu-\mu^T\Sigma^{-1}x_n+\mu^T\Sigma^{-1}\mu\right]$$

$$= -\frac{1}{2}\mu^T\left[\Sigma_0^{-1}+N\Sigma^{-1}\right]\mu+\mu^T\left[\Sigma_0^{-1}\mu_0+\Sigma^{-1}\sum_{i=1}^{n}x_n\right]+\text{const}$$

$$= -\frac{1}{2}\left[\Sigma_0^{-1}+N\Sigma^{-1}\right]\left[\mu^T\mu+2\mu^T\frac{\left[\Sigma_0^{-1}\mu_0+\Sigma^{-1}\sum_{i=1}^{n}x_n\right]}{\left[\Sigma_0^{-1}+N\Sigma^{-1}\right]}\right]+\text{const}$$

$$= -\frac{1}{2}\left[\Sigma_0^{-1}+N\Sigma^{-1}\right]\cdot\left(\mu+\mu^T\frac{\left[\Sigma_0^{-1}\mu_0+\Sigma^{-1}\sum_{i=1}^{n}x_n\right]}{2\left[\Sigma_0^{-1}+N\Sigma^{-1}\right]}\right)^2+\frac{\left(\mu^T\left[\Sigma_0^{-1}\mu_0+\Sigma^{-1}\sum_{i=1}^{n}x_n\right]\right)^2}{2\left[\Sigma_0^{-1}+N\Sigma^{-1}\right]}+\text{const}$$

of form $\frac{-1}{2\sigma^2}(x-\mu)^2$

$$\mu_N=\frac{\Sigma_0^{-1}\mu_0+\Sigma^{-1}\sum_{i=1}^{n}x_n}{\left[\Sigma_0^{-1}+N\Sigma^{-1}\right]}$$

$$\Sigma_N^{-1}=\left[\Sigma_0^{-1}+N\Sigma^{-1}\right]$$

2.41: Gamma Function $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$

Prove Normalization of: $Gam(\lambda|a,b) = \frac{1}{\Gamma(a)} b^a \lambda^{a-1} \exp(-b\lambda)$

Where: $\int_0^\infty Gam(\lambda|a,b) dx = 1$

Therefore: $\int_0^\infty \frac{1}{\Gamma(a)} b^a \lambda^{a-1} e^{-b\lambda} d\lambda = \frac{b^a}{\Gamma(a)} \int_0^\infty \lambda^{a-1} e^{-b\lambda} d\lambda$

Substitution: $\lambda \cdot b = x$; $b = \frac{x}{\lambda}$; $d\lambda b = dx$

Therefore: $\frac{b^a}{\Gamma(a)} \int_0^\infty \left(\frac{x}{b}\right)^{a-1} e^{-x} \frac{dx}{b} = \frac{\cancel{b^a}}{\Gamma(a)} \int_0^\infty \frac{1}{\cancel{b^{a-1}}} x^{a-1} e^{-x} \frac{dx}{\cancel{b}}$

$= \frac{\cancel{\Gamma(a)}}{\cancel{\Gamma(a)}}$

2.42. Mean $E[\lambda] = \int_0^\infty x \cdot G(\lambda|a,b) = \frac{b^a}{\Gamma(a)} \int_0^\infty \lambda \cdot \lambda^{a-1} e^{-b\lambda} d\lambda$ $\quad b\lambda = x$; $\lambda = \frac{x}{b}$; $d\lambda b = dx$

$= \frac{b^a}{\Gamma(a)} \int_0^\infty \left(\frac{x}{b}\right)^{(a+1)-1} \cdot e^{-x} \cdot \frac{d\lambda}{b} = \frac{b^a}{\Gamma(a)} \int_0^\infty \left(\frac{1}{b^{a+1-1}}\right) x^{(a+1)-1} e^{-x} \cdot \frac{d\lambda}{b}$

$= \frac{1}{b\Gamma(a)} \Gamma(a+1) = \frac{a}{b} \frac{\cancel{\Gamma(a)}}{\cancel{\Gamma(a)}}$

Variance: $Var[\lambda] = E[\lambda^2] - E[\lambda]^2$; $E[\lambda^2] = \frac{b^a}{\Gamma(a)} \int_0^\infty \lambda^2 (\lambda)^{a-1} e^{-b\lambda} d\lambda = \frac{b^a}{\Gamma(a)} \int_0^\infty (\lambda)^{a+2-1} e^{-b\lambda} d\lambda$

$= \frac{b^a}{\Gamma(a)} \int \left(\frac{x}{b}\right)^{a+2-1} e^{-x} \frac{dx}{b} = \frac{\cancel{b^a}}{\Gamma(a) \cancel{b^a} \cdot b^2} \Gamma(a+2)$

$= \frac{(a+1)\Gamma(a+1)}{\Gamma(a) b^2} = \frac{(a+1)}{b^2}$

$= \frac{(a+1)}{b^2} - \frac{a}{b^2} = \frac{1}{b^2}$

2.42 cont. Gamma Distribution Mode : $\frac{d}{d\lambda} \mathrm{Gam}(\lambda|a,b)$

$$= 0 = \frac{b^a}{\Gamma(a)} \left[(a-1)\lambda^{a-2} \exp(-b\lambda) + -b\lambda^{a-1} \exp(-\lambda)\right]$$

$$b\lambda^{a-1} \cancel{\exp(-b\lambda)} = (a-1)\lambda^{a-2} \cancel{\exp(-b\lambda)}$$

$$\lambda = \frac{(a-1)}{b}$$

2.43. $p(x|\sigma^2, q) = \frac{q}{2(2\sigma^2)^{1/q}\Gamma(1/q)} \exp\left(-\frac{|x|^p}{2\sigma^2}\right)$

Prove Normalization : $\int_{-\infty}^{\infty} p(x|\sigma^2, q)\, dx = 1$

$$\int_{-\infty}^{\infty} \frac{q}{2(2\sigma^2)^{1/q}\Gamma(1/q)} \exp\left(-\frac{|x|^p}{2\sigma^2}\right) dx$$

$$\frac{\cancel{2} q}{\cancel{2}(2\sigma^2)^{1/q}\Gamma(1/q)} \int_0^{\infty} e^{-\frac{|x|^p}{2\sigma^2}} dx = \frac{q}{(2\sigma^2)^{1/q}\Gamma(1/q)} \frac{1}{2}\sqrt{2\pi\sigma^2}$$

Stirling Approximation : $\Gamma(z) \approx \sqrt{2\pi}\, e^{-z} z^{z+1/2}$

$$= \frac{q^{1/2} \frac{1}{2}\sqrt{2\pi\sigma^2}}{(2\sigma^2)^{1/q} \cdot \sqrt{2\pi}\, e^{-1/q} \cdot \frac{1}{q}^{(1/q - 1/2)}}$$

$q = 1$, Normal.

Reduction when $q = 2$

$$p(x|\sigma^2, q) = \frac{\cancel{2}}{\cancel{2}(2\sigma^2)^{1/2}\Gamma(1/2)} \exp\left(-\frac{|x|}{2\sigma^2}\right) = \frac{1}{(2\sigma^2)^{1/2}\sqrt{\pi}} \exp^{-\frac{|x|^2}{2\sigma^2}}$$

Consider $t = y(x,w) + \epsilon$, ϵ random noise

Prove log likelihood function over w and σ^2

$$p(x, \sigma^2, q) = \frac{q}{2(2\sigma^2)^{1/q}\Gamma(1/q)} \exp\left(\frac{-|x|^q}{2\sigma^2}\right)$$

$$- \frac{[y(x,w) - \epsilon - (t)]^q}{2\sigma^2}$$

$$\ln p(x|\sigma^2, q) = \ln \frac{q}{2(2\sigma^2)^{1/q}\Gamma(1/q)} \exp - \frac{[y(x,w) + \cancel{\epsilon} - t]^q}{2\sigma^2}$$

$$= \ln \frac{q}{(2\sigma^2)^{1/q}\Gamma(1/q)} \exp$$

$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{N} |y(x,w) - t|^q - \frac{N}{q}\ln(2\sigma^2) + \text{const}$$

2.44. $N(x|\mu, \tau^{-1})$ $\quad p(\mu,\lambda) = N(\mu|\mu_0, (\beta\lambda)^{-1})\,Gam(\lambda|a,b)$

for $\vec{x} = \{x_1, \cdots, x_n\}$

$$= \left(\frac{\beta\lambda}{2\pi}\right)^{1/2} e^{-\frac{\beta\lambda}{2}(\mu-\mu_0)^2} \cdot \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$$

$$= \left(\frac{\beta\lambda}{2\pi}\right)^{1/2} \frac{b^a}{\Gamma(a)} \cdot \lambda^{a-1} \cdot e^{-\frac{\beta\lambda}{2}(\mu-\mu_0)^2 - b\lambda}$$

$$= \underbrace{\left(\frac{\beta\lambda}{2\pi}\right)^{1/2} \frac{b^a}{\Gamma(a)} \cdot \lambda^{a-1}}_{G} e^{-\left(\frac{\beta}{2}(\mu-\mu_0)^2 - b\right)\lambda}$$

$$e^{-\left(\frac{\beta}{2}(\mu^2 - 2\mu\mu_0 + \mu_0^2) - b\right)\lambda}$$

$$= G \cdot e^{-\left(\frac{\beta\lambda}{2}\mu^2\right)} \; e^{-\left[\frac{\beta}{2}[-2\mu\mu_0 + \mu_0^2] - b\right]\lambda}$$

$$= G \cdot e \quad e$$

$\sigma = \beta\lambda$

$Gam(\lambda | a, \frac{\beta}{2}[-2\mu\mu_0 + \mu_0^2] - b)$

2.45, Wishart Distribution

$$W(\Lambda|W,\nu) = B|\Lambda|^{(\nu-D-1)/2} \cdot e^{-\frac{1}{2}\mathrm{Tr}(W^{-1}\Lambda)}$$

Prove Wishart Distribution is a conjugate prior to a precision matrix of a Gaussian

$$N(x|\mu,\Lambda^{-1}) = N(\mu|\mu_0,(\beta\Lambda)^{-1}) \cdot W(\Lambda|W,\nu)$$

$$= N(\mu|\mu_0,(\beta\Lambda)^{-1}) \cdot B|\Lambda|^{(\nu-D-1)/2} e^{-\frac{1}{2}\mathrm{Tr}(W^{-1}\Lambda)}$$

$$= N(\mu|\mu_0,(\beta\Lambda)^{-1}) \cdot \frac{|\Lambda|^{(\nu-D-1)/2} e^{-\frac{1}{2}\mathrm{Tr}(W^{-1}\Lambda)}}{|W|^{-\nu/2}\left(2^{\nu D/2} \cdot \pi^{D(D-1)/4} \prod_{i=1}^{D}\Gamma\left(\frac{\nu+1-i}{2}\right)\right)}$$

$$= \left[\frac{\beta\Lambda}{2\pi}\right]^{-1/2} e^{-\frac{\beta}{2}(\mu-\mu_0)^T\Lambda(\mu-\mu_0)} \frac{|\Lambda|^{(\nu-D-1)/2} e^{-\frac{1}{2}\mathrm{Tr}(W^{-1}\Lambda)}}{|W|^{-\nu/2}\left(2^{\nu D/2}\pi^{D(D-1)/4}\prod_{i=1}^{D}\Gamma\left(\frac{\nu+1-i}{2}\right)\right)}$$

$\nu = 1$ Degree of Freedom

$$= \left[\frac{\beta\Lambda}{(2\pi)}\right]^{-1/2} e^{-\frac{\beta}{2}(\mu-\mu_0)^T\Lambda(\mu-\mu_0)} \frac{|\Lambda|^{D/2} \cdot e^{-1/2\,\mathrm{Tr}(W^{-1}\Lambda)}}{|W|^{-1/2}\left(2^{D/2}\pi^{D(D-1)/4}\prod_{i=1}^{D}\Gamma\left(\frac{2-i}{2}\right)\right)}$$

$$= \frac{(\beta\Lambda)^{-1/2}|\Lambda \cdot W|^{D-1} e^{-\frac{\beta}{2}(\mu-\mu_0)^T\Lambda(\mu-\mu_0)}\, e^{-1/2\,\mathrm{Tr}(W^{-1}\Lambda)}}{2^{D-1}\pi^{[2+(D(D+1))]/4}\prod_{i=1}^{D}\Gamma\left(\frac{2-i}{2}\right)}$$

$$= \frac{(\beta\Lambda)^{-1/2}|\Lambda \cdot W|^{D-1} e^{-\frac{1}{2}\left[\beta(\mu-\mu_0)^T(\mu-\mu_0)+W^{-1}\right]\Lambda}}{2^{D-1}\pi^{[2+(D(D+1))]/4}\cdot\prod_{i=1}^{D}\Gamma\left(\frac{2-i}{2}\right)}$$

2.46 $p(x|\mu, a, b) = \int_0^\infty N(x|\mu, \tau^{-1})\, \mathrm{Gam}(\tau|a, b)\, d\tau$

$$= \frac{b^a \sqrt{\tau}}{\sqrt{2\pi}\,\Gamma(a)} \int_0^\infty \tau^{a-1} e^{-\frac{\tau}{2}(x-\mu)^2} e^{-b\tau}\, d\tau$$

$$= \frac{b^a}{\sqrt{2\pi}\,\Gamma(a)} \int_0^\infty \tau^{(a+\frac{1}{2})-1} e^{-\left[\frac{(x-\mu)^2}{2}+b\right]\tau}\, d\tau$$

$$= \frac{b^a}{\sqrt{2\pi}\,\Gamma(a)} \frac{\Gamma(a+\frac{1}{2})}{\left[\frac{(x-\mu)^2}{2}+b\right]^{-a-1/2}} \quad ; z = \tau\left[\frac{(x-\mu)^2}{2}+b\right]$$

$v = 2a$

$\lambda = a/b$

$$\mathrm{St}(x|\mu, \lambda, v) = \frac{\Gamma(\frac{v}{2}+\frac{1}{2})}{\Gamma(v/2)} \left(\frac{\lambda}{\pi v}\right)^{1/2} \left[1 + \frac{\lambda(x-\mu)^2}{v}\right]^{-v/2-1/2}$$

2.47. $\lim_{v\to\infty} \mathrm{St}(x|\mu, \lambda, v) = \boxed{\frac{\Gamma(\infty)}{\Gamma(\infty)} \cancel{\left(\frac{\lambda}{\infty}\right)^{1/2}}} \left[1 + \frac{\lambda(x-\mu)^2}{\infty}\right]^{-\infty/2-1/2}$

$$= \text{Ignore} \left[1 + \frac{\lambda(x-\mu)^2}{\infty}\right]^{-\infty/2-1/2}$$

$$\propto e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \text{ as } \sigma^2 \to \infty$$

$$\int_0^\infty \left[1 + \cancel{\frac{z^T \Lambda z}{\nu}}\right]^{-\frac{D}{2}-\frac{\nu}{2}}$$

2.49 $St(x|\mu,\Lambda,\nu) = \int_0^\infty N(x|\mu,(\eta\Lambda)^{-1})\, Gam(\eta|\nu/2,\nu/2)\, d\eta$

$$E[z] = \int_0^\infty z\, N(x|\mu,(\eta\Lambda)^{-1})\, Gam(\eta|\nu/2,\nu/2)\, d\eta$$

$$= \frac{\Gamma(D/2+\nu/2)}{\Gamma(\nu/2)} \frac{|\Lambda|^{1/2}}{(\pi\nu)^{D/2}} \int_0^\infty \sqrt{t\nu}\left[1+\frac{z^2}{\nu}\right]^{-D/2-\nu/2} dz$$

$t = \frac{z^2}{\nu}$, $\sqrt{t\nu} = z$, $\frac{1}{2}(t\nu)^{-1/2} = dz$

$$\int_0^\infty \sqrt{t\nu}\,[1+t]^{-D/2-\nu/2}\, \frac{1}{2}(t\nu)^{-1/2}\, dt$$

$$= \frac{\Gamma(D/2+\nu/2)}{\Gamma(\nu/2)} \frac{|\Lambda|^{1/2}}{2(\pi\nu)^{D/2}} \int_0^\infty (t\nu)^{1/2-1/2}\,[1+t]^{-D/2-\nu/2}\, dt$$

$$= \int_0^\infty (t\nu)^{\ldots}\,[1+t]^{-D/2-\nu/2}\, dt$$

$$= \int_0^\infty (t\nu)^{\ldots - 1}\,[1+t]^{-(D+1)/2-(\nu-D)/2}\, dt$$

$$= B\left(\frac{D+1}{2}, \frac{\nu-D}{2}\right)$$

$$= \frac{\Gamma(D/2+\nu/2)|\Lambda|^{1/2}}{\Gamma(\nu/2)2(\pi\nu)^{D/2}}\left[\frac{\Gamma\left(\frac{D+1}{2}\right)\Gamma\left(\frac{\nu-D}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\right]^{-1}$$

$$= \frac{|\Lambda|^{1/2}\Gamma\left(\frac{D+1}{2}\right)\Gamma\left(\frac{\nu-D}{2}\right)}{\Gamma(\nu/2)2(\pi\nu)^{D/2}}$$ How is this μ?

$)^{\frac{\nu}{2}-1} d\eta$

$(x-\mu)\, d\eta$

$|\eta$

$d\eta$

$\left.\right]^{-D/2-1/2}$

$1/2$

$z = (x-\mu)$

sudo sh -c "sync && echo 3 > /proc/sys/vm/drop_caches"

dd if=./largefile of=/dev/null bs=4k

$$= \frac{\Gamma(D/2+V/2)V^{3/2}}{(\pi V)^{D/2}\cdot 2}\int_0^\infty (\cdot t)^{V2}[1+t]^{-D/2-V/2}\,dt$$

$$= \frac{\Gamma(D/2+V/2)V^{3/2}}{(\pi V)^{D/2}\cdot 2}\int_0^\infty t^{3/2-1}[1+t]^{-(D+1)/2+(V-D)/2}\,dt$$

$$= \frac{\Gamma(D/2+V/2)V^{3/2}}{\Gamma(V/2)(\pi V)^{D/2}\cdot 2}\cdot\frac{\Gamma\left(\frac{D+1}{2}\right)\Gamma\left(\frac{V-D}{2}\right)}{\Gamma\left[\frac{D+1}{2}-\frac{V+D}{2}\right]}$$

$$= \frac{\Gamma\left(\frac{D+1}{2}\right)\Gamma\left(\frac{V-D}{2}\right)V^{3/2}}{\Gamma(V/2)\,\pi^{D/2}\,V^{D/2}\cdot 2} = \frac{\Gamma\left(\frac{D}{2}+\frac{1}{2}\right)\Gamma\left(\frac{V}{2}+\frac{D}{2}\right)V}{\Gamma(V/2)\Gamma(D/2)\cdot 2}$$

2.50 $$\lim_{v\to\infty} St(x|\mu,\Lambda,v) = \lim_{v\to\infty}\frac{\Gamma(D/2+v/2)}{\Gamma(v/2)}\frac{|\Lambda|^{1/2}}{(\pi v)^{D/2}}\left[1+\frac{\Delta^2}{v}\right]^{-D/2-v/2}$$

$$\lim_{v\to\infty} \cancel{\log\left(St\left(x|\mu,\Lambda,v\right)\right)}$$

$$\lim_{v\to\infty}\log$$

$\boxed{\text{Graph?}}$ $\boxed{\text{LaHopitals Rule?}}$ Yes.

Next Question

2.51. $\exp(iA) = \cos A + i\sin A$ $\quad e^{iA}e^{-iA}=1=[\cos A+i\sin A][\cos A-i\sin A]$

$= \cos^2 A+\sin^2 A$

$\exp(iA)\exp(-iA)=1$

$\mathrm{Re}\, e^{i(A-B)} = e^{iA}e^{-iB} = \cos(A-B) = [\cos A\cos B+\sin A\sin B]$

$\mathrm{I}: e^{i(A-B)} = e^{iA-iB} = \sin(A-B) = \cos A\cdot\cos B + i\cos A\, i\sin B - i\sin B\cdot\cos A + \sin^2 B$

dumps.wikimedia.org/other/static_html_dumps/2008-03/

2.3. Complete the square of

$$\mu_N = \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{ML}$$

$$\frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2} ; \quad \sigma_N^2 = \frac{\sigma^2+N\sigma_0^2}{\sigma_0^2\sigma^2}$$

$$p(x) = N(\cancel{x}\mu_0 \mid \mu_N, \sigma_N^2)$$

$$= \frac{1}{\left[2\pi\left(\sigma_N^2\right)\right]^2} e^{\frac{1}{2\sigma^2}(\mu_0-\mu_N)^2}$$

$$(\mu_0-\mu_N)^2 = \mu_0^2 - 2\mu_N\mu_0 + \mu_N^2$$

$$= \mu_0^2 - 2\left[\frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{ML}\right]\mu_0 +$$

$$\left[\frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{ML}\right]^2$$

$$= \mu_0^2 - \frac{2\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0^2 - \frac{2N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{ML}\mu_0 +$$

$$\left(\frac{\sigma^2\mu_0}{N\sigma_0^2+\sigma^2}\right)^2 + 2\left(\frac{\sigma^2\mu_0}{N\sigma_0^2+\sigma^2}\right)\left(\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{ML}\right)$$

$$+ \left(\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\right)\mu_{ML}^2$$

2.38. Complete the square of: $p(x) = N(\mu_0 | \mu_N, \sigma_N^2)$

$$\text{where}: \mu_N = \frac{\sigma^2}{N\sigma_0^2 + \sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2}\mu_{ML}$$

$$\frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}$$

$$\therefore p(x) = N(\mu_0 | \mu_N, \sigma_N^2)$$

$$= \frac{1}{(2\pi\sigma_N^2)^{N/2}} e^{-\frac{1}{2\sigma_N^2}(\mu_0 - \mu_N)^2}$$

$$\text{where}: (\mu_0 - \mu_N)^2$$

$$= \mu_0^2 + -2\left[\cancel{\frac{\sigma^2}{N\sigma_0^2+\sigma^2}}\right]\mu_0^2 - 2\left[\cancel{\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}}\right]\mu_{ML}\mu_0$$

$$+ \left[\cancel{\frac{\sigma^2}{N\sigma_0^2+\sigma^2}}\right]\mu_0^2 + 2\left[\frac{\sigma^2}{N\sigma_0^2+\sigma^2}\right]\left[\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\right]\mu_{ML}\mu_0$$

$$+ \left(\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\right)\mu_{ML}^2$$

$$\cancel{\left[1 + \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\right]} = \left(\mu_0 - \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0 - \frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{ML}\right)\left(\mu_0 - \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0 - \frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{ML}\right)$$

$$= \mu_0^2 - \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0^2 - \frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_0\mu_{ML} - \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0^2 - \left[\frac{\sigma^2}{N\sigma_0^2+\sigma^2}\right]^2\mu_0^2$$

$$- \left[\frac{\sigma^2}{N\sigma_0^2+\sigma^2}\right]\left[\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\right]\mu_0\mu_{ML} - \left[\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\right]\mu_{ML}\mu_0 - \left[\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\right]\mu_{ML}$$

$$\times \left[\frac{\sigma^2}{N\sigma^2+\sigma^2}\right]\mu_0 + \left[\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\right]^2\mu_{ML}^2$$

$ax^2 + bx + c$

$a(x^2 + \frac{b}{a}x)$

$a(x + \frac{b}{2a}x)^2 - \frac{b^2}{2a}$

$$= \mu_0^2 - \left[\frac{2\sigma^2}{N\sigma_0^2+\sigma^2} - \left(\frac{\sigma^2}{N\sigma_0^2+\sigma^2}\right)^2\right]\mu_0^2 - \left[\frac{2N\sigma_0^2}{N\sigma_0^2+\sigma^2} - \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\right]\mu_{ML}\mu_0 -$$

$$= \left[1 - \frac{2\sigma^2}{N\sigma_0^2+\sigma^2} + \left(\frac{\sigma^2}{N\sigma_0^2+\sigma^2}\right)^2\right]\mu_0^2 - \left[\frac{2N\sigma_0^2}{N\sigma_0^2+\sigma^2} - \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\right]\mu_{ML}\cdot\mu_0 + \left[\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\right]^2\mu_{ML}^2$$

$$\cancel{\left[1 - \frac{2\sigma^2}{N\sigma^2+\sigma^2}\right.}$$

$$= \left[1 - \frac{2\sigma^2}{N\sigma_0^2+\sigma^2} + \left(\frac{\sigma^2}{N\sigma_0^2+\sigma^2}\right)^2\right]\left[\mu_0^2 - \frac{\left(\frac{2N\sigma_0^2}{N\sigma_0^2+\sigma^2} - \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\right)}{\left(1 - \frac{2\sigma^2}{N\sigma_0^2+\sigma^2} + \left(\frac{\sigma^2}{N\sigma_0^2+\sigma^2}\right)^2\right)}\mu_{ML}\cdot\mu_0\right] + \left[\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\right]^2\mu_{ML}^2$$

$$= \left[1 - \frac{2\sigma^2}{N\sigma_0^2+\sigma^2} + \left(\frac{\sigma^2}{N\sigma_0^2+\sigma^2}\right)^2\right]\left(\mu_0 - \frac{\left(\frac{2N\sigma_0^2}{N\sigma_0^2+\sigma^2} - \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\right)}{2\left(1 - \frac{2\sigma^2}{N\sigma_0^2+\sigma^2} + \left(\frac{\sigma^2}{N\sigma_0^2+\sigma^2}\right)^2\right)}\mu_{ML}\cdot\mu_0\right)^2 + \cdots$$

$$- \left[\frac{\left(\frac{2N\sigma_0^2}{N\sigma_0^2+\sigma^2} - \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\right)}{2\left(1 - \frac{2\sigma^2}{N\sigma_0^2+\sigma^2} + \left(\frac{\sigma^2}{N\sigma_0^2+\sigma^2}\right)^2\right)^2}\mu_{ML}\cdot\mu_0\right]^2 + \left[\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\right]^2\mu_{ML}^2$$

$$= a + b$$

Thus, $p(x) = N(\cancel{\mu} \mid \mu_N, \sigma_N^2)$

$$= \frac{1}{(2\pi\sigma_N^2)^{N/2}} \exp^{\frac{1}{2\sigma_N^2}[a+b]} = \frac{1}{(2\pi\sigma_N^2)^{N/2}} \exp^{\frac{1}{2\sigma_N^2}a} \cdot \exp^{\frac{1}{2\sigma_N^2}b}$$

2.49 Variance of Multivariate T-distribution:

$$\mathrm{Var}[x] = E[\Delta^2] - E[x]^2$$

$$E[\Delta^2] = \int_0^\infty x^2 \cdot S(x \mid \mu, \Lambda, \nu) = \frac{\Gamma(D/2+\nu/2)}{(\pi\nu)^{D/2}\Gamma(\nu/2)}\int_0^\infty x^2\left[1 + \frac{(x-\mu)^T\Lambda(x-\mu)}{\nu}\right]^{-D/2-\nu/2} dx$$

$$= \frac{\Gamma(D/2+\nu/2)}{(\pi\nu)^{D/2}\Gamma(\nu/2)}\int_0^\infty \Delta^2 \left[1 + \frac{\Delta^2}{\nu}\right]^{-D/2-\nu/2} d\Delta \qquad \frac{\Delta^2}{\nu} = t\,;\ \Delta = \sqrt{t\nu}$$

$$d\Delta = \frac{1}{2}(t\nu)^{-1/2}\nu\, dt$$

$$= \frac{\Gamma(D/2+\nu/2)}{(\pi\nu)^{D/2}\Gamma(\nu/2)}\int_0^\infty (\nu t)\left[1+t\right]^{-D/2-\nu/2}\frac{1}{2}\frac{(\nu)^{1/2}}{t^{1/2}}\, dt$$

$$= \frac{\Gamma(D/2+\nu/2)\,\nu^{3/2}}{(\pi\nu)^{D/2}\cdot 2}\int_0^\infty (\nu t)^{\nu/2}\,[1+t]^{-D/2-\nu/2}\,dt$$

$$= \frac{\Gamma(D/2+\nu/2)\,\nu^{3/2}}{(\pi\nu)^{D/2}\cdot 2}\int_0^\infty t^{3/2-1}\,[1+t]^{-(D+1)/2-(\nu-D)/2}\,dt$$

$$= \frac{\Gamma(D/2+\nu/2)\,\nu^{3/2}}{\Gamma(\nu/2)(\pi\nu)^{D/2}\cdot 2}\cdot\frac{\Gamma\left(\frac{D+1}{2}\right)\Gamma\left(\frac{\nu-D}{2}\right)}{\Gamma\left[\frac{D+1}{2}+\frac{\nu-D}{2}\right]}$$

$$= \frac{\Gamma\left(\frac{D+1}{2}\right)\Gamma\left(\frac{\nu-D}{2}\right)\nu^{3/2}}{\Gamma(\nu/2)\,\pi^{D/2}\,\nu^{D/2}\cdot 2} = \frac{\Gamma\left(\frac{D}{2}+\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}-\frac{D}{2}\right)\nu}{\Gamma(\nu/2)\,\Gamma(D/2)\cdot 2}$$

2.50 $\lim_{\nu\to\infty} St(x|\mu,\Lambda,\nu) = \lim_{\nu\to\infty}\frac{\Gamma(D/2+\nu/2)}{\Gamma(\nu/2)}\frac{|\Lambda|^{1/2}}{(\pi\nu)^{D/2}}\left[1+\frac{\Delta^2}{\nu}\right]^{-D/2-\nu/2}$

$\therefore \lim_{\nu\to\infty} \cancel{\log\left(St(x|\mu,\Lambda,\nu)\right)}$

$\lim_{\nu\to\infty} \log$

$\boxed{\text{Graph?}}$ $\boxed{\text{L'hopitals Rule?}}$ Yes.

Next Question

2.51. $\exp(iA) = \cos A + i\sin A \quad \therefore e^{iA}e^{-iA} = 1 = [\cos A + i\sin A][\cos A - i\sin A]$

$= \cos^2 A + \sin^2 A$

$\exp(iA)\exp(-iA) = 1$

$\mathrm{Re}\; e^{i(A-B)} = e^{iA}e^{-iB} = \cos(A-B) = [\cos A\cos B + \sin A\sin B]$

$\mathrm{Im}: e^{i(A-B)} = e^{iA-iB} = \sin(A-B) = \cos A\cdot\cos B + i\cos A\, i\sin B$

$-\sin B\cdot\cos A$

$+\sin^2 B$

2.52 $p(\theta, \theta_0, m) = \frac{1}{2\pi I_0(m)} \exp\{m\cos(\theta-\theta_0)\}$

if $\xi = m^{1/2}(\theta+\theta_0)$, and $\cos\alpha = 1 - \frac{\alpha^2}{2} + O(\alpha^4)$:

Prove $\lim_{m\to\infty} p(\theta, \theta_0, m) = N(x \mid \mu, \sigma^2)$

$$\lim_{m\to\infty} \frac{1}{2\pi I_0(m)} \exp\{m\cos(\theta-\theta_0)\}$$

$$\frac{1}{\cancel{2\pi}\frac{1}{2\pi}\int_0^{2\pi}\exp\{m\cos\theta\}d\theta} \exp\{m\cos(\theta-\theta_0)\}$$

$$\frac{1}{\int_0^{2\pi}\exp\{m(1-\frac{\theta-\theta_0}{2}+O(\alpha^4))\}} \exp\{m(1-\frac{\theta-\theta_0}{2})\}$$

$$\frac{1}{\int_0^{2\pi}\exp\{m-\frac{m(\theta-\theta_0)}{2}\}} \exp\{m-\frac{m(\theta-\theta_0)}{2}\}$$

$$\frac{1}{\cancel{\exp\{m\}}\int_0^{2\pi}\exp\{\frac{-\xi_m}{2(\theta-\theta_0)}\}d\theta} \cancel{\exp}\cdot\exp^{-\frac{\xi_m^2}{2(\theta+\theta_0)}}$$

Gaussm

2.53 $\sin(A-B) = \cos B\sin A - \cos A\sin B$ $\quad \sin(\theta_n-\theta_0) = \cos\theta_n\sin\theta_0 - \cos\theta_n\sin\theta_0$

$\sum_{n=1}^{N}\sin(\theta_n-\theta_0)=0$ $\quad \sum_{n=1}^{N}\sin(\theta_n-\theta_0) = \sum_{n=1}^{N}\{\cos\theta_0\sin\theta_n - \cos\theta_n\sin\theta_0\} = 0$

$$\theta_{ML} = \tan^{-1}\frac{\sum\sin\theta_n}{\sum\cos\theta_n}$$

2.54. $\frac{d}{d\theta}p(\theta|\theta_0,m) = \frac{d}{d\theta}\frac{1}{2\pi I_0(m)}\exp\{m\cos(\theta-\theta_0)\} = \frac{-1}{2\pi \cancel{I_0(m)}} e^{m\cancel{\cos(\theta-\theta_0)}} \cdot m(-\sin(\theta-\theta_0)) = 0$

$\theta = \theta_0$: Max

$\frac{d^2}{d^2\theta}p(\theta|\theta_0,m) = \frac{1}{2\pi \cancel{I}(m)}\left[\exp\{m\cos(\theta-\theta_0)\}m^2(-\sin(\theta-\theta_0)) + \exp\{m\cos(\theta-\theta_0)\}m\cos(\theta-\theta_0)\right]$

$= 0$ $\therefore \theta = \theta_0$ its Th[illegible] (2π) ... ① Log ② Remove $\log(\cos(\theta-\theta_0))$

2.58. $-\nabla \ln g(\eta) = E[u(x)]$

$$g(\eta)\int h(x)\exp\{\eta^T u(x)\}dx = 1$$

$$-\nabla\nabla \ln g(\eta) = E[u(x)u(x)^T] - E[u(x)]E[u(x)^T] = \mathrm{cov}[u(x)]$$

$$\ln g(\eta) + \int \ln \int h(x)dx + \ln\int \exp \eta^T u(x)\, dx = \ln(1)$$

$$-\nabla \ln(g(\eta)) = E[u(x)] = \nabla\left[\ln\int h(x)dx + \ln\int \exp \eta^T u(x) dx\right]$$

$$g(\eta)\int h(x)\exp\{\eta^T u(x)\}dx = 1$$

$$\nabla g(\eta)\int h(x)\exp\{\eta^T u(x)\}dx + g(\eta)\int h(x)\exp\{\eta^T u(x)\}u(x)dx = 0$$

$$-\frac{1}{g(\eta)}\nabla g(\eta) = g(\eta)\int h(x)\exp\{\eta^T u(x)\}u(x)dx = E[u(x)]$$

$$-\frac{1}{g(\eta)}\nabla^2 g(\eta) + \nabla g(\eta)\nabla\frac{1}{g(\eta)} =$$

$$-\frac{1}{g(\eta)}\left[\nabla^2 g(\eta) + \frac{\nabla g(\eta)}{g(\eta)^2}\right] = \cancel{\nabla \ln g(\eta)\cdot\nabla g(\eta) - \frac{\nabla \ln g(\eta)}{g(\eta)}}$$

$$= \cancel{\nabla\nabla \ln g(\eta)\, g(\eta) - \nabla \ln g(\eta)}$$

$$-\frac{1}{g(\eta)}\left[E[u(x)u(x)^T] - E[u(x)]^T[E[u(x)]]\right] - \frac{E[u(x)]}{g(\eta)}$$

2.59. $p(x|\sigma) = \frac{1}{\sigma}f\left(\frac{x}{\sigma}\right) \quad y = x/\sigma$

$$= \frac{1}{\sigma}f(y) \qquad dy = \frac{dx}{\sigma}$$

$$\int p(x|\sigma)dx = 1 = \int \frac{1}{\sigma}f(y)\sigma\, dy = \int f(y)dy$$

2.60

$p(x) = h_1, h_2, h_3, h_4, h_5 + \cdots + h_i$ $\quad \Delta_i$

$N = 1\ 2,3\ 4\ 5 + \cdots + n_i$

$\int_1^{n_i} p(x)\,dx = 1$ $\qquad \sum_i p(x)\Delta_i = 1$

$$\frac{d}{dh}\left[\sum_{i=1}^{n} \ln p(x_i) + \lambda\left(\sum_{i=1}^{n} p(x_i)\Delta_i - 1\right)\right] = 0$$

$$\frac{d}{dh}\left[\sum_{i=1}^{n} \ln h_k + \lambda\left(\sum_{i=1}^{n} h_i\Delta_i - 1\right)\right] = 0$$

$$\frac{n_i}{h_i} + \lambda\Delta_k = 0$$

$$h_k = \frac{n_i}{\lambda\Delta_k}$$

2.61 $\quad p(x|C_k) = \dfrac{K_k}{N_k V}$ $\quad \displaystyle\int_{-\infty}^{\infty} p(x|C_k)\,dx = \int_{-\infty}^{\infty} \frac{K_k}{N_k V}\,dx = \int_0^{\infty} \frac{\sum_{i=1}^{N} k\left(\frac{x-x_i}{h}\right)}{N_k V}\,dx$

$= \infty$

Chapter 3

1. $\sigma(a) = \dfrac{1}{1+\exp(-a)}$ $\quad 2\sigma(2a) - 1 = \dfrac{2}{1+\exp(-2a)} - 1 = \dfrac{1 - \exp(-2a)}{1+\exp(-2a)} = \dfrac{1 - \exp(-a)\cdot\exp(-a)}{1+\exp(-a)\exp(-a)}$

$= \dfrac{\exp(a) - \exp(-a)}{\exp(a) + \exp(-a)} = \tanh(a)$

$$y(x, w) = w_0 + \sum_{j=1}^{M} w_j \sigma\left(\frac{x - \mu_j}{s}\right)$$

$\vec{w} = \{w_1, w_2, w_3 \cdots w_m\}$

$\vec{u} = \{u_1, u_2, u_3 \cdots u_m\} = \left\{\frac{w_1+1}{2}, \frac{w_2+1}{2}, \frac{w_3+1}{2} + \cdots + \frac{w_m+1}{2}\right\}$

3.2 $\Phi(\Phi^T\Phi)^{-1}\Phi^T$ $\cancel{W_{ML} = (\Phi^T\Phi)^{-1}\Phi t}$

$\cancel{\frac{W_{ML}}{t} = (\Phi^T\Phi)^{-1}\Phi}$

$\longrightarrow \frac{W_{ML}}{t} \cdot \Phi^T$

$\Phi(\Phi^T\Phi)^{-1}\Phi^T$

$= \Phi\Phi^{-1} = I$, $W = (\Phi^T\Phi)^{-1}\Phi^T t$

$= \Phi^{\dagger} t = \Phi^{-1} t$

3.3 $E_D(W) = \frac{1}{2}\sum_{n=1}^{N} r_n \{t_n - W^T\phi(x_n)\}^2$

$t_n \sim r_n > 0$

Prove w^* which minimizes this Erf

$\frac{dE_D(W)}{\omega^*} = 0 = t_n - w^*\phi(x_n)$

$w^* = \frac{t_n}{\phi(x_n)}$

1. This function approximates data dependent noise variance

2. Replicated datapoints: increase (or decrease) weighting form

3.4. $y(x, W) = W_0 + \sum_{i=1}^{D} W_i x_i$ $E_D(W) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, w) - t_n\}^2$

$E[\epsilon_i] = 0$, $E[\epsilon_i \epsilon_j] = \delta_{ij}\sigma^2$

$y(X_{\epsilon i}, W) = W_0 + \sum_{i=1}^{D} W_i (X_i + \epsilon_i) = W_0 + \sum_{i=1}^{D} W_i (X_{\epsilon i})$

Prove Minimizing E_D is equal to minimizing sum-of-square error without noise, but weight-decay regularization

$\frac{\partial E_D}{\partial W} = 0 = y(X_{\epsilon i}, W) - t_n$

$\frac{\partial E_D}{\partial W} = 0 > y(X_{\epsilon i}, W) - t_n$

3.5 Lagrange Multiplier

Prove Minimization of $\frac{1}{2}\sum_{n=1}^{N}\{t_n - W^T\phi(x_n)\}^2 + \frac{\lambda}{2}\sum_{j=1}^{M}|w_j|^q = f$ ①

3.5 is equal to

Minimization of $E_D(W) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - W^T\phi(x_n)\}^2$ with constraint ②

① $\frac{\partial f}{\partial W} = 0 = \frac{q\lambda}{2}\sum_{j=1}^{M}|W_j|^{q-1} - \sum_{n=1}^{N}\{t_n - W^T\phi(x_n)\}\phi(x_n)$

② $\frac{\partial E_{D,\lambda}}{\partial W} = 0 = \lambda W^T W - \sum_{n=1}^{N}\{t_n - W^T\phi(x_n)\}\phi(x_n)$ E.11

$L(W,\lambda) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - W^T\phi(x_n)\}^2 + \frac{\lambda}{2}\left(\sum_{j=1}^{N}|W_j|^q - \eta\right)$

$\sum_{j=1}^{M}|W_j|^2 = \eta$

3.6. $p(t|W,\Sigma) = N(t|y(x,W),\Sigma)$

where $y(x,W) = W^T\phi(x)$ $\phi(x_n) = \{x_1, x_2, x_3 + \cdots + x_n\}$

$t_n = \{t_1, t_2, t_3 + \cdots + t_n\}$

Prove W_{ML} for W is $W_{ML} = (\Phi^T\Phi)^{-1}\Phi^T t$

$$p(t|W,\Sigma) = \frac{1}{(2\pi|\Sigma|)^{N/2}} e^{\Sigma(t_n - y(x,W))\Sigma^{-1}(t_n - y(x,W))/2}$$

$$= \frac{1}{(2\pi|\Sigma|)^{N/2}} e^{\Sigma(t_n - W^T\phi(x))\Sigma^{-1}(t_n - W^T\phi(x))/2}$$

Log liklihood:

$$\ln p(t|W,\Sigma) = -\frac{N}{2}\ln(2\pi|\Sigma|) - \Sigma^{-1}(t_n - W^T\phi(x))\Sigma^{-1}(t_n - W^T\phi(x))/2$$

$$\frac{\partial \ln p(t|W,\Sigma)}{\partial W} = \Sigma^{-1}(t_n - W^T\phi(x))(-\phi(x))^T = 0$$

$t_n = W^T\phi(x)$; $\phi^T(x)t = W^T\phi(x)\phi^T$

$W = \phi^+ t$

Maximum likelihood of $\Sigma^{-} = E[[t - y(x,W)]]^2 = \frac{1}{N}\int(t - W^T\phi(x))^T(t - W^T\phi(x))$

3.7 $p(w|t) = N(w|m_n, S_n)$

$m_n = S_N(S_0^{-1} m_0 + \beta\phi^T t)$

$S_N^{-1} = S_0^{-1} + \beta\phi^T\phi$

Complete square of $p(w|t)$ with respect to w

$$p(w|t) = \frac{1}{(2\pi S_n)^{n/2}} \exp\Big\{\underbrace{-(w-m_n)^T S_n^{-1}(w-m_n)/2}_{a}\Big\}$$

$$a : (w-m_n)^T S_n^{-1}(w-m_n)/2 = \frac{w^T S_n^{-1} w - m_n^T S_n^{-1} w - w^T S_n^{-1} m_n + m_n^T S_n m_n}{2}$$

$$= \frac{w^T S_n^{-1} w}{2} - m_n^T S_N^{-1} w + m_n^T S_n^{-1} m_n$$

$$= \frac{S_N^{-1}}{2}\left(w^T w - 2m_n^T w\right) + m_n^T S_n^{-1} m_n$$

$$= \frac{S_N^{-1}}{2}\left(w + m_n^T w\right)^2 + \frac{t(m_n^T w)^2}{2} + m_n^T S_n^{-1} m_n$$

$$\Rightarrow p(w|t) = \frac{1}{(2\pi S_n)^{n/2}} \exp\left\{\frac{S_n^{-1}}{2}\left(w - \frac{m_n w}{2}\right)^2 + \frac{S_n^{-1} m_n^T m_n^T m^2}{-4} + m_n^T S_n m_n\right\}$$

3.8 N Data Points, $\vec{w}$

$P(w|t) = N(w|m_n, S_n)$

N+1 Datapoints, (x_{N+1}, t_{N+1})

$$P(w_{n+1}, t_{N+1}) \sim p(w|t_n) : \quad p(w|t) \cdot p(w_n|t_n) = p(w_{n+1}, t_{N+1})$$

$$\frac{1}{(2\pi S)^{1/2}} \exp\{(w-m)S^{-1}(w-m)/2\} \cdot \frac{1}{(2\pi S_n)^{n/2}} \exp\{(w-m_n)S_n^{-1}(w-m_n)/2\}$$

$$= \frac{1}{(2\pi S)^{(n+1)/2}} \exp\{(w-m)S(w-m)/2 + (w-m_n)S_n^{-1}(w-m_n)/2\}$$

$$= \frac{1}{(2\pi S)^{(n+1)/2}} \exp\left\{\frac{w^T S^{-1} w}{2} - m^T S^{-1} w + \frac{m^T S^{-1} m}{2} + \frac{w^T S_n^{-1} w}{2} - m_n^T S_n^{-1} m_n + \frac{m_n^T S^{-1} m_n}{2}\right\}$$

$$= \frac{1}{(2\pi S)^{(n+1)/2}} \exp\left\{\frac{w^T S_{n+1}^{-1} w}{2} - m_{N+1}^T S_{N+1}^{-1} w + \frac{m_{n+1}^T S_{N+1}^{-1} m_{n+1}}{2}\right\}$$

$$= \frac{1}{(2\pi S)^{(n+1)/2}} \exp\left\{\frac{S_{n+1}^{-1}}{2}\left(w - \frac{m_{n+1}}{2}\right)^2 - \frac{S_{n+1}^{-1}}{2}\left(\frac{m_{N+1}}{2}\right)^2 + m_{N+1}^T S_{N+1}^{-1} m_{N+1}\right\}$$

$$= N(w|m_{N+1}, S_{N+1}) \cdot \text{Liklihood}$$

3.9 $p(w|m_{N+1}, S_{N+1}) = p(w|m, S) \cdot p(w|m_N, S_N)$

$= N(w|m, S)\, N(w|m_N, S_N)$ = where, $m_N = S_N(S_0^{-1} m_0 + \beta\phi^T t)$.

$S_N^{-1} = S_0^{-1} + \beta \Phi^T \Phi$

Proof $p(z) = p(y) \cdot p(x|y)$

$= N(x|\mu, \Lambda^{-1}) \cdot N(y|Ax+b, L^{-1})$

$\tilde{=} N(w|m, S) \cdot N(w|S_N(S_0^{-1} m_0 + \beta\phi^T t), (S_0^{-1} + \beta \Phi^T \Phi)^{-1})$

$N(w|m_{N+1}, S_{N+1}) = N(w|S_{N+1}(S_0^{-1} m_0 + \beta\phi^T t), (S_0^{-1} + \beta\phi^T\phi)^{-1})$

3.10 $p(y) = N(y|A\mu + b, L^{-1} + A\Lambda^{-1}A^T)$

$p(t|X, \alpha, \beta) = \int p(t|W, \beta)\, p(w|t, \alpha, \beta)\, dw$ $\boxed{\text{Baysian Counterparts}}$

$= \int p(t|X, W, \beta)\, p(W|\ldots)\, dx$

$= \int N(t|y(x,w), \beta^{-1}) \cdot N(w|m_N, S_N)\, dw$ $(w|0, \alpha^{-1} I)$

$= \int N(t|w^T\phi(x), \beta^{-1}) \cdot N(w|m_n, S_n)\, dw$

$\doteq \int N(t| \cancel{\phi(x) m_n^T}, \beta^{-1} + \cancel{w}\, \phi(x) S_n \cancel{w}^T \phi(x)^T)\, dw$

$\Rightarrow N(t|\phi(x) m_n^T, \underbrace{\beta^{-1} + \phi(x) S_n \phi(x)^T}_{\sigma_N^2})$

3.11 $(M + vv^T)^{-1} = M^{-1} - \dfrac{(M^{-1}v)(v^T M^{-1})}{1 + v^T M^{-1} v}$ $S_N^{-1} = S_0^{-1} + \beta\phi^T\phi = S_0 - \dfrac{S_0 \phi \phi^T S_0^{-1}}{1 + \phi^T S_0 \phi} =$

Prove uncertainty $\sigma_N^2 \geq \sigma_{N+1}^2$

$\frac{1}{\beta} + \phi(x)^T S_N \phi(x)$: $\lim_{n\to\infty} \frac{1}{\beta} + \phi(x)^T S_n \phi(x) = \lim_{n\to\infty} \left(\frac{1}{\beta} + \underbrace{\phi(x)^T \frac{(1 + \phi^T S_0 \phi)}{S_0 \phi \phi^T S_0^{-1}}\phi(x)}_{=0}\right)$

ie β governance of noise.

3.12 $p(t|X, w, \beta) = \prod_{n=1}^{N} N(t_n | w^T \phi(x_n), \beta^{-1})$

Conjugate prior: $p(w, \beta) = N(w | m_0, \beta^{-1} S_0)\, Gam(\beta | a_0, b_0)$

Prove posterior is of form:

$$p(w, \beta | t) = N(w | m_N, \beta^{-1} S_n)\, Gam(\beta | a_n, b_n)$$

Then, $\underbrace{p(w, \beta | t)}_{\text{Posterior}} = \underbrace{p(w, \beta)}_{\text{Likelihood}} \underbrace{p(t | X, w, \beta)}_{\text{prior}}$

$= N(w | m_0, \beta^{-1} S_0)\, Gam(\beta | a_0, b_0) \prod_{n=1}^{N} N(t_n | w^T \phi(x_n), \beta^{-1})$

$= \cancel{N(w | m_0, \beta^{-1} S_0) \prod_{n=1}^{N} N(t_n | \sum_{j=1}^{M} w_j \phi_j(x_n), \beta^{-1})}\, Gam(\beta | a_0, b_0)$

$= N(w | \beta S_0 \Phi^T t, \beta^{-1}(S_0)) \prod_{n=1}^{N} N(t_n | \sum_{j=1}^{M} w_j \phi_j(x_n), \beta^{-1})\, Gam(\beta | a_0, b_0)$

$= \cancel{N(w | \beta S_0 \Phi^T t, \beta^{-1}(S_0))\, Gam(t_n | a_n, b_n)}$

$= \cancel{N(w, \beta | m_0, a_0}$,

$= Gam(w$

3.13. $p(t | x, t)$?

Prove $p(t | x, t) = St(t | \mu, \lambda, v)$

$= \int_0^\infty N(w | m_n, \beta^{-1} S_n)\, Gam(\beta | a_n, b_n)\, d\beta$

$= \frac{(S_n / \beta)^{n/2}}{(2\pi)^{n/2}} \exp\left\{ \frac{S_n}{2\beta} (w - m_n)^2 \right\} \cdot \frac{1}{\Gamma(a_n)} b_n^{a_n} \beta^{a_n - 1} \exp\{-b_n \beta\}\, d\beta$

Substituting: $v = 2a_n$; $S_n = b_n / a_n$; $\beta =$

Re-attempt

3.14. $k(x,x') = \beta\phi(x)^T S_N \phi(x')$ $\quad S_N^{-1} = \alpha I + \beta\phi^T\phi$

Suppose $\phi_j(x)$ $\phi_0(x) = 1$

$$\sum_{n=1}^{N} \psi_j(x_n)\psi_k(x_n) = I_{jk} \begin{Bmatrix} j=k & 0 \\ j\neq k & 1 \end{Bmatrix}$$

$$\psi_0(x) = 1$$

Show that $\alpha = 0$, the equivalent kernel can be written

$$k(x,x') = \psi(x)^T\psi(x') \qquad \psi = (\psi_1, \ldots, \psi_m)^T$$

Therefore,

$$k(x,x') = \beta\phi(x)^T S_n \phi(x') \text{ for } y(x,m_n) = \sum_{n=1}^{N} k(x,x)t_n$$

$$= \sum_{n=1}^{N} \beta\phi(x)^T S_n \phi(x') t_n$$

$$= \sum_{n=1}^{N} \beta\phi(x)^T[\alpha I + \beta\psi^T\psi]\phi(x') t_n$$

$$= \sum_{n=1}^{N} \beta\phi(x)^T[\beta\psi^T\psi]\phi(x') t_n$$

$$= k(x,x') = \psi(x)_j^T \psi(x')_k$$

if $j=k$, then $k(x,x') = 1$

$$= \sum_{n=1}^{N} \beta\phi(x)^T\phi(x')$$

$$= 1$$

3.24 $p(t) = \dfrac{p(t|W,\beta)\,p(W,\beta)}{p(W,\beta|t)} = \dfrac{\cancel{\left(\frac{\beta}{2\pi}\right)^{N/2}} \exp\left\{\frac{-\beta}{2}(t-\phi W)^T(t-\phi W)\right\}\cdot\left(\frac{\beta}{2\pi}\right)^{M/2}|S_0|^{-1/2}\exp\left\{\frac{-\beta}{2}(W-m_0)^T S_0^{-1}(W-m_0)\right\}}{\cancel{\left(\frac{\beta}{2\pi}\right)^{N/2}}|S_N|^{-1/2}\exp\left\{\frac{-\beta}{2}(W-m_n)^T S_N^{-1}(W-m_n)\right\}\frac{b_N^{a_n}}{\Gamma(a_n)}\beta^{a_n-1}e^{-b_n\beta}}$

Not enough Room.

$$= \frac{N(t|W,\beta)\,N(W|m_0,\beta^{-1}S_0)\,\mathrm{Gam}(\beta|a_0,b_0)}{N(W|m_n,\beta^{-1}S_N)\,\mathrm{Gam}(\beta|a_n,b_n)}$$

$$= \frac{\cancel{\left(\frac{\beta}{2\pi}\right)^{N/2}}\exp\left\{\frac{-\beta}{2}\cancel{(t-\phi W)^T}(t-\phi W)\right\}\cdot\left(\frac{\beta}{2\pi}\right)^{M/2}|S_0|^{-1/2}\exp\left\{\cancel{\frac{-\beta}{2}(W-m_0)^T}S_0^{-1}(W-m_0)\right\}\frac{b_0^{a_0}}{\Gamma(a_0)}\beta^{a_0-1}\cancel{e^{-b_0\beta}}}{\cancel{\left(\frac{\beta}{2\pi}\right)^{N/2}}|S_N|^{-1/2}\exp\left\{\frac{-\beta}{S_N}\cancel{(W-m_n)^T}(W-m_n)\right\}\frac{b_n^{a_n}}{\Gamma(a_n)}\beta^{a_n-1}\cancel{e^{-b_N\beta}}}$$

4.1 Continued. $X = \sum_n \alpha_n x_n$; $\alpha_n \geq 0$; $\sum_n \alpha_n = 1$

$\hat{w}^T x_n + w_0 > 0$
$\hat{w}^T y_n + w < 0$
Prove if $\hat{w}^T x_n + w_0 = \hat{w}^T y_n + w_0$ the sets of points do not intersect.

$w^T(x_n - y_N) > 0 \quad X_N > y_N$;

$w^T(x_n - y_N) < 0$; $x_N < y_N$; $f(\ldots)$

$w^T(x_n - y_N) = 0$; $x_N = y_N$; $f(x_N) = f(y_N)$

4.2 $E_D(\tilde{W}) = \frac{1}{2}\mathrm{Tr}\left\{(\tilde{X}\tilde{W} - T)^T(\tilde{X}\tilde{W} - T)\right\}$; $a^T t_n + b = 0$; $t_n = T_{nx}$.

Prove $y(x) = \tilde{W}^T\tilde{x} = T^T(\tilde{X}^+)^T\tilde{x}$ and $a^T y(x) + b = 0$

How? $\phi_0(x) = 1$ and w_0

$$\frac{dE_D(\tilde{W})}{d\tilde{W}} = \frac{d}{d\tilde{w}}\frac{1}{2}\mathrm{Tr}\left\{(\tilde{X}\tilde{W} - T)^T(\tilde{X}\tilde{W} - T)\right\} = (\tilde{X}\tilde{W} - T)^T\tilde{X} = 0$$

$\tilde{X}^T\tilde{W}X - \tilde{T}^T X = 0$

$\tilde{W} = (\tilde{T}^T X)(X^T X)^{-1}$

$= \tilde{T}^T\tilde{X}^+$

$y(x) = \hat{W}^T\tilde{x} = T^T\tilde{X}^+ x$

3.15 $E(m_n) = \frac{\beta}{2}\|t - \phi m_n\|^2 + \frac{\alpha}{2} m_N^T m_N$

Prove $2E(m_n) = N$

$\boxed{3.92 \,\&\, 3.95}$ $\alpha = \frac{\gamma}{m_N^T m_n}$ $\frac{1}{\beta} = \frac{1}{N-\gamma}\sum_{n=1}^{N}\{t_n - m_N^T \phi(x_n)\}^2$

$$E(m_n) = \frac{1}{\frac{2}{N-\gamma}\cancel{\sum_{n=1}^{N}\{t_n - m_N^T\phi(x_n)\}^2}}\cancel{\|t-\phi m_n\|^2} + \frac{\gamma}{2\cancel{m_N^T m_n}}\cancel{m_N^T m_N}$$

$$= \frac{N-\gamma}{2} + \frac{\gamma}{2} = \frac{N}{2} - \cancel{\frac{\gamma}{2}} + \cancel{\frac{\gamma}{2}}$$

$$\boxed{2E(m_n) = \frac{2N}{2} = N}$$

3.16 $p(t|\alpha,\beta) = \int p(t|w,\beta)\,p(w|\alpha)\,dw$

$p(y) = N(y|A\mu + b, L^{-1} + A\Lambda^{-1}A^T)$

$p(x|y) = N(x|\Sigma\{A^T L(y-b) + \Lambda\mu\}, \Sigma)$ $\Sigma = (\Lambda + A^T L A)^{-1}$

$$= \left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\alpha}{2\pi}\right)^{M/2}\int \exp\left\{-\frac{\beta}{2}\sum\{t_n - w^T\phi(x_n)\}^2\right\}\exp\left\{-\frac{\alpha}{2} I w^T w\right\} dw$$

$$= \left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\alpha}{2\pi}\right)^{M/2}\int \exp\left\{-\frac{\beta}{2}\sum\{t_n - w^T\phi(x_n)\}^2 - \frac{\alpha}{2} I w^T w\right\} dw$$

"Zero Mean Isotropic Gaussian"

$$= \left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\alpha}{2\pi}\right)^{M/2}\int \exp\left\{-\frac{\beta}{2}\sum\{t_n^2 - 2t_n w^T\phi(x_n) + w^T\phi(x_n) w^T\phi(x_n)\} - \frac{\alpha}{2} I w^T w\right\} dw$$

$$= \left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\alpha}{2\pi}\right)^{M/2}\int \exp\underbrace{\left\{-\frac{\alpha}{2} I w^T w - \frac{\beta}{2}\sum\{2t_n w^T\phi(x_n) + w^T\phi(x_n) w^T\phi(x_n)\}\right\}}_{\text{Complete the Square}} \cdot \exp^{-\frac{\beta}{2}\sum t_n^2}\, dw$$

3.4 cont...

$$\frac{\partial E_{D,x_{\epsilon}}}{\cancel{\partial W}} = \frac{\cancel{\partial}}{\partial W}\left[\frac{1}{2}\sum\{y_{\epsilon i}(x_{\epsilon i}, W) - t_n\}^2\right] = \frac{\partial}{\partial W}\left[\frac{1}{2}\sum_{i=1}^{N}\{(W_0 + \sum_{i=1}^{D} W_i(x_{ni} + \epsilon_{ni})) - t_n\}^2\right]$$

$$= \frac{\cancel{\partial}}{\cancel{\partial W}}\left[\frac{1}{2}\sum_{i=1}^{N}\{(W_0 + \sum_{i}^{N} W_i x_{\epsilon i}) - t\}^2\right] = \frac{\partial}{\partial W}\left[\frac{1}{2}\sum_{i=1}^{N}(y_{\epsilon i}^2 - 2y_{\epsilon i}t + t_n^2)\right]$$

$$= \frac{\cancel{\partial}}{\partial W}\left[\frac{1}{2}\sum_{i=1}^{N}(W_0 + \sum_{i=1}^{N} W_i x_{\epsilon i})^2 - \sum_{i=1}^{N}(W_0 + \sum_{i=1}^{N} W_i x_{\epsilon i})t_n + \frac{1}{2}\sum_{i=1}^{N} t_n^2\right]$$

$$= \frac{\cancel{\partial}}{\partial W}\left[\frac{1}{2}\sum_{i=1}^{N}(W_0^2 + 2\sum_{i}^{N} W_i x_i \cdot W_0 + \sum_{i=1}^{N}(W_i x_i)^2) - \sum_{i=1}^{N} W_0 t_n - \sum_{i=1}^{N} W_i x_{\epsilon i} t_n + \frac{1}{2}\sum_{i=1}^{N} t_n^2\right]$$

$$E\left[\left(\sum_{i=1}^{D} W_i \epsilon_{ni}\right)^2\right] = \sum_{i=1}^{D} W_i^2 \sigma^2$$

Again...

$$y_{\epsilon i} = W_0 + \sum_{i=1}^{N} W_i(x_i + \epsilon_i) = W_0 + \sum_{i=1}^{N} W_i x_{\epsilon i}$$

$$\tilde{E} = \frac{1}{2}\sum_{n=1}^{N}(y_{\epsilon i} - t_n)^2$$

$$= \frac{1}{2}\sum_{n=1}^{N}\{y_{\epsilon i}^2 - 2y_{\epsilon i}t_n + t_n^2\} = \frac{1}{2}\sum_{n=1}^{N}(\{y_N + \sum W_i x_\epsilon\}^2 - 2(y_N + \sum W_i x_\epsilon)t_n + t_n^2\}$$

$$= \frac{1}{2}\sum_{n=1}^{N}(y_N^2 + \cancel{2y_N\sum W_i x_{\epsilon i}} + (\sum W_i x_\epsilon)^2 - (2y_N - \cancel{2\sum W_i x_\epsilon})t_n + t_n^2)$$

$$= \frac{1}{2}\sum_{n=1}^{N}(y_N^2 + \underbrace{(\sum W_i x_\epsilon)^2}_{} - 2y_N t_n + t_n^2)$$

$$\frac{1}{2}\sum_{n=1}^{N}\left(\sum W_i x_\epsilon\right)^2 = E[(\sum W_i x_\epsilon)^2] = \sum_{i=1}^{D} W_i^2\sigma^2$$

$$= \frac{1}{2}\sum_{n=1}^{N}(y_N^2 - 2y_N t_n + t_n^2) + \sum_{i}^{N} W_i\sigma^2$$

$$\Rightarrow E[\tilde{E}_D] = E_D + \frac{1}{2}\sum_{i=1}^{D} W_i\sigma^2$$

$$\cancel{=\left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\alpha}{2\pi}\right)^{M/2}\int}$$

~~Back of Book~~

$$= \frac{\beta}{2}(t - w^T\phi(x))^2 + \frac{\alpha}{2}w^Tw = \frac{\beta}{2}(t_1^2 - 2w^T\phi(x)t) + (w^T\phi(x))^2\frac{\beta}{2} + \frac{\alpha}{2}w^Tw$$

$$= \frac{\beta}{2}(t^2 - 2w^T\phi(x)t) + w^TAw/2$$

$$= \frac{\beta}{2}(t^2 - 2w^T\phi(x)A^{-1}At) + w^TAw/2$$

$$= \frac{\beta}{2}(t^2) - 2m_nAw + w^TAw/2 + \frac{m_n^TAm_n}{2} - \frac{m_n^TAm_n}{2}$$

$$= \frac{\beta}{2}(t^2) - \frac{m_n^TAm_n}{2} + (w-m_n)^TA(w-m_n)$$

$$= \frac{\beta}{2}\left(t^2 - \frac{2m_n^TAm_n}{\beta} + \frac{m_n^TAm_n}{\beta}\right) + (w-m_n)^TA(w-m_n),$$

$$= \frac{\beta}{2}\left(t^2 - \frac{2m_n^TAm_n}{\beta} + \frac{m_n^T(\alpha I + \beta\phi^T\phi)m_n}{\beta}\right) + (w-m_n)^TA(w-m_n)$$

$$= \frac{\beta}{2}\left(t^2 - \frac{2m_n^TAm_n}{\beta} + \frac{m_n^T\alpha m_n}{\beta} + \frac{m_n^T\beta\phi^T\phi m_n}{\beta}\right) + (w-m_n)^TA(w-m_n)$$

$$= \frac{\beta}{2}(t - w\phi(t))^2 + \frac{\alpha}{2}m_n^Tm_n + (w-m_n)^TA(w-m_n)$$

3.17 $$p(t|\alpha,\beta) = \left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\alpha}{2\pi}\right)^{M/2}\int \exp\{-E(w)\}dw$$

$$= \int \underbrace{p(t|w,\beta)}_{n\times n}\underbrace{p(w|\alpha)}_{m\times m}dw = \left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\alpha}{2\pi}\right)^{M/2}\int \exp^{-\frac{\beta}{2}\sum(t-w\phi(x))^2} \cdot \exp^{-\frac{\alpha}{2}(w)^2} dw$$

$$= \left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\alpha}{2\pi}\right)^{M/2}\int \exp^{-\left(\frac{\beta}{2}\sum(t-w\phi(x))^2 + \frac{\alpha}{2}(w)^2\right)} dw$$

therefore $\boxed{E(w) = \frac{\beta}{2}\sum(t - w^T\phi(x))^2 + \frac{\alpha}{2}w^Tw}$

3.18, Complete the square for $E(w)$

$$\frac{\beta}{2}\left[\sum t_n^2 - 2\sum t_n w^T\phi(x) + \sum (w^T\phi(x))^T w^T\phi(x)\right] + \frac{\alpha}{2}[w^Tw]$$

$$= \frac{\beta}{2}\left[\sum t_n^2 - 2\sum t_n w^T\phi(x)\right] + \frac{w^T}{2}[\beta\phi^T(x)\phi(x) + \alpha]w = \frac{\beta}{2}\sum(t_n - w^T\phi(x))^2 - \frac{\beta(w^T\phi(x))^2}{4}$$

$$+ \frac{w^T}{2}[\beta\phi^T(x)\phi(x) + \alpha]w$$

$$\cancel{\frac{\beta}{2}\sum(t_n - w^T\phi(x))^2 - \frac{\beta(w^T\phi(x))^2}{2} + \frac{w^T}{2}(\beta\phi(x)^T\phi(x)+\alpha)w}$$

$$\frac{\beta}{2}\sum[t_n^2 - 2t_n w^T\phi(x)] + \frac{w^T}{2}(\beta\phi(x)^T\phi(x)+\alpha)w$$

$$\frac{\beta}{2}\sum[t_n^2 - 2t_n \underbrace{(\lambda I+\phi^T\phi)^{-1}}_{A=(\lambda I+\phi^T\phi)}\phi^T t\cdot\phi(x)] + \frac{w^T}{2}(\beta\phi(x)^T\phi(x)+\alpha)w$$

$$\frac{\beta}{2}\sum[t_n^2 - 2t_n A^{-1}\phi^T\cdot t_n\cdot\phi(x)] + \frac{w^T}{2}(\beta\phi(x)^T\phi(x)+\alpha)w$$

$$\frac{\beta}{2}\sum t_n^2 - \frac{\beta}{2}\sum t_n \underbrace{A^{-1}\phi^T\cdot t_n}_{m_n}\cdot\phi(x) + \frac{w^T}{2}(\beta\phi(x)^T\phi(x)+\alpha)w$$

$$\frac{\beta}{2}\sum t_n^2 - \beta\sum t_n m_n\cdot\phi(x) + \frac{w^T}{2}(\beta\phi(x)^T\phi(x)+\alpha)w$$

$$\frac{\beta}{2}\sum(t_n - m_n\phi(x))^2 - \frac{\beta m_n^T\phi(x)^T m_n\phi(x)}{2} + \frac{w^T}{2}(\beta\phi(x)^T\phi(x)+\alpha)w$$

$$\frac{\beta}{2}\sum(t_n - m_n\phi(x))^2 - \frac{\beta m_n^T\phi(x)^T m_n\phi(x)}{2} + \frac{((\lambda I+\phi^T\phi)\phi^T t)^T}{2}(\beta\phi(x)^T\phi(x)+\alpha)w$$

$$\frac{\beta}{2}\sum(t_n - m_n\phi(x))^2 - \frac{\beta m_n^T\phi(x)^T m_n\phi(x)}{2} + \frac{A\phi^T t}{2}(\beta\phi(x)^T\phi(x)+\alpha)w$$

Re-done $\frac{\beta}{2}\sum(t - w^T\phi(x))^2 + \frac{\alpha}{2}w^Tw = \frac{\beta}{2}\sum(t_n^2 - 2w^T\phi(x) + [w^T\phi(x)]^2) + \frac{\alpha}{2}w^Tw$

$$= \frac{\beta}{2}(t_n^2 - 2((\alpha I+\beta\phi^T\phi)^{-1}\phi^T t\,\phi(x)) + ((\alpha I+\beta\phi^T\phi)^{-1}\phi(x))^2) + \frac{\alpha}{2}[(\alpha I+\beta\phi^T\phi)^{-1}\phi^T t]^T[(\alpha I+\beta\phi^T\phi)^{-1}\phi^T t]$$

$$= \frac{\beta}{2}\sum(t_n^2 - 2A^{-1}\phi^T t\,\phi(x) + [A^{-1}\phi^T t\,\phi(x)]^2) + \frac{\alpha}{2}[A^{-1}\phi^T t]^T[A^{-1}\phi^T t]$$

$$= \frac{\beta}{2}\sum t_n^2 - \sum\beta A^{-1}\phi^T t\,\phi(x) + \sum\frac{\beta}{2}[A^{-1}\phi^T t\,\phi(x)]^2 + \frac{\alpha}{2}[A^{-1}\phi^T t][A^{-1}\phi^T t]$$

$$= \frac{\beta}{2}\sum t_n^2 - \sum m_n\phi(x) + \sum\frac{\beta}{2}[A^{-1}\phi^T t\,\phi(x)]^2 + \frac{\alpha}{2}\frac{m_n^T m_n}{\beta^2}$$

$$= \frac{\beta}{2}\sum t_n^2 + \sum\frac{m_n}{2}A^{-1}\phi^T\phi(x)t^2 - \sum m_n\phi(x) + \frac{\alpha}{2}\frac{m_n^T m_n}{\beta^2}$$

3.21 $\alpha = \dfrac{\text{const}}{m_n^T m_n}$ Alternative, $\dfrac{d}{d\alpha}\ln|A| = \mathrm{Tr}\left(A^{-1}\dfrac{d}{d\alpha}A\right)$

$$\frac{\partial}{\partial x}(A^{-1}A) = \frac{\partial}{\partial x}I = \ln A \cdot \frac{\partial}{\partial x}A \cdot A + A^{-1}\cdot\frac{\partial}{\partial x}A = 0$$

$$\ln A \frac{\partial}{\partial x} A \cancel{A A^{-1}} + A^{-1}\frac{\partial}{\partial x}\cancel{A A^{-1}} = 0$$

$$\ln A \frac{\partial}{\partial x} A + A^{-1}\frac{\partial}{\partial x} = 0$$

$$\ln A \frac{\partial}{\partial x} = -A^{-1}\frac{\partial}{\partial x}A^{-1}$$

$$\frac{\partial}{\partial A_{ij}}\mathrm{Tr}(AB) = \frac{\partial}{\partial A_{ij}}\left[A_{11}B_{11} + \cdots + A_{ij}B_{ji}\right] = B_{ji}$$

$$\frac{\partial}{\partial A}\mathrm{Tr}(AB) = B^T$$

$$\frac{\partial}{\partial A}\mathrm{Tr}(A^TB) = B \quad ; \quad \frac{\partial}{\partial A}\mathrm{Tr}(A) = I$$

$$\frac{\partial}{\partial A}\mathrm{Tr}(ABA^T) = A(B+B^T)$$

$$\frac{\partial}{\partial A}\ln A = (A^{-1})^T$$

$$= \mathrm{Tr}\left(A^{-1}\frac{\partial A}{\partial A}\right)$$

$$\frac{\partial}{\partial \alpha}\ln|A| = \mathrm{Tr}\left(A^{-1}\frac{\partial}{\partial \alpha}A\right)$$

$$\frac{\partial}{\partial \alpha}\ln p(t|\alpha,\beta) = \mathrm{Tr}\left(p(t|\alpha,\beta)^{-1}\frac{\partial}{\partial \alpha}p(t|\alpha,\beta)\right)$$

$$= \mathrm{Tr}\left(p(t|\alpha,\beta)^{-1}\cdot\left(\frac{\beta}{2\pi}\right)^{N/2}\cdot\frac{M}{4\pi}\left(\frac{\alpha}{2\pi}\right)^{-M/2}\cdot\frac{\partial}{\partial \alpha}e^{-\frac{\beta}{2}(t-\phi(\beta A^{-1}\phi^T t))^2 + \frac{\alpha}{2}(\beta A^{-1}\phi t)^2 - \frac{1}{2}(w-\beta A^{-1}\phi^T t)A(w-\beta A^{-1}\phi^T t)}\right)$$

$$\frac{\partial}{\partial \alpha}\ln|A| = \mathrm{Tr}\left(A^{-1}\frac{\partial}{\partial \alpha}A\right) = \mathrm{Tr}\left(\frac{I}{\alpha I + \beta\phi\phi}\right) =$$

$$= \mathrm{Tr}\left(\frac{\cancel{(\beta/2\pi)^{N/2}}\frac{M}{4\pi}\cancel{\left(\frac{\alpha}{2\pi}\right)^{M/2}}\cancel{e^{-\beta/2(t-\phi(\beta A^{-1}\phi^T t))^2 + \frac{\alpha}{2}(\beta A^{-1}\phi t)^2 - \frac{1}{2}(w-\beta A^{-1}\phi t)A(w-\beta A^{-1}\phi t)}}}{\cancel{(\beta/2\pi)^{N/2}}\cancel{\left(\frac{\alpha}{2\pi}\right)^{M/2}}\cancel{e^{-\beta/2(t-\phi(\beta A^{-1}\phi^T t))^2 + \frac{\alpha}{2}(\beta A^{-1}\phi t)^2 - \frac{1}{2}(w-\beta A^{-1}\phi t)A(w-\beta A^{-1}\phi t)}}}\right.$$

$$\cdot\frac{\partial}{\partial \alpha}\left(-\beta/2(t-\phi(\beta A^{-1}\phi^T t))^2 + \frac{\alpha}{2}(\beta A^{-1}\phi t)^2 - \frac{1}{2}(w-\beta A^{-1}\phi^T t)A(w-\beta A^{-1}\phi t)\right)$$

$$\Rightarrow \alpha = \frac{\gamma}{m_n^T m_n}$$

$$= \cancel{\frac{\beta}{2}\sum(t_n}$$

Again, $= \frac{\beta}{2}\sum(t - w^T\phi(w))^2 + \frac{\alpha}{2} w^T w$

$$w = (\alpha I + \phi^T\phi)^{-1}\phi t$$

$$A = \alpha I + \beta\phi^T\phi$$

$$m_N = \beta A^{-1}\phi^T t$$

$$= \frac{\beta}{2}\sum t_n^2 - \beta\sum w^T\phi(x) t_n + \frac{\beta}{2}\sum(w^T\phi(x))^2 + \frac{\alpha}{2}w^T w$$

$$= \frac{\alpha}{2}w^T w + \frac{\beta}{2}\sum[w^T\phi(x)]^2 - \beta\sum w^T\phi(x) t_n + \frac{\beta}{2}\sum t_n^2$$

$$\cancel{= \frac{\alpha}{2}w^T w + \frac{\beta}{2}\sum[w^T\phi(x)]^2 - \beta\sum(\alpha I + \beta\phi^T\phi)^{-1}\phi t\,\phi(x) t_n + \frac{\beta}{2}\sum t_n^2}$$

$$= \frac{\beta}{2}\sum t_n^2 - \beta\sum w^T\phi(x) t_n + \frac{w^T}{2}[\alpha + \beta\phi^T\phi]w$$

$$= \frac{\beta}{2}\sum(t_n - w^T\phi(x))^2 - \frac{\beta}{2}(w^T\phi(x))^2 + \frac{w^T}{2}[A]w = \frac{\beta}{2}\sum(t_n - w^T\phi(x))^2 - \frac{A}{2}(w - \beta A^{-1}\phi(x))^2 +$$

$$= \frac{\beta}{2}\sum(t_n - w^T\phi(x))^2 - \frac{A}{2}\left(w - \frac{m_n}{t}\right)^2 - \frac{m_n^T m_n}{t^2}A$$

$\boxed{\text{Unusual}}$

3.19.
$$\int \exp\{-E(w)\}dw = \exp\{-E(m_n)\}\int \exp\{-\tfrac{1}{2}(w-m_n)^T A(w-m_n)\}dw$$

$$z = w - m_n$$

$$= \exp\{-E(m_n)\}\int \exp\{-\tfrac{1}{2}z^T A z\}dz$$

$$= \exp\{-E(m_n)\}\sqrt[M]{\frac{2\pi}{A}}$$

$$\ln p(t|\alpha,\beta) = \ln\left(\frac{\beta}{2\pi}\right)^{N/2} + \ln\left(\frac{\alpha}{2\pi}\right)^{M/2} - E(m_n) - \frac{M}{2}\ln|A| + \frac{M}{2}\ln(2\pi)$$

$$\left(\frac{N}{2}\ln\beta + \frac{M}{2}\ln\alpha - E(m_n) - \frac{M}{2}\ln|A| - \frac{N}{2}\ln(2\pi)\right)$$

3.20 $\ln p(t|\alpha,\beta) = \frac{M}{2}\ln\alpha + \frac{N}{2}\ln\beta - E(m_n) - \frac{1}{2}\ln|A| - \frac{N}{2}\ln(2\pi)$

$$\frac{\partial}{\partial\alpha}\ln p(t|\alpha,\beta) = 0 = \frac{M}{2\alpha} - \frac{m_N^T m_n}{2} - \frac{1}{2}\sum\frac{1}{\lambda_i + \alpha_i} \Rightarrow \alpha m_N^T m_N = M - \alpha\sum\frac{1}{\lambda_i + \alpha} = \text{const}$$

$$\boxed{\alpha = \frac{\text{const}}{m_n^T m_n}}$$

$$p(t|\alpha,\beta) = \frac{M}{2}\ln\alpha + \frac{N}{2}\ln\beta - E(m_n) - \frac{1}{2}\ln|A| - \frac{N}{2}\ln(2\pi)$$

$$p(t|\alpha,\beta) = \frac{\partial}{\partial\beta}\left[\cancel{\frac{M}{2}\ln\alpha} + \frac{N}{2}\ln\beta - E(m_n) - \frac{1}{2}\ln|A| - \cancel{\frac{N}{2}\ln(2\pi)}\right]$$

$$= \frac{N}{2\beta} - \frac{\partial}{\partial\beta}\left[\frac{\beta}{2}(t - \phi\beta A^{-1}\phi^T t)^2\right] + \alpha\cdot\beta(A^{-1}\phi^T t)^2$$

$$0 = \frac{N}{2\beta} - \frac{1}{2}(t - \phi\beta A^{-1}\phi^T t)^2 - \frac{d}{d\beta}\frac{1}{2}\sum\ln|A| = \frac{N}{2\beta} - \frac{1}{2}(t - \phi\beta A^{-1}\phi^T t)$$

$$= \frac{N}{2\beta} - \frac{1}{2}(t - \phi\beta A^{-1}\phi^T t)^2$$

$$= \frac{N}{2\beta} - \frac{1}{2}(t - \phi\beta A^{-1}\phi^T t)^2$$

$$\frac{1}{N-\gamma}(t - \phi\beta A^{-1}\phi^T t)^2 = \frac{1}{\beta}$$

$$p(t) = \frac{1}{(2\pi)^{N/2}}\frac{b_0^{a_0}}{b_n^{a_n}}\frac{\Gamma(a_n)}{\Gamma(a_0)}\frac{|S_N|^{1/2}}{|S_0|^{1/2}} \leftarrow \text{Prove this}$$

$$p(w,\beta|t) = N(w|m_N, \beta^{-1}S_n)\,\mathrm{Gam}(\beta|a_n, b_n)$$

$$\frac{\partial p(w,\beta|t)}{\partial w} = \frac{\partial}{\partial\beta}\left[\frac{\beta^{N/2}}{(2\pi S_N)^{N/2}}\exp\left\{\frac{\beta}{2S_N}(w-m_N)^T(w-m_N)\right\}\frac{d}{dw}\frac{\beta}{2S_N}(w-m_N)^T(w-\right.$$

$$= \frac{\partial}{\partial\beta}\left[\frac{\beta^{N/2}}{(2\pi S_N)^{N/2}}\exp\left\{\frac{\beta}{2S_N}(w-m_N)^T(w-m_N)\right\}\frac{\beta}{S_N}(w-m_N)\,\mathrm{Gam}(\beta\right.$$

$$= \frac{\partial}{\partial\beta}\left[\frac{\beta^{N/2}}{(2\pi S_N)^{N/2}}\exp\left\{\frac{\beta}{2S_N}(w-m_N)^T(w-m_N)\right\}\frac{\beta}{S_N}(w-m_N)\frac{1}{\Gamma(a_n)}b_n^{a}\right.$$

$$p(t) = \iint p(w,\beta|t) = \iint N(w|m_n, \beta^{-1}S_n)\,\mathrm{Gam}(\beta|a_n, b_n)\,dw\,d\beta$$

$$= \iint\left(\frac{\beta}{2\pi}\right)^{N/2}\exp\left\{-\frac{\beta}{2}(t-\phi w)^T(t-\phi w)\right\}$$

$$\left(\frac{\beta}{2\pi}\right)^{M/2}|S_0|^{-1/2}\exp\left\{-\frac{\beta}{2}(w-m_0)^T S_0^{-1}(w\right.$$

$$\frac{b_0^{a_0}}{\Gamma(a_0)^{-1}}\beta^{a_0-1}\exp(-b_0\beta)\,d\beta$$

3.23 Cont.. $\frac{-1}{2S_N}(W - m_N)^T(W \cdot m_n) - \frac{1}{2S_N}\beta m_N^T m_n - \frac{\beta}{2}(t^2 + m_0 S_0^{-1})$

$\frac{-1}{2S_N}(W - m_N)^T(W - m_n) - \frac{\beta}{2}(t^2 + m_0^T S_0 m_0 + m_N^T S_N^{-1} m_N)$

$$= \frac{b_0^{a_n}}{\Gamma(a_n)} \iint \exp\left\{-\frac{\beta}{2}(W - m_N)^T S_N^{-1}(W - m_N) - \frac{\beta}{2}\left(t^2 + m_0^T S_0^{-1} m_0 + m_N^T S_N^{-1} m_N\right)\right\} dW \left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\beta}{2\pi}\right)^{M/2} \beta^{a_n - 1} e^{-b_n \beta} d\beta$$

$$= \frac{b^{a_N}}{\Gamma(a_N)} \iint \exp^{-\frac{\beta}{2} z^T S_N^{-1} z} dz \cdot \exp^{-\frac{\beta}{2}(t^2 + m_0^T S_0^{-1} m_0 + m_N^T S_N^{-1} m_N)} \left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\beta}{2\pi}\right)^{M/2} \beta^{a_0 - 1} e^{-b_0 \beta} d\beta$$

$$= \frac{b^{a_N}}{\Gamma(a_n)} \cancel{\sqrt[M]{\frac{2\pi A}{\beta}}} \int \exp^{-\frac{\beta}{2}(t^2 + m_0^T S_0^{-1} m_0 + m_N^T S_N^{-1} m_N)} \left(\frac{\beta}{2\pi}\right)^{N/2} \cancel{\left(\frac{\beta}{2\pi}\right)^{M/2}} \beta^{a_0 - 1} e^{-b_0 \beta} d\beta$$

$$= \frac{b^{a_N}}{\Gamma(a_n)} \int \exp^{-\beta\left(\frac{1}{2}[t^2 + m_0^T S_0^{-1} m_0 + m_N^T S_N^{-1} m_N] + b_N\right)} \cdot \beta^{a_N - 1} \cdot e^{-b_0 \beta} d\beta$$] Fuzz u2 terms

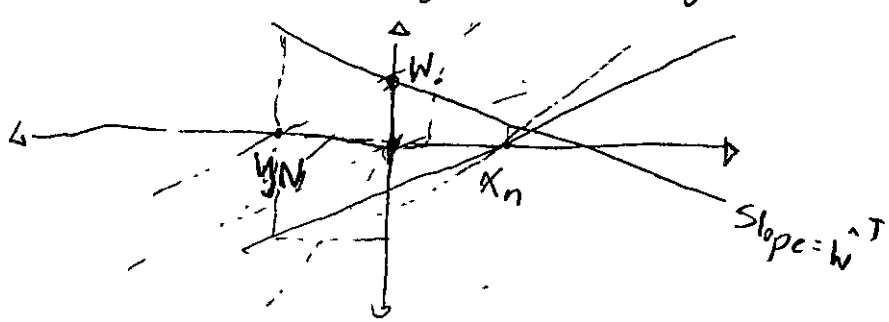
$$= \frac{b^{a_N}}{\Gamma(a_n)} \frac{1}{(2\pi)^{(M+N)/2} |S_0|^{1/2}} (2\pi)^{M/2} |S_N|^{1/2} \int \exp^{-\beta b_N} \beta^{a_n - 1} d\beta$$

$$= \frac{b_0^{a_0} (2\pi)^{M/2} |S_N|^{1/2}}{\Gamma(a_0)(2\pi)^{(M+N)/2} |S_0|^{1/2}} \frac{\Gamma(a_n)}{b_N^{a_N}}$$

Chapter 4.

4.1 $x_n = \sum_n \alpha_n x_n$; $\alpha \geq 0$, $\sum \alpha_n = 1$; y_n

iff $\hat{W}^T x_n + W_0 > 0 \ \forall x_n \wedge \hat{W}^T y_n + W_0 < 0 \ \forall y_n$



$$a^T t_n + b = 0$$

$$y(x) = T^T (\tilde{X}^\dagger) \tilde{x} = \sum_{i=1}^{n} t_n^T (\tilde{X}^\dagger) \tilde{x}$$

$$\frac{y(x)}{(\tilde{X}^\dagger)\tilde{x}} = \sum_{i=1}^{n} t_n^T \quad \dots \qquad a^T t_n + b = 0$$

$$a^T \frac{y(x)}{(\tilde{X}^\dagger)\tilde{x}} + b = 0$$

4.2 $E_D(\tilde{W}) = \frac{1}{2} \mathrm{Tr}\{(XW + 1w_0^T - T)^T (XW + 1w_0^T - T)\}$

$$\frac{\partial E_D(\tilde{W})}{\partial w_0} = (XW + 1w_0 - T) \cdot 1 = (XW - T)1 + \vec{1}^T \vec{1} w_0 = 0$$

$$w_0 = -(XW - T)\vec{1} / \vec{1}^T\vec{1}$$
$$= \tilde{T} - \tilde{X} W^T$$

Back Substituting: $E_D(\tilde{W}) = \frac{1}{2} \mathrm{Tr}\{(XW + 1(\tilde{T} - \tilde{X}W^T) - T)^T (XW + 1(\tilde{T} - \tilde{X}W)^T - T)\}$

$$= \frac{1}{2} \mathrm{Tr}\{(XW + 1 \cdot \tilde{T} - 1 \tilde{X} W^T - T)^T (XW + 1 \cdot \tilde{T} - 1 \cdot \tilde{X} W^T - T)\}$$

$$= \frac{1}{2} \mathrm{Tr}\{(XW + \bar{T} - \bar{X} W^T - T)^T (XW + \bar{T} - \bar{X} W^T - T)\}$$

From 4.

$$W = (X^T X)^{-1} X \hat{T} = \hat{X}^\dagger \hat{T}$$

$$y(x^*) = W^T x^* + w_0 = W^T x^* + \tilde{T} - x W^T$$
$$= W^T (x^* - x) + \tilde{T}$$
$$= (\tilde{X}^\dagger T)^T (x^* - x) + \tilde{T}$$

If $a^T t + b = 0$

$$a^T t = -b = a^T \frac{\vec{1}}{1} T^T$$

$$a^T y(x^*) = a^T(\tilde{T} + (X^\dagger T)^T (x^* - x)) = a^T \tilde{T} + a^T \underbrace{(X^\dagger \tilde{T})^T}_{(T - \tilde{T})^T = 0^T} (x^* - x) = -b = a^T t$$

$$S_W = \sum (x_n - m_1)(x_n - m_1)^T + \sum (x_n - m_2)(x_n - m_2)^T$$

$$S_B = (m_2 - m_1)(m_2 - m_1)^T$$

$$\sum (w^T x_n + w_0 - t_n) x_n = 0$$

$$w_0 = -w^T m \;;\; t_n = N_1 \frac{N}{N_1} - N_2 \frac{N}{N_2} \;;\; m = \frac{1}{N}(N_1 m_1 + N_2 m_2)$$

$$\sum \left(w^T x_n - w^T m - t_n \right) x_n = 0$$

$$\left(w^T (x_n - m) - t_n \right) x_n = 0$$

$$\left(w^T (x_n - m) - \left[N_1 \frac{N}{N_1} - N_2 \frac{N}{N_2} \right] \right) x_n = 0$$

$$\left(w^T \left(x_n - \frac{1}{N}(N_1 m_1 + N_2 m_2) \right) - \left[N_1 \frac{N}{N_1} - N_2 \frac{N}{N_2} \right] \right) x_n = 0$$

$$\left(w^T \left(x_n - \frac{1}{N}(N_1 m_1 + N_2 m_2) \right) \right) x_n = \left[N_1 \frac{N}{N_1} - N_2 \frac{N}{N_2} \right] x_n$$

$$w^T \left(x_n^T x - \frac{1}{N}(N_1 m_1 + N_2 m_2) x_n \right) = N \left[N_1 m_1 - N_2 m_2 \right]$$

$$w^T \left(x - \frac{(N_1 m_1 + N_2 m_2)}{2N} \right)^2 - w^T \frac{(N_1 m_1 + N_2 m_2)^2 x}{4N^2} = N \left[N_1 \cancel{m_1} - N_2 m_2 \right]$$

$$w^T \left(x - \frac{N_1 m_1}{2N} - \frac{N_2 m_2}{2N} \right.$$

$$\sum \left(w^T x_n - w^T m - N_1 \frac{N}{N_1} + N_2 \frac{N}{N_2} \right) x_n = 0$$

$$\left(w^T x_n^T x_n - \left(w^T m - N_1 \frac{N}{N_1} + N_2 \frac{N}{N_2} \right) x_n \right) = w^T \left(x - \frac{\left(w^T m - N_1 \frac{N}{N_1} + N_2 \frac{N}{N_2} \right)}{2 w^T} \right)^2 - \frac{\left(w^T m - N_1 \frac{N}{N_1} + N_2 \frac{N}{N_2} \right)^2 x}{2 w^T}$$

$$N \left(m_1 - \frac{N_2}{N_1} m_2 \right)$$

$$\left(S_W + \frac{N_1 N_2}{N} S_B \right) w = N (m_1 - m_2)$$

$$\left[(x_n - m_1)(x_n - m_1)^T + (x_n - m_2)(x_n - m_2) + \frac{N_1 N_2}{N} (m_2 - m_1)(m_2 - m_1) \right] w = N (m_1 - m_2)$$

$$= N \left(\frac{x}{N_1} - \frac{x}{N_2} \right)$$

$$\sum (w^T x + w_0 - t_n) x_n = 0$$

$$w_0 = -w^T m \;;\; t_n = N_1 \cdot \frac{N}{N_1} - N_2 \frac{N}{N_2} \;;\; m = \frac{1}{N}(N_1 m_1 + N_2 m_2)$$

$$S_B = (m_2 - m_1)(m_2 - m_1)^T \;;\; S_W = (x_n - m_1)(x_n - m_1)^T + (\cancel{x_0} - m_2)(x_n - m_2)^T$$

$$\sum (w^T x_n + w_0 - t_n) x_n = 0$$

$$\sum (w^T x_n - w^T m - t_n) x_n = w^T x_n x_n^T - w^T m x_n - t_n x_n = w^T (x_n x^T - m x_n) - t_n x_n$$

$$= w^T \left(x_n - \frac{m}{2}\right)^2 - w^T \left(\frac{m}{2}\right)^2 - t_n x_n = 0$$

$$= w^T \left(x_n - \frac{N_1 m_1}{2N} - \frac{N_2 m_2}{2N}\right)^2 - w^T \left(\frac{m}{2}\right)^2 - t_n x_n$$

$$= w^T \left(x_n - \frac{N_1 x_n}{2NN_1} - \frac{N_2 x_n}{2NN_1}\right)^2 - w^T \left(\frac{m}{2}\right)^2 - t_n x_n$$

Backwards:

$$\left(S_W + \frac{N_1 N_2}{N} S_B\right) w = N(m_2 - m_1) =$$

$$\left[(x_n - m_1)(x_n - m_1)^T + (x_n - m_2)(x_n - m_2)^T + \frac{N_1 N_2}{N}(m_2 - m_1)(m_2 - m_1)^T\right] w = N(m_2 - m_1)$$

$$\left[(x_n - m_1)(x_n - m_1)^T + (x - m_2)(x_n - m_2)^T + \frac{N_1 N_2}{N}\left[m_2^2 - 2m_1 m_2 + m_1^2\right]\right] w =$$

$$\cancel{\frac{N_1 N_2}{N}\left[\left(\frac{x_n}{N_2}\right)^2 - 2\frac{x_n^2}{N_1 N_2} + \left(\frac{x_n}{N_1}\right)^2\right]}$$

$$\cancel{\frac{x_n^2}{N}\left[\frac{N_1 \cdot N_2}{N_2 N_2} - \frac{2 N_1 N_2}{N_1 N_2} + \frac{N_1 N_2}{N_1 N_1}\right]}$$

$$\cancel{\frac{x_n^2}{N}\left[\frac{N_1^2}{N_1 N_2} + \frac{N_2^2}{N_1 N_2}\right]}$$

Could be $\frac{b^2}{2a}$

$$\frac{N_1 N_2}{N}\left[m_2^2 - 2m_1 m_2 + m_1^2\right]$$

$$-m_2^2 + m_2^2$$

Finish later

4.3 $a_m^T t_n + b = 0$ & $a_m^T y(x) + b = 0$

$A_{mn} = \begin{bmatrix} a_{11} & & \\ & \ddots & \\ & & a_{mn} \end{bmatrix}$; then $A_n^T t_n + b = 0$

$A_m^T y(x) + b = 0$

$$E_D(\tilde{W}) = \frac{1}{2} Tr\left[(XW - T)^T (XW - T)\right]$$

$$\simeq \frac{1}{2} Tr\left[(XW + \vec{1}w_0^T - T)^T (XW + \vec{1}w_0^T - T)\right]$$

$$\frac{dE_D(\tilde{W})}{dw_0} = \cancel{(XW + \vec{1}w_0^T - T)} \cdot \vec{1}$$

$$= (XW - T)\vec{1} + \vec{1}w_0 \Rightarrow -\vec{1}w_0 = (XW - T)\vec{1}$$

$$w_0 = -(XW - T)\frac{\vec{1}}{\vec{1}}$$

$$= (T - XW)\frac{\vec{1}}{\vec{1}}$$

$$E_D(\tilde{W}) = \frac{1}{2} Tr\left[(XW + (T - XW)\vec{1} - T)^T (XW + (T - XW)\vec{1} - T)\right]$$

If $W = (X^T X)^{-1} X^T T = X^{\dagger} T$

$$= (X - \bar{X})^{\dagger} (T - \bar{T})$$

$$y(x^*) = W^T x^* + w_0$$

$$= W^T x^* + (T - XW)\frac{\vec{1}}{\vec{1}}$$

$$= X^{\dagger} T x^* + T\frac{\vec{1}}{\vec{1}} - XW^T\frac{\vec{1}}{\vec{1}}$$

$$= -X^{\dagger} T (x^* - x) + T$$

$A^T t = -b$

$$A^T y(x^*) = A^T t + A^T \underbrace{X^{\dagger} T}_{\hat{T} = (T - T) = 0^T} (x^* - x) = -b \quad \therefore A^T t = -b$$

4.4 $m_k = w^T m_k$ $\quad w^T w = 1$ $\quad x^2 - \lambda x + \lambda$

Prove: $w \propto (m_2 - m_1)$ $\quad (x - \frac{\lambda}{2})^2$

$\mathcal{L}(w, \lambda) = w^T m_k + \lambda(1 - w^T w) \Rightarrow w^T m_k + \lambda - \lambda w^T w = -\lambda w^T w + w^T m_k + \lambda$

$= -\lambda\left(w^T - \frac{m_k}{2\lambda}\right)^2 + \lambda\left(\frac{m_k}{2\lambda}\right)^2 + \lambda$ $\quad = -\lambda\left(w^T - \frac{m_k}{2\lambda}\right)^2 - \lambda\left(\frac{m_k}{2\lambda}\right)^2 + \lambda$

$1 - w^T w = 0$

$2\lambda w^T + m_k = 0$

$\lambda = 0$

$w^T \propto \frac{m_k}{2\lambda} \propto (m_2 - m_1)$

$w^T m_k$ $\quad \frac{m_k^2}{1\lambda} + \frac{m_k^3}{4\lambda}$

$L(x, \lambda) = 1 - x_1^2 - x_2^2 + \lambda(x_1 + x_2 - 1)$

$= 1 - x_1^2 - x_2^2 + \lambda x_1 + \lambda x_2 - \lambda$

$= 1 - (x_1^2 - \lambda x_1) - (x_2^2 - \lambda x_2) - \lambda$

$= 1 - (x_1 - \frac{\lambda}{2})^2 - (\frac{\lambda}{2})^2 - (x_2 - \frac{\lambda}{2})^2 - (\frac{\lambda}{2})^2 - \lambda$

$-2x_1 + \lambda$ $\quad = 1 - (x_1 - \frac{\lambda}{2})^2 - (x_2 - \frac{\lambda}{2})^2 - 2(\frac{\lambda}{2})^2 - \lambda$

$-2x_2 + \lambda$

$x_1 + x_2 - 1$

4.5 #4.20 $y = w^T x$

#4.23 $m_k = w^T m_k$

#4.24 $s_k^2 = \sum_{n \in C_k} (y_n - m_k)^2$

Prove Fischer criterion $J(w) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2}$

Can be written as $J(w) = \frac{(w^T S_B w)}{(w^T S_W w)}$

Where $S_B = (m_2 - m_1)(m_2 - m_1)^T$

$S_W = \sum (x_n - m_1)(x_n - m_1)^T + \sum (x - m_2)(x_n - m_2)^T$

4.7. $\frac{1}{\sigma(y)} = \cancel{1+e^{-y} = 1 + e^{\ln\left(\frac{\sigma}{1-\sigma}\right)} = 1 + \sigma/1-\sigma = 1-\sigma} + \sigma = \frac{1}{1-\sigma}$

$= 1 + e^{-y} = \ln(e^{-y}) \frac{e^{-y}}{e^{-y}} + e^{-y} = \frac{\ln(e^{-y})e^{-y} + 1}{e^{-y}}$

$= \frac{\cancel{e^{-y} + \ln y \, e^{-y} + 1}}{e^{-y}} = e^{-y}\cancel{(1+\ln y) + 1}$

$\sigma(-a) = 1 - \sigma(a)$

$\frac{1}{\sigma(a)} = \frac{1+e^{-a}}{1} \cdots \frac{1+e^{a}}{e^{a}} \cdots = e^{a} - \sigma(-a) - a\sigma(-a) = e^{a}(1-\sigma(-a))$

$\sigma(a) = e^{a}(1-\sigma(-a))$

4.8. $p(C_1|x) = \frac{p(x|C_1)p(C_1)}{p(x|C_1)p(C_1) + p(x|C_2)p(C_2)}$

$= \frac{1}{1+\exp(-a)}$; $a = \ln \frac{p(x|C_1)p(C_1)}{p(x|C_2)p(C_2)}$

$\boxed{a = \ln \frac{\sigma(a)}{1-\sigma(a)}}$

Prove: $p(C_1|x) = \sigma(W^T x + w_0)$

Then $W^T = \Sigma^{-1}(\mu_1 - \mu_2)$; $w_0 = -\frac{1}{2}\mu_1^T \Sigma^{-1}\mu_1 + \frac{1}{2}\mu_2^T \Sigma^{-1}\mu_2 + \ln \frac{p(C_1)}{p(C_2)}$

$p(C_1|x) = \frac{1}{1+\exp(-a)}$; $a = \ln \frac{p(x|C_1)p(C_1)}{p(x|C_2)p(C_2)}$

$= \ln \frac{(x-\mu_1)^T \Sigma^{-1}(x-\mu_1) \cdot p(C_1)}{(x-\mu_2)\Sigma^{-1}(x-\mu_2) \cdot p(C_2)}$

$= \ln \frac{e^{(x-\mu_1)^T \Sigma^{-1}(x-\mu_2)}}{e^{(x-\mu_2)\Sigma^{-1}(x-\mu_2)}} + \ln \frac{p(C_1)}{p(C_2)}$

$$\frac{(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)}{(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)} + \ln\frac{p(C_1)}{p(C_2)}$$

$$p(C_1|x) = \frac{p(x|C_1)p(C_1)}{p(x|C_1)p(C_1)+p(x|C_2)p(C_2)} = \frac{1}{1+e^{\ln p(x|C_1)p(C_1)/[p(x|C_1)p(C_1)+p(x|C_2)p(C_2)]}}$$

$$a = \ln\frac{p(x|C_1)}{p(x|C_2)} + \ln\frac{p(C_1)}{p(C_2)}$$

$$= \frac{(x-\mu_1)\Sigma^{-1}(x-\mu_1)}{(x-\mu_2)\Sigma^{-1}(x-\mu_2)} + \ln\frac{p(C_1)}{p(C_2)} = \frac{-\frac{1}{2}\Sigma^{-1}(x^Tx-2\mu_1^Tx+\mu_1^T\mu_1)}{\Sigma^{-1}(x^Tx-2\mu_2x+\mu_2^T\mu_2)} + \ln\frac{p(C_1)}{p(C_2)}$$

$$= -\frac{1}{2}\Sigma^{-1}[(x^Tx-2\mu_1x+\mu_1^T\mu_1)+(x^Tx-2\mu_2x+\mu_2^T\mu_2)] + \ln\frac{p(C_1)}{p(C_2)}$$

$$= -\frac{1}{2}\Sigma^{-1}(-2\mu_1x+2\mu_2x+\mu_1^T\mu_1-\mu_2^T\mu_2) + \ln\frac{p(C_1)}{p(C_2)}$$

$$= -\frac{1}{2}\Sigma^{-1}(2\mu_1x-2\mu_2x-\mu_1^T\mu_1+\mu_2^T\mu_2) + \ln\frac{p(C_1)}{p(C_2)}$$

$$\boxed{= \underbrace{-\frac{1}{2}\Sigma^{-1}(\mu_1-\mu_2)}_{W}\underbrace{x}_{x} \underbrace{-\frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 - \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 + \ln\frac{p(C_1)}{p(C_2)}}_{W_0}}$$

4.9. $p(C_k) = \pi_k \quad p(\phi|C_k) \quad \vec{\phi} = \{\phi_1, \phi_2, \dots \phi_n\}$

Suppose $\{\phi_n, t_n\}$ $n = 1, \dots, N$

$t_{nj} = I_{jk}$ if pattern n of C_k

Prove maximum liklihood solution for prior probability is

$$\pi_k = \frac{N_k}{N}$$

posterior = liklihood · prior / Evidence

$$p(C_k|\phi) = \frac{p(\phi|C_k) \cdot p(C_k)}{\sum p(\phi_n|C_k) p(C_k)}$$

$$\frac{p(C_k|\phi) \cdot \sum p(\phi_n|C_k) p(C_k)}{p(\phi_n|C_k)} = p(C_k)$$

$$1 \cdot \frac{N}{N_k} \quad ? \qquad = \pi_k$$

Where is the Derivative?

OH! I haven't read this section

$$p(C_k|\phi) = \frac{p(\phi|C_k) p(C_k)}{\sum p(\phi_n|C_k) p(C_k)} = \frac{p(C_k) p(\phi|C_k)}{\sum p(C_k) p(\phi_n|C_k)} = \frac{\pi N(\phi_n|\mu_1, \Sigma)}{\pi N(\phi_n|\mu_1, \Sigma) + (1-\pi) N(\phi_n|\mu_2, \Sigma)}$$ Baysian - Continuous

$$p(C_k|\phi) = \prod_{n=1}^{N} [\pi N(\phi_n|\mu_1, \Sigma)]^{t_n} [(1-\pi) N(\phi_n|\mu_2, \Sigma)]^{1-t_n}$$

$$\frac{d \ln p(C_k|\phi)}{d\pi} = \sum_{n=1}^{N} t_n \ln \pi + (1-t_n) \ln(1-\pi) = 0$$

$$t_n (\ln \pi - \ln(1-\pi)) + \ln(1-\pi) = t_n \ln \pi/(1-\pi) + \ln(1-\pi) = 0$$ Generative Model - Nonce...

$1 - t_n \quad \ln(\pi) \quad = t_n$

t_n

$$\frac{d\ln p(C_k|\phi_n)}{d\pi} = \sum t_n \ln\pi + (1-t_n)\ln(1-\pi) = 0$$

$$\sum_i \ln\pi + \frac{\ln(1-\pi)}{t_n} - \ln(1-\pi) = 0$$

$$\sum \cancel{\ln\pi + \ln(1-\pi)} \left(\frac{1}{t_n} - 1\right) = 0$$

$$-\frac{\ln\pi}{\ln(1-\pi)} + \frac{1}{t_n} = \frac{1}{t_n} \qquad 1-t_n$$

$\boxed{\text{Fail}}$ Reason? No.

$$p(\{\phi_n, t_n\}|\{\pi_k\}) = \prod_{n=1}^{N}\prod_{k=1}^{K}\{p(\phi_n|C_k)\pi_k\}^{t_{nk}}$$

$$\ln p(\{\phi_n, t_n\}|\{\pi_k\}) = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}\{\ln p(\phi_n|C_k) + \ln\pi_k\}$$

$$\mathcal{L}(\ln p(\{\phi_n, t_n\}|\pi_k), \pi) = \frac{\partial}{\partial\pi}\left\{\ln p(\{\phi_n, t_n\}|\{\pi_k\}) + \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right)\right\}$$

$$= \frac{\sum t_{nk}}{\pi} + \lambda \; ; \; \pi\lambda = -\sum t_{nk} = N$$

$$= 0 \qquad ; \; \lambda = -N \quad \text{because } \sum \pi_k = 1$$

$$; \; \pi_k = \frac{N}{N_k}$$

$$= \iint \left(\frac{\beta}{2\pi}\right)^{N/2} \exp\left\{\frac{-\beta}{2}(t-\phi w)^T(t-\phi w)\right\} \left(\frac{\beta}{2\pi}\right)^{M/2} |S_0|^{-1/2} \exp\left\{\frac{-\beta}{2}(w-m_0)^T S_0^{-1}(w-m_0)\right\} dw$$

$$\cdot \frac{b_0^{a_n}}{\Gamma(a_n)} \beta^{a_n-1} e^{-b_n\beta} d\beta$$

$$= \frac{b_0^{a_n}}{\Gamma(a_n)} \iint \exp\left\{\frac{-\beta}{2}(t-\phi w)^T(t-\phi w) - \frac{\beta}{2}(w-m_0)^T S_0^{-1}(w-m_0)\right\} dw$$

$$\cdot \left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\beta}{2\pi}\right)^{M/2} \beta^{a_n-1} e^{-b_n\beta} d\beta$$

$$= \frac{b_0^{a_n}}{\Gamma(a_n)} \iint \exp\left\{\frac{-\beta}{2}\left[t^2 - 2\phi w t + (\phi w)^2\right] - \frac{\beta}{2}\left[w^T w - 2 w m_0 + m_0^2\right] S_0^{-1}\right\}$$

$$\left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\beta}{2\pi}\right)^{M/2} \beta^{a_n-1} e^{-b_n\beta} d\beta$$

Completing the square: $m_n = S_n\left[S_0^{-1} m_0 + \phi^T t\right]$ $\quad S_N^{-1} = \cancel{\beta}\left(S_0^{-1} + \phi^T\phi\right)$

$a_n = a_0 + \frac{N}{2}$ $\quad b_n = b_0 + \frac{1}{2}\left(m_0^T S_0^{-1} m_0 - m_N^T S_N^{-1} m_N + \sum_{i=1}^{N} t^2\right)$

$$\underbrace{\frac{-\beta}{2}\left[t^2\right]}_{t} + \underbrace{\frac{\beta}{2}\cancel{w(2\phi w t)} - \frac{\beta}{2}\cancel{(\phi w)^2} - \left[\frac{\beta}{2}\cancel{w^T w} - \frac{\beta}{2}\cancel{2 w m_0}\right.}_{w} - \left.\frac{\beta}{2} m_0^2\right] S_0^{-1}$$

$$\frac{-\beta}{2}(\phi w)^2 - \frac{\beta}{2} w^T w S_0^{-1} + \frac{\beta}{2} 2 w m_0 S_0^{-1} + \frac{\beta}{2} 2\phi w t - \frac{\beta}{2} m_0^2 S_0^{-1} - \frac{\beta}{2} t^2$$

$$-\frac{w^T}{2}\left[\beta S_0^{-1} + \beta\phi^T\phi\right] w^T + \frac{\beta}{2}\left[2 m_0 S_0^{-1} + 2\phi t\right] w - \frac{\beta}{2} m_0^2 S_0^{-1} - \frac{\beta}{2} t^2$$

$$\frac{-w^T S_N^{-1} w}{2} + \beta\left[m_0 S_0^{-1} + \phi t\right] - \frac{\beta}{2} m_0^2 S_0^{-1} - \frac{\beta}{2} t^2$$

$$\cancel{-\frac{S_N}{2}\left[w^T w +\right.}$$

$$\frac{-\beta}{2}\left(w^T\left[S_0^{-1} + (\phi^T\phi)\right] w - 2\left[m_0 S_0^{-1} + \phi t\right] w - \frac{\beta}{2}\left(t^2 + m_0^2 S_0^{-1}\right)\right.$$

Note Answer Key Correct $\longrightarrow$ $\frac{-\beta}{2}\left(w^T\left[\beta/S_n\right] w - 2\left[m_N/S_N\right]\right)$

$$\frac{-1}{2S_N}\left(w^T w - 2 m_N w\right) - \frac{\beta}{2}\left(t^2 + m_0^2 S_0^{-1}\right)$$

4.10 $p(\{\phi_n, t_n\} \mid \{\pi_k\}) = \prod_{n=1}^{N} \prod_{k=1}^{K} \{p(\phi_n \mid C_k)\pi_k\}^{t_{nk}}$

$\ln p(\{\phi, t_n\} \mid \{\pi_k\}) = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \{\ln p(\phi_n \mid C_k) + \ln \pi_k\}$

$$\ln p(\phi_n, t_n \mid \pi_k) + \lambda \sum_{k=1}^{K} \pi_k - 1$$

Now, $p(\{\phi, t_n\} \mid \pi_k) = N(\phi \mid \mu_k, \Sigma)$

Prove $\underbrace{\text{Maximum likelihood}}$ Solution is : $\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} t_{nk} \phi_n$

$$\frac{d \ln p(\{\phi, t_n\} \mid \{\pi_k\})}{d\mu_k} = \frac{d}{d\mu_k} \ln \prod_{n=1}^{N} \prod_{k=1}^{K} \left\{ e^{-(\phi_n - \mu_k)^T \Sigma^{-1} (\phi_n - \mu_k)/2} \right\}^{t_{nk}}$$

$$= \frac{d}{d\mu_k} - \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} (\phi_n - \mu_k)^T \Sigma^{-1} (\phi_n - \mu_k)/2$$

if Max : $0 = \sum_{n=1}^{N} t_{nk} (\phi - \mu_k) \cdot \phi_n = \sum_{n=1}^{N} t_{nk} \phi_n \cdot \phi_n - \sum_{n=1}^{N} \mu_k \phi_n$

$$\boxed{\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} t_{nk} \phi_n}$$

Prove Maximum likelihood Solution is : $\Sigma = \sum_{k=1}^{K} \frac{N_k}{N} S_k$

Where : $S_k = \frac{1}{N_k} \sum_{n=1}^{N} t_{nk} (\phi_n - \mu_k)(\phi_n - \mu_k)^T$

$$\therefore \frac{d \ln p(\{\phi, t_n\} \mid \{\pi_k\})}{d\Sigma} = \frac{d}{d\Sigma} \ln \prod_{n=1}^{N} \prod_{k=1}^{K} \left\{ e^{-(\phi_n - \mu_k)^T \Sigma^{-1} (\phi_n - \mu_k)/2} \right\}^{t_{nk}}$$

$$= \frac{d}{d\Sigma} - \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} (\phi_n - \mu_k)^T \Sigma^{-1} (\phi_n - \mu_k)/2$$

$$\boxed{N\Sigma = \sum_{n=K}^{K} t_k (\phi - \mu_k)(\phi - \mu_k)^T}$$

4.11 $\phi_k = \{\phi_M = (\phi_1, \phi_2, \phi_3, \cdots, \phi_M)\}$; $p(C_k|x) = \dfrac{p(x|C_k)p(C_k)}{\sum_j p(x|C_j)p(C_j)}$

$$= \frac{\exp(a_k)}{\sum_j \exp(a_j)}$$

$$a_k = \ln p(x|C_k)p(C_k)$$

$$p(\phi|\pi_k) = \prod_{k=1}^{K} p(\phi_k|\pi_k) = \prod_{L=1}^{L}\prod_{M=1}^{M} p(\phi_M = (\phi_1, \phi_2, \phi_3, \cdots, \phi_M)|\pi_M)$$

$$= p(\phi_1|\pi_1)p(\phi_2|\pi_1)\cdots$$

$$a_k = \ln p(\phi|\pi_1)p(C_k) = \ln p(\phi|\pi) + \ln p(C_k)$$

$$\boxed{= \underbrace{\frac{1}{\pi_k}}_{w} \cdot \phi + \underbrace{\ln p(C_k)}_{}}$$

$$= \cancel{W_k} \cdot \phi_k + W_k$$

"Linear"

4.12 $\dfrac{d\sigma}{da} = \sigma(1-\sigma)$; $\sigma(a) = \dfrac{1}{1+\exp(-a)}$

$$= \frac{d}{da}(1+e^{-a})^{-1} = -(1+e^{-a})^{-2}(-e^{-a}) = \boxed{\sigma(1-\sigma)}$$

$1 - \dfrac{1}{1+e^{-a}}$

4.13 $\dfrac{d\sigma}{da} = \sigma(1-\sigma)$ Prove $E(w) = -\ln p(t|w) = -\sum_{n=1}^{N} t_n \ln y_n + (1-t_n)\ln(1-t_n)$

$$\nabla E(w) = \sum_{n=1}^{N}(y_n - t_n)\phi_n$$

Where $y_n = \sigma(a_n)$ and $a_n = w^T\phi_n$

4.15 H for $\sigma(a)$ is positive. $R_{nn} = y_n(1-y_n)$ $y_n = w^T x$

$H = \nabla\nabla E(w) = \sum y_n(1-y_n)\phi_n\phi_n^T = \phi^T R \phi$.

$\nabla E(w) = \sum_n (y_n - t_n)\phi_n = \phi^T(y - T)$

Kullback-Liebler:

$$R_{KL}(y|t_n) = \int p(x) \ln \frac{P(x)}{Q(x)} dx$$

With property $0 < y_n < 1$, $0 < H < 1$, $\boxed{0 < \nabla\nabla E(w) < 1}$;

$\boxed{0 < \sum y_n(1-y_n)\phi_n\phi_n^T < 1} \to 0 < \sum y_n(1-y_n) < 1$ ϕ^T

$\boxed{0 < \phi^T R \phi < 1}$

4.16. $p(x|t_i) = p(x|0)$ $\quad x_1, x_2, x_3, \dots x_n$

$p(x|t_i) = p(x|1)$ $\quad \downarrow\pi_1 \ \downarrow\pi_2 \ \downarrow\pi_3 \ \dots \ \downarrow\pi_n$ $\quad \phi = (x_n|t_n)$

$\quad t_{1/0} \ t_{1/0} \ t_{1/0} \ \dots \ t_{1/0}$

$$p(t|\phi) = p(\phi|t)p(\phi)$$

$$\ln p(t|\phi) = \ln p(\phi|t) + \ln p(\phi)$$

4.17 $p(C_k|\phi) = y_k(\phi) = \dfrac{\exp(a_k)}{\sum_j \exp(a_j)}$ $\quad a_k = W_k^T\phi$

Prove $\dfrac{\partial p(C_k|\phi)}{\partial a_k} = \dfrac{\partial y_k(\phi)}{\partial a_k} = \dfrac{\partial}{\partial a_k}\left[\dfrac{\exp(a_k)}{\sum_j \exp(a_j)}\right] = \dfrac{\exp(a_k)\cdot\sum \exp(a_j) - \exp(a_k)^2}{(\sum_j \exp(a_j))^2}$

$$= \frac{\exp(a_k)}{\sum_j \exp(a_j)} - \frac{\exp(a_k)^2}{\sum_j \exp(a_j)^2}$$

$$\boxed{= y(1-y)}$$

$\dfrac{\partial p(C_k|\phi)}{\partial a_j} = \dfrac{\partial y_k(\phi)}{\partial a_j} = \dfrac{\partial}{\partial a_j}\left[\dfrac{\exp(a_k)}{\sum_j \exp(a_j)}\right] = \boxed{-y^2}$

4.18 $\nabla E(w) = \sum_{n=1}^{N} (y_n - t_n)\phi_n$

$$E(w_1, \cdots, w_K) = -\ln p(T|w_1, \cdots, w_K) = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \ln y_{nk}$$

$$\nabla E(w_1, \cdots, w_K) = \sum_{n=1}^{N} (y_{nj} - t_{nj})\phi_n$$

$$\nabla E(w_1, \cdots, w_K) = \nabla y_k(\phi) = -\nabla \ln p(T|w_1, \cdots, w_K) = -\nabla \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \ln y_{nk}$$

$$= -\frac{\partial}{\partial w_j} \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \ln y_{nk} = -\sum_{n=1}^{N}\sum_{k=1}^{K} \frac{t_{nk}}{\cancel{y_{nk}}} \cancel{y_k} (I_{kj} - y_j) \cdot \phi_n$$

$$\boxed{= \sum_{n=1}^{N} (y_n - t_n) \cdot \phi_n}$$

4.19. $\phi(a) = \int_{-\infty}^{a} N(\theta|0,1)d\theta$ $\therefore$ $\nabla \ln \phi(a) = \nabla \ln \int_{-\infty}^{a} N(\theta|0,1)d\theta$

$$\phi(a) = \frac{1}{2}\left\{1 + \frac{1}{\sqrt{2}}\mathrm{erf}(a)\right\}, \quad \mathrm{erf}(a) = \frac{2}{\sqrt{\pi}}\int_0^a \exp(-\theta^2/2)d\theta$$

$$\frac{\partial y_n}{\partial a_n} = \frac{\partial \phi(a)}{\partial a_n} = \frac{\partial}{\partial a_n}\left\{\frac{1}{2}\left(1 + \frac{1}{\sqrt{2}}\mathrm{erf}(a)\right)\right\}$$

$$= \frac{\partial}{\partial a_n}\left\{\frac{1}{2}\left(1 + \frac{1}{\sqrt{2}}\frac{2}{\sqrt{\pi}}\int_0^{a_n} e^{-\theta^2/2} d\theta\right)\right\}$$

$$\cancel{= \frac{\partial}{\partial a_n}\left[\frac{1}{\sqrt{2}}\frac{2}{\sqrt{\pi}}\sum_0^{\infty}\right]}$$

$$\nabla E = \frac{\partial E}{\partial a_n} = \frac{\partial E}{\partial y}\frac{\partial y}{\partial a_n}\frac{\partial a_n}{\partial W} = \sum \frac{y_n - t_n}{y_n(1-y_n)} \frac{1}{\sqrt{2\pi}} e^{-a_n^2} \phi_n$$

$$\nabla\nabla E = \nabla\left[\sum \frac{y_n - t_n}{y_n(1-y_n)} \cdot \frac{1}{\sqrt{2\pi}} e^{-a_n^2} \phi_n\right]$$

$$= \frac{\partial}{\partial y_n}\left[\sum \frac{y_n - t_n}{y_n(1-y_n)} \cdot \frac{1}{\sqrt{2\pi}} e^{-a_n^2} \cdot \phi_n \nabla y_n\right] + \left[\frac{y_n - t_n}{y_n(1-t_n)} \frac{1}{\sqrt{2\pi}} e^{-a_n^2} \cdot (-2a_n)\phi_n \nabla a_n\right]$$

$$\frac{y_n(1-y_n) + (y_n - t_n)(-2y_n)}{y_n(1-y_n)^2} \cdot 1$$

$$E(w) = -\ln p(t|W) = -\sum_{n=1}^{N} t_n \ln y_n + (1-t_n)\ln(1-t_n)$$

$$\nabla E(w) = \frac{\partial}{\partial w} E(w) = \frac{\partial}{\partial w} -\sum_{n=1}^{N} t_n \ln \sigma(w^T\phi_n) + (1-t_n)\ln(1-t_n)$$

$$= -\sum_{n=1}^{N} t_n \frac{1}{\sigma(w^T\phi_n)} \cdot \frac{\partial}{\partial w}\sigma \quad \boxed{= -\sum_{n=1}^{N} t_n \frac{1}{\sigma(w^T\phi_n)} \sigma(1-\sigma)\phi_n} \; t_n -$$

Where does t_n of first term eliminate?

4.14 $\frac{d \ln \sigma}{da} = 1-\sigma = 1 - \frac{1}{1+e^{-w^T\phi(x)}} \Rightarrow 1 = e^{-w^T\phi(x)}, \; \ln(1) = \boxed{-w^T\phi(x) = 0}$

4.15. Prove H for $\sigma(a)$ is $a^T y(x) + b = 0$ is positive

Here $R_{nn} = y_n(1-y_n)$, Hence show Error function is a concave fxn of W

4.20 $\nabla_{w_k}\nabla_{w_j} E(w_1, \dots, w_K) = -\sum_{n=1}^{N} y_{nk}(I_{kj} - y_{nj})\phi_n \phi_n^T$

Prove positive and semidefinite: H_{MK} ; $u^T H u$; u_{MK} ;

Assuming $0 < y_n < 1$: $0 < \nabla\nabla E(w_1, \dots, w_k) < 1$

$; 0 < -\sum_{n=1}^{N} y_{nk}(I_{kj} - y_{nj})\phi_n\phi_n^T < 1$

$; 0 < H < 1$

$0 < u^T H u < 1$ $\quad \boxed{0 < u^T H u < u^T u}$

Jensens Inequality: $\boxed{E[g(x)] \geq g(E[X])}$

4.21: $\phi(a) = \int_{-\infty}^{a} N(\theta|0,1)\,d\theta$; $\mathrm{erf}(a) = \frac{2}{\sqrt{\pi}}\int_0^a \exp(-\theta^2/2)\,d\theta$

$$\vec{\Phi}(a) = \frac{1}{2}\left\{1 + \frac{1}{\sqrt{2}}\mathrm{erf}(a)\right\}$$

$\cancel{\frac{d}{da}\vec{\Phi}(a) = \frac{d}{da}\frac{2}{\sqrt{\pi}}\int_0^a \exp(-\theta^2/2)\,d\theta} =$ $\quad \cancel{\mathrm{erf}(x) = 2\phi(a)+1}$

$$\boxed{\frac{1}{2}(\mathrm{erf}(x) + 1) = \phi(a)}$$

4.22 $Z = \int f(z)dz$

$$\simeq f(z_0)\int \exp\left\{-\frac{1}{2}(z-z_0)^T A(z-z_0)\right\}dz$$

$$= f(z_0)\frac{(2\pi)^{M/2}}{|A|^{1/2}}$$

Derive: $\ln p(D) \simeq \ln p(D|\theta_{MAP}) + \ln p(\theta_{MAP}) + \frac{M}{2}\ln(2\pi) - \frac{1}{2}\ln|A|$

$p(D) = Z \quad \therefore \ln p(D) = \ln \int f(z)dz = \ln f(z_0)\frac{(2\pi)^{M/2}}{|A|^{1/2}}$

$$= \ln p(D|\theta_{MAP})\cdot p(\theta_{MAP})\frac{(2\pi)^{M/2}}{|A|^{1/2}}$$

$$= \ln p(D|\theta_{MAP}) + \underbrace{\ln p(\theta_{MAP}) + \frac{M}{2}\ln(2\pi) - \frac{1}{2}\ln|A|}_{\text{Occam. Factor.}}$$

4.23. Prove $\ln p(D) \simeq \ln p(D|\theta_{MAP}) - \frac{1}{2}M\ln N$

from $\ln p(D) \simeq \ln p(D|\theta_{MAP}) + \ln p(\theta_{MAP}) + \frac{M}{2}\ln(2\pi) - \frac{1}{2}\ln|A|$

Show that if prior Gaussian $p(\theta) = N(\theta|m, V_0)$

the log model: takes the form:

$$\ln p(D) \simeq \ln p(D|\theta_{MAP}) - \frac{1}{2}(\theta_{MAP}-m)^T V_0^{-1}(\theta_{MAP}-m) - \frac{1}{2}\ln|H| + const$$

Where $H = -\nabla\nabla \ln p(D|\theta)$

Assume prior is broad, so V^{-1} is small & const is neglected.

Prove $\ln p(D) \simeq \ln p(D|\theta_{MAP}) - \frac{1}{2}M\ln N$

If $A = -\nabla\nabla \ln p(D|\theta_{MAP})p(\theta_{MAP}) = -\nabla\nabla \ln p(D|\theta_{MAP}) - \nabla\nabla \ln p(\theta_{MAP})$

$$= H - \nabla\nabla \ln p(\theta_{MAP})$$

$$= H - \nabla\nabla \ln N(\theta|m, V_0)$$

$$= H - \nabla\nabla(\theta-m)^T V_0^{-1}(\theta-m)$$

$$= H - \nabla(\theta-m)V_0^{-1} = H - V_0^{-1}$$

4.20. $p(\phi|C_k) = N(\phi|\mu_k, \Sigma)$

$$\sum_{n=1}^{N} t_n \ln p(\phi|C_k)$$

4.23 cont... $\ln p(D) \simeq \ln p(D|\theta_{MAP}) + \ln p(\theta_{MAP}) + \frac{M}{2}\ln(2\pi) - \frac{1}{2}\ln(\cancel{H+V_0}) + \text{const}$

$$\tilde{H} = \sum H_n = N\hat{H}$$

$$\hat{H} = \frac{1}{N}\sum H_N$$

$$\ln p(D) \cong \ln p(D|\theta_{MAP}) + \ln p(\theta_{MAP}) + \frac{M}{2}\ln(2\pi) - \frac{1}{2}\ln(M^{M}\hat{H})$$

$$\cong \ln p(D|\theta_{MAP}) + \ln p(\cancel{\theta_{MAP}}) + \cancel{\frac{M}{2}\ln(2\pi)} - \frac{M}{2}\ln N - \cancel{\frac{1}{2}\ln H}$$

$$\boxed{\cong \ln p(D|\theta_{MAP}) - \frac{M}{2}\ln N}$$

4.24 $p(C_1|t) = \int \sigma(a)p(a)\,da = \int \sigma(a) N(a|\mu_a, \sigma_a^2)\,da$

$p(a) = \int \delta(a - w^T\phi)\, q(w)\,dw$; $p(x_a) = \int p(x_a, x_b)\,dx_b$

if $a = w^T\phi$. $p(a) = \int \delta(a - w^T\phi)\, q(a)\,dw$

$$\delta(w^T\phi - w^T\phi) = 1 @ 0$$

$$= \int q(a)\,dw = \int N(a|\mu_a, \sigma_a^2)$$

Thus, $\boxed{\int \sigma(a) N(a|\mu_a, \sigma_a^2)\,da}$

4.25. $\sigma(a) = \frac{1}{1+\exp(-a)}$ scaled by $\phi(\lambda a)$, where $\phi(a) = \int_{-\infty}^{a} N(\theta|0,1)\,d\theta$

Prove if λ is chosen the derivatives of the two functions are equal at $a=0$, then $\lambda^2 = \pi/8$

$$\Phi(a) = \int_{-\infty}^{a} \sigma(\lambda a) N(\theta|0,1)\,d\theta = \int_{-\infty}^{a} \frac{1}{1+\exp(-\lambda a)} N(\theta|0,1)\,d\theta;$$

$$= \int_{-\infty}^{a} \frac{1}{1+\exp\left(-\sqrt{\frac{\pi}{8}}a\right)} N(\theta|0,1)\,d\theta =$$

4.25 $\phi'(\lambda a) = \dfrac{\lambda e^{-\lambda a}}{(1+e^{-\lambda a})^2} = \dfrac{\sqrt{\pi/8}\, e^{0}}{(1+1)^2} = \sqrt{\dfrac{\pi}{16 \cdot 8}}$

$\phi'(a) = \dfrac{d}{da} N(a|0,1) = \dfrac{-1}{\sqrt{2\pi}} \cdot e^{-\frac{a^2}{2}} = \dfrac{-1}{\sqrt{2\pi}}$ $\boxed{\text{Huh}}$

~~4.26~~ $\int \underbrace{\phi(\lambda a)}_{\text{a Probit}} \underbrace{N(a|\mu,\sigma^2)}_{\text{Gaussian}}\, da = \phi\left(\dfrac{\mu}{(\lambda^{-2}+\sigma^2)^{1/2}}\right)$

if $a = \mu + \sigma z$, $\int \phi(\lambda\mu + \lambda\sigma z) N(\mu+\sigma z|\mu,\sigma^2)\, d\mu$

$= \int \dfrac{1}{1+e^{\lambda(\mu+\sigma z)}} \cdot e^{\frac{-(\mu+\sigma z-\mu)^2}{2\sigma^2}}\, d\mu$

$= \int \dfrac{1}{1+e^{\lambda(\mu+\sigma z)}}\, e^{\frac{-\mu^2}{2\sigma^2}}\, d\mu$; $uv - \int v\,du$; $u = \dfrac{1}{1+e^{\lambda(\mu+\sigma z)}}$, $du = \dfrac{\lambda e^{-\lambda(\mu+\sigma z)}}{(1+e^{-\lambda(\mu+\sigma z)})^2}$

$\dfrac{d}{d\mu}\left[\dfrac{1}{1+e^{-\lambda(\mu+\sigma z)}}\, e^{\frac{-(\sigma z)^2}{2\sigma^2}}\right] = \dfrac{\lambda e^{-\lambda(\mu+\sigma z)}}{(1+e^{-\lambda(\mu+\sigma z)})^2}\, e^{\frac{-z^2}{2}}$

$= \dfrac{\lambda e^{-\lambda(\mu+\sigma z)} \cdot e^{-z^2/2\sigma}}{(1+e^{-\lambda(\mu+\sigma z)})^2}$

Diff term - ...

$\dfrac{d}{d\mu}\phi\left(\dfrac{\mu}{(\lambda^{-2}+\sigma^2)^{1/2}}\right) = \dfrac{d}{d\mu} \dfrac{1}{1+e^{(\mu/\frac{1}{\lambda^2}+\sigma^2)^{1/2}}}$

$= \dfrac{\left(\dfrac{1}{\frac{1}{\lambda^2}+\sigma^2}\right)^{1/2} e^{-\mu/(\frac{1}{\lambda^2}+\sigma^2)}}{(1+e^{-\mu/(\frac{1}{\lambda^2}+\sigma^2)})^2} =$

4.25 $\phi(\lambda a) = \frac{1}{1+e^{-\lambda a}}$; $\frac{d}{da}\left[\phi(\lambda a)\right] = \frac{d}{da}\left[\frac{1}{1+e^{-\lambda a}}\right] = \frac{\lambda e^{-\lambda a}}{(1+e^{-\lambda a})^2}$

$$\int_{-\infty}^{a} N(a|0,1) = \int_{-\infty}^{a} \frac{e^{-a^2}\, da}{\pi};$$

Derivative of Error Function

$$e^{z} = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \frac{z^3}{3!} \text{ where } z = -\frac{u^2}{2}$$

$$= 1 - \frac{u^2}{2 \cdot 1!} + \frac{u^4}{2^2 2!} - \frac{u^6}{2^3 3!} = \sum_{k=0}^{\infty} \frac{u^{2k}}{2^k \cdot k!}(-1)^k \Rightarrow \int_{-\infty}^{a} \sum_{k=0}^{\infty} \frac{u^{2k}}{2^k k!}(-1)^k du = \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k k!} \int_{-\infty}^{a} u^{2k} du$$

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{2^k k!} \left[\int_{-\infty}^{0} u^{2k} du + \int_{0}^{\lambda a} u^{2k} du \right]$$

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{2k \cdot k!} \left[+1/2 + \frac{\lambda a^{2k+1}}{2k+1} \right]$$

$$\frac{d}{da} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k \cdot k!} \left[+1/2 + \frac{a^{2k+1}}{2k+1} \right] = \sum_{k=1}^{\infty} \frac{(-1)^k}{2k \cdot k!} \lambda a^{2k}$$

@ $a = 0$; $\sum_{k=1}^{\infty} \frac{(-1)^k}{2k \cdot k!}$ $\quad 0 \quad 1/2\pi$.

$\lambda = \sqrt{\frac{\pi}{8}}$; $\phi'(\lambda a) = \frac{\sqrt{\pi/8}}{4}$,

$a = 0$;

$\boxed{\text{Wrong } \phi(\lambda a)?}$ $\boxed{\text{Possible completion}}$

5.1 $\sigma(a) = \{1 + \exp(-a)\}^{-1}$

$\tanh = \dfrac{e^{+a} - e^{-a}}{e^{a} + e^{-a}}$ $\qquad \sigma(a) = \dfrac{1}{1+e^{-a}} = \dfrac{e^{x}}{e^{x}+1}$

$$= \frac{e^{2x} - 1}{e^{2x}+1} = \frac{1-e^{-2x}}{1+e^{-2x}} = \frac{e^{2x}}{e^{2x}+1} - \frac{1}{e^{2x}+1} = \frac{e^{x} \cdot e^{x}}{e^{x} \cdot e^{x} + \frac{e^{x}}{e^{x}}} - \frac{1}{e^{x} \cdot e^{x} + \frac{e^{x}}{e^{x}}}$$

$$= \frac{e^{x}}{e^{x} + \frac{1}{e^{x}}} \cdot e^{x}$$

$$= \frac{e^{2x}}{e^{2x}+1} - \frac{1}{e^{2x}+1} = \boxed{\sigma(2x) - \frac{1}{e^{2x}+1}}$$

$$y_k(x,w) = \sigma\left(\sum_{j=1}^{M} w_{kj}^{(2)} h\left(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right)$$

Thus, $y_k(x,w) = \tanh\left(\dfrac{\left(\sum_{j=1}^{M} w_{kj}^{(2)} h\left(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right)}{2}\right) + \dfrac{1}{e^{\left(\sum_{j=1}^{M} w_{kj}^{(2)} h\left(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right)/2} + 1}$

5.2 $p(t|x,w) = N(t|y(x,w), \beta^{-1} I)$; $E(w) = \frac{1}{2}\sum_{n=1}^{N} \|y(x_n,w) - t_n\|^2$

Prove $\left(\dfrac{d\,p(t|x,w)}{dt} \text{ @ } t = 0\right) = \left(\dfrac{dE(w)}{dt_n} \text{ @ } t=0\right)$

$$\frac{d}{dt}\left(e^{-(t-y(x,w))^2 \beta/2}\right) = -\frac{\beta}{2} e^{-(t-y(x,w))^2\beta/2} = 0$$

$$\frac{dE(w)}{dt} = \sum_{n=1}^{N} \|y(x_n,w) - t_n\|^2 = 0$$

$$-(t - y(x,w))^2 = 0$$

$$\boxed{t_n = \frac{-y(x,w) \pm \sqrt{y(x,w)^2 - 4(1)y(x,w)^2}}{2a}}$$

5.3. $p(t|x,w) = N(t|y(x,w), \Sigma)$

$E(w) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n,w) - t_n\}^2$ $\qquad \Sigma_{ML} = \dfrac{1}{NK}\sum_{n=1}^{N}\|y(x_n,w_{ML}) - t_n\|^2$ independent

$$\Sigma_{ML} = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}\,(y(x_n,w_{ML}) - t_n)(y(x,w_{ML}) - t_n)^T$$

Dependent on size of W

5.1 $y_k(x,w) = \sigma\left(\sum_{j=1}^{m} w_{kj}^{(2)} h\left(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right)$; where $g(\cdot) = \sigma(a) = (1+\exp(-a))^{-1}$

Prove $\tanh(a) = \dfrac{e^a - e^{-a}}{e^a + e^{-a}}$;

5.4. $t \in \{0,1\}$; $y(x,w)$; $p(t=1|x)$; Binary Distribution ; $p(t|x) = y(x,w)^t (1-y(x,w))^{1-t}$

$$E(w) = -\ln p(t|x) = -\sum_{n=1}^{N} t_n \ln y_n + (1-t_n)\ln(1-y_n)$$

$$E(w) = -\ln p(t+\epsilon|x) = -\sum_{n=1}^{N} (t_n+\epsilon)\ln y_n + (1-t_n)\ln(1-y_n)$$

5.5 $t_k \in \{0,1\}$ $y(x,w) = p(t_k=1|x)$

$$= \prod_{k=1}^{K} y_k(x,w)^{t_k}[1-y_k(x,w)]^{1-t_k} ..$$

Prove $E(w) = -\sum_{n=1}^{N}\sum_{k=1}^{K}\{t_{nk}\ln y_{nk} + (1-t_{nk})\ln(1-y_{nk})\}$

$\hookrightarrow \dfrac{dE(w)}{dw} = \dfrac{d}{dw}\left[-\sum_{n=1}^{N}\sum_{k=1}^{K}\{t_{nk}\ln y_{nk} + (1-t_{nk})\ln(1-y_{nk})\}\right]$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K}\left[t_n / y_{nk} + \frac{1-t_{nk}}{1-y_{nk}}\right]\cancel{y_{nk}'} = 0$$

$$\frac{1-y_n}{y_n} = \frac{1-t_n}{t_n}$$

$\boxed{\dfrac{dE(w)}{dw} = y_n - t_n}$ should be equal to $\dfrac{dp(t_k=1|x)}{dt_k} = 0$

$$\frac{dp(t_k=1|x)}{dw} = \frac{d}{dw}\left[\prod_{k=1}^{K} y_k(x,w)^{t_k}\cancel{[1-y_k(x,w)]}^{1-t_k}\right] \checkmark$$

$$= \cancel{t_k y_k(x,w)^{t_k-1}[1-y_k(x,w)]^{1-t_k}} + \cancel{y_k(x,w)^{t_k}(1-t_k)[1-y(x,w)]^{-t_k}}$$

$= 0$

$\boxed{= y_k^{t_k} + t_k\ ?}$ Close - move on.

5.6: $E(w) = -\sum_{n=1}^{N} t_n \ln y_n + (1-t_n)\ln(1-y_n)$; $\frac{dE}{da_k} = \frac{d}{da_k}\left[-\sum_{n=1}^{N} t_n \ln y_n + (1-t_n)\ln(1-y_n)\right]$

$$= \left[-\sum \frac{t_n}{y_n} - \frac{(1-t_n)}{1-y_n}\right]\frac{dy}{da_k}$$

$$= \cancel{\frac{1-t_n}{1-y_n} - \frac{t_n}{y_n} = \frac{(1-t_n)y_n}{(y_n - y_n)^2}}$$

$$= \left[\frac{1-t_n}{1-y_n} - \frac{t_n}{y_n}\right] y_n(1-y_n) = (1-t_n)y - t_n(1-y_n)$$

$$= y_n - \cancel{t_n y} - t_n + \cancel{t_n y_n}$$

$$\boxed{= y_n - t_n}$$

5.7. $E(W) = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{kn} \ln y_{nk}(x_n, W)$; $\frac{dE}{da_k} = \frac{d}{da_k}\left[-\sum_{n=1}^{N}\sum_{k=1}^{K} t_{kn} \ln y_{nk}(x_n, W)\right] = -\frac{t_k}{y_k} \cdot y' = -\frac{t_k}{\cancel{y_k}} \cancel{y}(1-y_n)$

$$\boxed{= y_n - t_k}$$

5.8, $\frac{d\sigma}{da} = \sigma(1-\sigma)$; $\frac{d\tanh(a)}{da_k} = \frac{d}{a_k}\left[\frac{e^a - e^{-a}}{e^a + e^{-a}}\right] = \frac{(e^a + e^{-a})(e^a + e^{-a}) + (e^a - e^{-a})(e^a - e^{-a})}{(e^a + e^{-a})^2}$

$$\boxed{= 1 - \tanh^2(a)}$$

5.9 $E(w) = -\sum_{n=1}^{N}\{t_n \ln y_n + (1-t_n)\ln(1-y_n)\}$; $0 \leq y(x,w) \leq 1$; $t \in \{0,1\}$

Prove $E(w)$ for $-1 \leq y(x,w) \leq 1$ with $t=1$ for C_1 and $t=-1$ for C_2

~~$E(w) = -\sum_{n=1}^{N}(t_n - 1)\ln y_n + (1-t_n)\ln(1-y_n)$~~

~~$E(w) = -\sum_{n=1}^{N}(t_n - 1)\ln(y_n - 1) + (1 - t_n$~~ → ~~$E(w) = \sum_{n=1}^{N}(t_n - 1)$~~

$$\boxed{= -\sum_{n=1}^{N}(1-t_n)\ln(y_n - 1) + (t_n - 1)\ln(1-y_n)}$$

$\boxed{\sigma(a) = 2\sigma(a) - 1}$ Correction $y(a) = 2\sigma(a) - 1$
$= \tanh(a/2)$

5.10 $\boxed{Hu_i = \lambda u_i}$; $v^T H v = \sum_i c_i^2 \lambda_i$; $v = \sum_i c_i u_i$

$$= \sum_i c_i u_i^T \cdot \lambda \cdot \sum c_i u_i$$

$$\boxed{= \sum_i c_i^2 \lambda_i}$$

5.13 Prove $E(w) = E(\hat{w}) + (w-\hat{w})^T b + \frac{1}{2}(w-\hat{w})^T H (w-\hat{w})$ is $W(W+3)/2 = \frac{W^2+3W}{2}$

m^2 $(W \times m)(m \times W)$ $(W \times m)$

$\frac{1}{2}\cancel{(n \times n)}(n \times p)(p \times o)$; $m = p$; $n = p$ $\frac{1}{2}(W \times m)(m \times m)(m \times W)$

$\frac{1}{2}\cancel{(n \times n)(n \times p)}(n \times m)$ $\boxed{\frac{1}{2}W^2}$ for H

$\frac{1}{2}(m \times \cancel{n)(n \times n)(n} \times m)$

$\frac{1}{2}(m \times \cancel{m)} = \frac{m^2}{2}$ $\cancel{*W}$

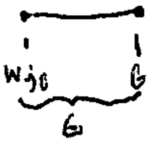
$\boxed{\text{Partial Solve}}$

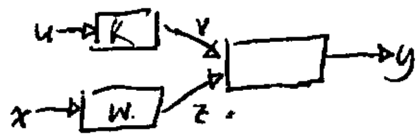
5.14 $\frac{\partial E_n}{\partial w_{ji}} = \frac{E_n(w_{ji}+\epsilon) - E_n(w_{ji}-\epsilon)}{2\epsilon} + O(\epsilon^2) = \frac{E_n(w_{ji}+\epsilon) - E_n(w_{ji}-\epsilon)}{w_{ji}+\epsilon - w_{ji}+\epsilon} + \cancel{O(\epsilon)} - \cancel{O(\epsilon)} + O(\epsilon)^2$

Taylor Expansion : $f(x) = \sum_{x=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$

$= f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots$

$\frac{\partial E_n}{\partial w_{ji}} = \frac{E_n(w_{ji}+\epsilon) - E_n(w_j)}{\epsilon} + O(\epsilon)$





5.15 $J_{ki} \equiv \frac{\partial y_k}{\partial x_i}$; $\frac{\partial E}{\partial w} = \sum_{n,j} \frac{\partial E}{\partial y_k}\frac{\partial y_k}{\partial z_j}\frac{\partial z_j}{\partial w}$; $\frac{\partial y_k}{\partial x_i} = \sum \frac{\partial y}{\partial a_j}\frac{\partial a_j}{\partial x_i} = \sum_j w_j \frac{\partial y_k}{\partial a_j}$

$\frac{\partial y_k}{\partial a_j} = \sum_l \frac{\partial y_k}{\partial a_l}\frac{\partial a_l}{\partial a_j} = h'(a_j)\sum_l w_{lj}\frac{\partial y_k}{\partial a_l} = h'(a_j)\sum w_{lj}\delta_{kj}\sigma'(a_j)$

$= h'(a_j)\sum_l w_{lj}\delta_{kj} y(1-y)_j$

Forward : $\frac{\partial y_k}{\partial x_i} = \frac{y_k(x_i+\epsilon) - y_k(x_i-\epsilon)}{2\epsilon} + O(\epsilon^2)$

$z = \sum_i w_i \cdot x_i$; $v = \sum_i k_i u_i$

$y(v, z) = \sum(w_i x_i + k_i u_i)$

5.16 $H \simeq \sum_{n=1}^{N} b_n b_n^T$; $b_n = \nabla y_n = \nabla a_n$

$E = \frac{1}{2}\sum_{n=1}^{N}(y_n - t_n)^2$; $y_n = \sum w_i x_i$

$E(w) = \sum_{k=1}^{K} E_k(w)$; $E_k(w) = \sum_{n=1}^{N}(y_{nk} - t_{nk})^2$; $H_k = \nabla\nabla E_k = \sum_{n=1}^{N} \nabla y_{nk} \nabla y_{nk} + \sum_{n=1}^{N}(y_{nk} - t_{nk})\nabla\nabla y_{nk}$

$$\boxed{H = \sum_{k=1}^{K} H_k = \sum_{k=1}^{K}\sum_{n=1}^{N} b_{nk}^T b_{nk}}$$

5.17. $E = \frac{1}{2}\iint \{y(x,w) - t\}^2 p(x,t)\,dx\,dt$

$y(x) = \frac{\int t p(x,t)\,dt}{p(x)} = \int t p(t|x)\,dt = E_t[t|x]$

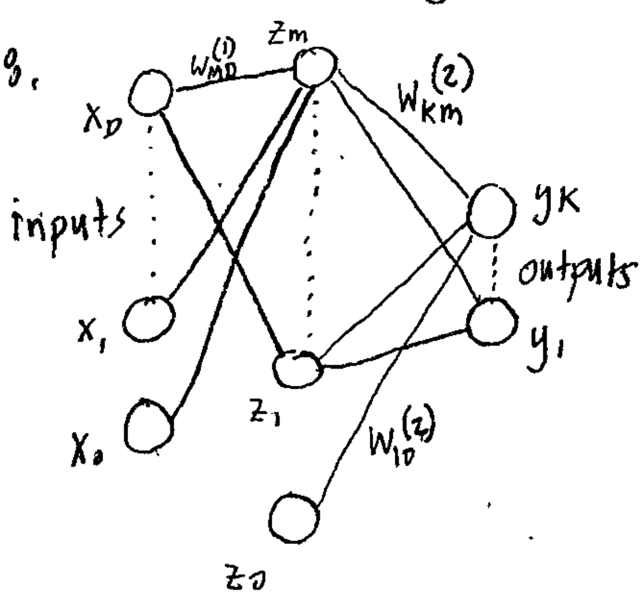
$\frac{\partial^2 E}{\partial w_r \partial w_s} = \frac{\partial}{\partial w_r}\frac{\partial}{\partial w_s} E = \frac{\partial}{\partial w_r}\frac{\partial}{\partial w_s}\frac{1}{2}\iint \{y(x,w) - t\}^2 p(x,t)\,dx\,dt$

$\frac{\partial}{\partial w_r}\frac{\partial}{\partial w_j} = \frac{1}{2}\iint \{y(x,w)^2 - 2y(x,w)t + t^2\} p(x,t)\,dx\,dt$

$\cancel{\frac{\partial}{\partial w_r}\frac{\partial}{\partial w_j} = \frac{1}{2}\left[\int [y(x,w)^2 - E[t|x,w]]^2 p(x)\,dx + \int \{E[t|x] - t\}^2 p(x)\,dx\right]}$

$\frac{\partial}{\partial w_r}\frac{\partial}{\partial w_j}\int y(x,w)^2 p(x)\,dx = \int \frac{\partial y}{\partial w_r}(x,w)\frac{\partial}{\partial w_j} y(x,w) p(x)\,dx$

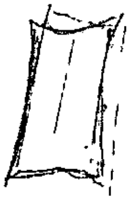
5.18.



+ Skip Layer

$a_j = \sum_D w_{MD}^{(i)} x_i$; $z_M = h(a_j)$; $y_K = \sum_M W_{KM} \cdot z_m$

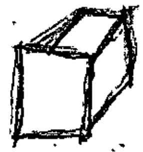
$$\boxed{\frac{\partial E}{\partial W_{km}} = \sum \left[\frac{\partial E}{\partial y}\frac{\partial y}{\partial z}\frac{\partial z}{\partial w} - \frac{\partial E}{\partial y}\frac{\partial y}{\partial a_j}\frac{\partial a_j}{\partial x}\right]}$$



$$= \frac{\partial}{\partial w_j}\left[(1-h(a_j)^2)\sum_{k=1}^{K} w_{kj}(y_k-t_k)x_i\right]$$

$$= -2h(a_j)h'(a_j)\sum_{k=1}^{K} w_{kj}(1-h(a_j)$$

5.23.



m V ↑↓

1:2

C – O

2:1

4:1

H

$\boxed{22.1L = 1mol}$

m V ↑↓ m V ↑↓

$H_2 + CO \rightarrow CO_2$

$PV = nRT$

P.V = + PV = nRT

$H_2 + O_2 \rightarrow 2H_2O$

$O_2 + C \rightarrow$

$CO_2 \rightarrow C + O_2$

1 : 1

$CH_4 \rightarrow CO_2$

$\% CH_4 + O_2 \rightarrow CO_2 + H_2$

$CH_4 + 3O_2 \rightarrow \cancel{2}CO_2\ 2H_2O$

$2CH_4 + \cancel{4O_2} \longrightarrow 2CO_2 + 4H_2O$			
%	%		

5.11 $E(W) = E(W^*) + \frac{1}{2}(W-W^*)^T H (W-W^*)$; $Hu_i = \lambda_i u_i$

$(W-W^*) = v = \sum_i c_i u_i$

$\to v^T H v = \sum_i c_i^2 \lambda_i$

$$\boxed{\sqrt{\frac{v^T H v}{\lambda}} = c_i}$$

5.12. $E(W) = E(W^*) + \frac{1}{2}(W-W^*)^T H (W-W^*)$ Prove sufficient condition when

$$(H)_{ij} \equiv \frac{\partial E}{\partial w_i \partial w_j}\Big|_{W^* = \hat{W}}$$

if $W = W^*$; then $E(W) = E(W^*)$

$\boxed{\frac{E(W)}{E(W^*)} > 95\%}$; sufficient

else if $(W \mathrel{!}= W^*)$; then $E(W) = E(W^*) + \frac{1}{2}(W-W^*)^T H (W-W^*)$

$$\boxed{\frac{E(W)}{E(W^*) + \frac{1}{2}(W-W^*)^T H (W-W^*)}}$$ possibly $< 95\%$

5.13. ~~$E(W) \simeq E(\hat{W}) + (W-\hat{W})^T b + \frac{1}{2}(W-\hat{W})^T H (W-\hat{W})$~~

5.19. $H \simeq \sum_{n=1}^{N} y_n(1-y_n) b_n b_n^T$; $\sigma(a) = \frac{1}{1+e^{-a}}$; $\nabla E(W) = \sum_{n=1}^{N} \frac{\partial E}{\partial a_n} \nabla a_n = \sum_{n=1}^{N} (y_n - t_n) \nabla a_n$

$H = \nabla\nabla E(W) = \cancel{\sum_{n=1}^{N} \nabla a_n \nabla a_n + \sum_{n=1}^{N} (y_n - t_n) \nabla\nabla a_n = \sum_{n=1}^{N} b_n b_n^T + \sum_{n=1}^{N} (y_n - t_n) \nabla\nabla a_n}$

$= \sum_{n=1}^{N} \frac{\partial y_n}{\partial a_n} \nabla a_n \nabla a_n + \sum_{n=1}^{N} \cancelto{0}{(y_n - t_n) \nabla\nabla a_n}$; where $y_n = \frac{1}{1+e^{-a}}$; $y'(a) = y(1-y)$

$= \sum_{n=1}^{N} y(1-y) \nabla a_n \nabla a_n = \sum_{n=1}^{N} y(1-y) b_n b_n^T$

5.20 $\frac{\exp(a)}{1 + \sum_j (\exp(a))}$ $H = \nabla\nabla E(W) = \sum \frac{\partial y_n}{\partial a_n} \nabla a_n \nabla a_n + \sum_{n=1}^{N} \cancelto{0}{(y_n - t_n) \nabla\nabla a_n}$

$$\boxed{= \sum \frac{\exp(a)^2 - \exp(a)(1 + \sum \exp(a))}{(1 + \sum \exp(a))^2} b_n b_n^T}$$

$$\mathrm{erf}(a) = \int_{-\infty}^{a} N(a|0,1)\,da = \int_{-\infty}^{a} 2e^{-a^2}/\sqrt{\pi}\,da = \frac{2}{\sqrt{\pi}}\left[\int_{-\infty}^{0} e^{-a^2}da + \int_{0}^{a} e^{-a^2}da\right] = \frac{2}{\sqrt{\pi}}\left[\tfrac{1}{2} + \int_0^a e^{-a^2}da\right]$$

$$e^{x} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

$$x = \frac{-a^2}{2};\ e^{-a^2} \cdot 1 - \frac{a^2}{2\cdot 1!} + \frac{a^4}{2\cdot 2!} - \frac{a^6}{2\cdot 3!} = \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k (k)!}$$

$$\mathrm{erf}(a) = \int_{-\infty}^{a} N(x|0,1)\,dx = \frac{2}{\sqrt{\pi}}\int_{-\infty}^{a} e^{-\frac{x^2}{2}}dx = \frac{2}{\sqrt{\pi}}\left[\int_{-\infty}^{0} e^{-\frac{x^2}{2}}dx + \int_0^a e^{-\frac{x^2}{2}}dx\right]$$

$$e^{u} = 1 + \frac{u}{1!} + \frac{u^2}{2!} + \frac{u^3}{3!} + \cdots;\quad u = -\frac{x^2}{2};\quad e^{-x^2/2} = 1 - \frac{x^2}{2\cdot 1!} + \frac{x^4}{2^2\cdot 2!} - \frac{x^6}{2^3\cdot 3!}$$

$$= \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k\cdot k!}\cdot x^{2k}$$

$$\frac{2}{\sqrt{\pi}}\left[\tfrac{1}{2} + \sum_{k=0}^{\infty}\frac{(-1)^k}{2^k k!}\int_0^a x^{2k}dx\right]$$

$$= \frac{2}{\sqrt{\pi}}\left[\tfrac{1}{2} + \right.$$

5.22 $E_n = \frac{1}{2}\sum_{n=1}^{N}(y_n - t_n)^2$; $y_n = \sum w_{kj} z_j$; $\delta_j = h'(a_j)\sum w_{kj}\delta_k$; $\delta_k = y_k - t_k$

① Both weights in second layer

$$\frac{\partial^2 E_n}{\partial w_{kj}^{(2)}\partial w_{k'j'}^{(2)}} = \frac{\partial E}{\partial w_{kj}^{(2)}}\frac{\partial E}{\partial w_{k'j'}^{(2)}} = \underbrace{\frac{\partial E}{\partial y_n}\frac{\partial y_n}{\partial w_{kj}}}_{z_{j'}}\frac{\partial w_{kj}}{\partial a_k}\cdot\underbrace{\frac{\partial E}{\partial y_n}\frac{\partial y_n}{\partial k_{j'}}}_{z_{j'}}\frac{\partial w_{kj}}{a_{k'}} = \boxed{z_j z_{j'} M_{kk'}}$$

② Both weights in first layer

$$\frac{\partial^2 E_n}{\partial w_{kj}^{(1)}\partial_{k'j'}^{(1)}} = \frac{\partial E}{\partial w_{kj}^{(1)}}\frac{\partial E}{\partial w_{k'j'}^{(2)}} = \cancel{\frac{\partial^2 E_n}{\partial w_{kj}^{(2)}}\frac{\partial a_j}{\partial a_j}\frac{\partial}{\partial w_j}} = \cancel{\frac{\partial}{\partial w_{kj}^{(2)}}\delta_j z_i} = \frac{\partial E_n}{\partial a_j}\frac{\partial a_j}{\partial w}\frac{\partial E_n}{\partial a_j}\frac{\partial a_j}{\partial w}$$

$$= \frac{\partial^2 E}{\partial a_j^2} z_j^2 = \frac{\partial^2}{\partial a_j^2} h^2(a_j) = \frac{\partial}{\partial a_j} h'(a)\sum_k w_{kj}\delta_k \cdot h^2(a_j)$$

$$= h''(a)\cdot\cancel{(z_i)\sum w_{kj}\delta_k\cdot h^2(a_j)} + h(a_j)\,h'(a_{j'})\sum_k w_{kj} w_k \boxed{\frac{\partial w_{ki}}{\partial m}}$$

$$\cancel{= h''(a)\,x_i x_i} =$$

$$= \cancel{\frac{\partial}{\partial w_{kj}}[\delta_j x_i]} = \frac{\partial}{\partial w_{jd}}\left[(1 - z_j^2)\sum_{k=1}^{K} w_{kj}\delta_k\, x_i\right]$$

Transformation: $y = \sum_i w_{kj} z_i + w_{k0}$

$y_k = \tilde{y}_k = C y_k + d$; $\boxed{\tilde{y}_k = C\left(\sum w_{kj} z_i + w_{k0}\right) + d}$

$w_{kj} = \tilde{w}_{kj} = C w_{kj}$; $\boxed{y_k = C \sum w_{kj} z_i + w_{k0}}$

$w_{k0} = \tilde{w}_{k0} = C w_{k0} + d$; $\boxed{y_k = \sum w_{kj} z_i + C w_{k0} + d}$

5.25 $E = E_0 + \frac{1}{2}(w-w^*)H(w-w^*)$

$v^T H v > 0$: Suppose $w^{(0)}$ is at origin and is updated by

$w^{(\tau)} = w^{(\tau-1)} - \rho \nabla E$

τ = step number

ρ = learning rate

(1) Prove τ steps, a $w \parallel \lambda_H = w_j^{(\tau)} = \{1-(1-\rho\eta_j)^\tau\} w_j^*$

where $w_j = w^T u_j$ & u_j and η_j are eigenvalues and eigenvectors

So that $H u_j = \eta_j u_j$

(2) Show $\tau \to \infty$, $w^{(\tau)} \to w^*$ provided $|1-\rho\eta_j| < 1$

(3) Now suppose training stops after τ steps.

Show $w \parallel \lambda_H = w_j^{(\tau)} \simeq w_j^*$ when $\eta_j \gg (\rho\tau)^{-1}$

$|w_j^{(\tau)}| \ll |w_j^*|$ when $\eta_j \ll (\rho\tau)^{-1}$

(1) ~~$w^{(\tau-1)} = 0$ @ $\langle 0,0 \rangle$ i.e. origin~~

$w^{(\tau)}$ ~~$= -\rho\nabla E = -\rho\frac{\partial E}{\partial w} = -\rho(w-w^*)^T H = -\rho(w-w^*)^T \eta_j$~~ ; $w^{(\tau)} = w^{(\tau-1)} - \rho(w-w^*)^T H = w^{(\tau-1)} - \rho\eta_i (w^{(\tau-1)}-w^*)^T$

(2) $\boxed{\lim_{\tau\to\infty} w_j^{(\tau)} = \lim_{\tau\to\infty} \{1-(1-\rho\eta_j)^\tau\} w_j^* = 1 \cdot w_j^* = w_j^*}$

4.26 $\int \phi(\lambda a) N(a|\mu,\sigma^2)\,da = \phi\left(\frac{1}{(\lambda^{-2}+\sigma^2)^{1/2}}\right)$; $a = \mu + \sigma z$

$$\frac{d}{d\mu}\left[\int \phi(\lambda a) N(a|\mu,\sigma^2)\,da\right] = \frac{d}{d\mu}\left[\int \phi(\lambda(\mu+\sigma z)) \cdot N(\mu+\sigma z|\mu,\sigma^2)\,d\mu\right]$$

$$= \frac{d}{d\mu}\left[\int_{-\infty}^{\infty} (1+e^{-\lambda(\mu+\sigma z)})^{-1} \cdot e^{-\frac{1}{2}(\mu+\sigma z-\mu)^2/\sigma}\,d\mu\right] = \frac{d}{d\mu}\left[\int_{-\infty}^{\infty} (1+e^{-\lambda(\mu+\sigma z)})^{-1} \cdot e^{-\frac{1}{2}(\sigma z)^2/\sigma}\,d\mu\right]$$

$$= \frac{d}{d\mu}\left[e^{-\frac{1}{2}\sigma z^2} \cdot \int_{-\infty}^{\infty} (1+e^{-\lambda(\mu+\sigma z)})^{-1}\,d\mu\right] = \frac{d}{d\mu}\left[e^{-\frac{1}{2}\sigma z^2} \cdot \frac{-1}{\lambda}\left(-\lambda(\mu+\sigma z) - \ln|1+e^{-\lambda(\mu+\sigma z)}|\right)\right]$$

$$= \left[-\frac{1}{\lambda}\left(-\lambda - \frac{-\lambda(e^{-\lambda(\mu+\sigma z)})}{1+e^{-\lambda(\mu+\sigma z)}}\right)\right]$$

$$= 1 + \frac{e^{-\lambda(\mu+\sigma z)}}{1+e^{-\lambda(\mu+\sigma z)}}$$

$$\frac{d}{d\mu}\phi\left(\frac{\mu}{(\lambda^{-2}+\sigma^2)^{1/2}}\right) = \frac{d}{d\mu}\left[\frac{\mu}{1+e^{\sqrt{\lambda^{-2}+\sigma^2}}}\right] \frac{d}{d\mu}\left[(1+e^{\sqrt{\lambda^2+\sigma^2}})^{-1}\right]$$ $\boxed{\text{unsure}}$

$\boxed{\text{Review or write in back of book}}$

~~5.20~~ 5.21 $H_N = \sum_{n=1}^{N} b_n b_n^T$; $b_n \equiv \nabla_w a_n$; Prove $H_{L+1} = H_L + b_{L+1} b_{L+1}^T$

$$H_L = \sum_{n=1}^{L} b_n b_n^T ; \boxed{H_{L+1} = \sum_{n=1}^{L+1} b_n b_n^T = \sum_{n=1}^{L} b_n b_n^T + b_{L+1} b_{L+1}^T}$$

$$(M+vv^T)^{-1} = M^{-1} - \frac{(M^{-1}v)(v^T M^{-1})}{1+v^T M^{-1} v} ; \boxed{(H_{L+1})^{-1} = (H_L + b_{L+1} b_{L+1}^T)^{-1} = H_L^{-1} - \frac{(H_L^{-1} b_{L+1})(b_{L+1}^T H_L^{-1})}{1 + b_{L+1}^T H_L^{-1} b_{L+1}}}$$

5.22. $\frac{\partial^2 E_n}{\partial w_{kj}^{(2)} \partial w_{k'j'}^{(2)}} = z_j z_{j'} M_{kk'}$; $\frac{\partial^2 E_n}{\partial w_{ji}^{(1)} \partial w_{j'i'}^{(1)}} = x_i x_{i'} h''(a_{j'}) I_{jj'} \sum_k w_{kj}^{(2)} \delta_k$

$$+ x_i x_{i'} h'(a_{j'}) h'(a_j) \sum_k \sum_{k'} w_{k'j'}^{(2)} w_{kj}^{(2)} M_{kk'}$$

$$\frac{\partial^2 E_n}{\partial w_{ji}^{(1)} \partial w_{kj'}^{(2)}} = x_i h'(a_{j'}) \left\{ \delta_k I_{jj'} + z_j \sum_{k'} w_{k'j'} H_{kk'} \right\}$$

5.22 cont... 5.94. Prove $\frac{\partial^2 E}{\partial w_{ji}^{(1)} \partial w_{j'i'}^{(1)}} = x_i x_{i'} h''(a_j) I_{jj'} \sum_K w_{kj'}^{(2)} \delta_k$

$+ x_i x_{i'} h'(a_{j'}) h'(a_j) \sum_K \sum_{K'} w_{K'j'}^{(2)} w_{KJ}^{(2)} M_{KK'}$

$a_j = \sum w_{ji} z_i \; ; \; z_j = h(a_j)$

$$\frac{\partial^2 E}{\partial w_{ji} \partial w_{ji}} = \frac{\partial}{\partial w_{ji}} \left[\frac{\partial E_n}{\partial a_j} \frac{\partial a_j}{\partial w_{ji}} \right] = \frac{\partial}{\partial w_j} \left[\delta_j z_i \right]$$

$$= \frac{\partial}{\partial w_j} \left[\sum_K \frac{\partial E_n}{\partial a_K} \frac{\partial a_K}{\partial a_j} \right] z_i$$

$$= \left[\sum_K \frac{\partial E_n}{\partial a_K} \frac{\partial a_K}{\partial a_j} \right]' z_i + \left[\sum_K \frac{\partial E_n}{\partial a_K} \frac{\partial a_K}{\partial a_j} \right] z_i'$$

$$= \left[\sum_K \delta_K \right.$$

$$\frac{\partial^2 E}{\partial w_{ji}^{(1)} \partial w_{j'i'}^{(1)}} = \frac{\partial^2 E}{\partial a_j \partial a_{j'}} \frac{\partial a_j}{\partial w_{ji}} \cdot \frac{\partial a_{j'}}{\partial w_{j'i'}} = \frac{\partial^2 E_n}{\partial a_j \partial a_{j'}} z_i^2$$

$$= \frac{\partial E}{\partial a_j} h'(a_j) \sum_K w_{kj} \delta_K z_i^2$$

$$\boxed{= h'(a_j)^2 \sum_{K'} \sum_K w_{kj'} w_{kj} z_i^2 + h''(a_j) \sum_{a_j} w_{kj} z_i^2 = \cdots}$$

③ One weight per layer

$$\frac{\partial^2 E}{\partial w_{ji}^{(1)} \partial w_{kj'}^{(1)}} = \frac{\partial^2 E}{\partial a_j \partial a_k} \frac{\partial a_j}{\partial w_{ji}^{(1)}} \frac{\partial a_k}{\partial w_{kj'}^{(1)}} = \boxed{h'(a_j) \left\{ \delta_K z_i + \sum w_{k'j'}^{(2)} H_{kk'} \right\}}$$

③ $\eta_j \gg (\rho\tau)^{-1}$ $\boxed{w^{(\tau)} = w^{(\tau-1)} - \rho\nabla E = \{1-(1-\rho\eta_j)^{\tau}\} w_j^* \simeq w_j^* + \underbrace{\epsilon}_{\text{small error}}}$

$\eta_j \ll (\rho\tau)^{-1}$ $\boxed{w^{(\tau)} = w^{(\tau-1)} - \rho\nabla E = \{1-(1-\rho\eta_j)^{\tau}\} w_j^* \simeq w_j^* + \infty}$

5.26. $\widehat{E} = E + \lambda\Omega$; $\Omega = \frac{1}{2}\sum_n\sum_k \left(\frac{\partial y_{nk}}{\partial \xi}\Big|_{\xi=0} \right)^2 = \frac{1}{2}\sum_n\sum_k \left(\sum_{i=1}^{D} J_{nki}\tau_{ni} \right)^2$

Prove $\Omega_n = \frac{1}{2}(G_{yk})^2$; $G_{yk} = \sum_i \tau \frac{\partial}{\partial x_i}$; $\Omega = \frac{1}{2}\left(\begin{bmatrix} \sum_i \tau_{ni}\frac{\partial}{\partial x_i} & & \\ & \ddots & \\ & & \end{bmatrix}_{n\times k} y_{nk} \right)^2$

$z_j = h(a_j)$; $a_j = \sum w_{ji} z_i$: Prove Ω_n evaluation by:

$\alpha_j = h'(a_j)\beta_j$; $\beta_j = \sum w_{ji}\alpha_i$

where $\alpha_j \equiv G_{z_j}$; $\beta_j \equiv G_{a_j}$

$\Omega_n = \frac{1}{2}\sum_i (G_n y_k)^2 = \frac{1}{2}\sum_i \left(G_n \left(\sum w_{ji} z_i + w_{j0} \right) \right)^2$

$= \frac{1}{2}\sum_i \left(\begin{bmatrix} \tau\frac{\partial}{\partial x_i} & & \\ & \ddots & \\ & & \end{bmatrix} \left(\sum w_{ji} z_i + w_{j0} \right) \right)^2$

$\boxed{= \left(\sum_i G_i y_{nk} \right) \cdot h'(a_j) \cdot \beta_j = \boxed{\sum (\tau_i G_n y_{nk}) \cdot h'(a_j) \cdot \sum w_j \alpha_i}}$

Prove $\frac{\partial \Omega_n}{\partial w_{rs}} = \sum_k \alpha_k \{ \phi_{kr} z_s + \delta_{kr} \cdot \alpha_s \}$; $\delta_{kr} = \frac{\partial y_k}{\partial a_r}$; $\phi_{kr} = G\delta_{kr}$

$= \left(\sum_k G_i y_{nk} \right) \left(G_{\partial_i} y_{nk} \right) = \sum_k G_i y_{nk} \cdot G \underbrace{\frac{\partial}{\partial a_r} y_k}_{\delta_k} = \sum G_i y_{nk} \phi_{kr}$

$= \sum G_i (w_j z_i + w_{j0}) \phi_{kr} = \sum G (w_j z_i \phi_{kr} + w_{j0}\phi_{kr})$

$= \sum G (w_j z_i \phi_{kr} + G\delta_k w_{j0}) = \sum w_j \alpha_k$

5.27 $x \to x+\xi$; $\xi = N(x|0,1)$; $\Omega = \frac{1}{2}\int \|\nabla y(x)\|^2 p(x)\,dx$

$$\tilde{E} = E + \lambda\Omega$$

$$= \frac{1}{2}\iiint \{y(s(x,\xi)) - t\}^2 p(t|x)\,p(x)\,p(\xi)\,dx\,dt\,d\xi$$

$$\frac{\partial y}{\partial \xi_i} = \sum \frac{\partial y}{\partial s}\frac{\partial s}{\partial \xi} = b_i$$

$$\cancel{\frac{\partial y}{\partial \xi_i \partial \xi_j} = \frac{\partial b_i}{\partial \xi_i} = \sum \frac{\partial b_i}{\partial s}\frac{\partial s}{\partial \xi_j} = B_{ij}}$$

$$\cancel{s(x,\xi) = s(x,0) + \xi\frac{\partial}{\partial \xi}s(x,\xi)\Big|_{\xi=0} + \frac{\xi^2}{2}\frac{\partial^2}{\partial \xi^2}s(x,\xi)\Big|_{\xi=0} + O(\xi^3)}$$

$$y(s(x,\xi)) = y(x) + \xi\,\nabla y(x)\frac{\partial x}{\partial \xi} + \frac{\xi^2}{2}\left[\frac{\partial^2 x}{\partial \xi^2}\nabla y(x) + \frac{\partial x}{\partial \xi}\nabla\nabla y(x)\frac{\partial x}{\partial \xi}\right] + \cdots$$

$$\tilde{E} = \cancel{\frac{1}{2}\iint\{y(x)-t\}^2 p(t|x)p(x)\,dx\,dt}$$

$$\cancel{+\iint} = \frac{1}{2}\iiint\{a+b+c-t\}^2 p(t|x)\,p(x)\,p(\xi)\,dx\,dt\,d\xi$$

$$= \frac{1}{2}\iint\cancel{\{a^2+ab+ac-at+ab+b^2+bc-bt+ca+bc+c^2-ct^2 - ta - tb - tc + t^2\}p(t|x)p(x)p(\xi)\,dx\,dt\,d\xi}$$

$$\cancel{a = y(x)\ ;\ b = \xi\,\nabla y(x)\frac{\partial x}{\partial \xi}\ ;\ c = \frac{\xi^2}{2}\left[\frac{\partial^2 x}{\partial \xi^2}\nabla y(x) + \frac{\partial x}{\partial \xi}\nabla\nabla y(x)\frac{\partial x}{\partial \xi}\right]}$$

$$\cancel{= \frac{1}{2}\iiint\{y(x)^2 + y(x}$$

$$\left(y(s(x,\xi)) - t\right)^2 = (y(x)-t)^2 + \left(\xi\nabla y(x)\frac{\partial x}{\partial \xi} - t\right)^2 + \left(\frac{\xi^2}{2}\left[\frac{\partial^2 x}{\partial \xi^2}\nabla y(x) + \frac{\partial x}{\partial \xi}\nabla\nabla y(x)\frac{\partial x}{\partial \xi}\right] - t\right)^2$$

$$\xi b^T = \xi\frac{\partial y}{\partial \xi}$$

$$\tilde{E} = \frac{1}{2}\iiint(y(x)-t)^2 p(t|x)p(x)p(\xi)\,dx\,dt\,d\xi$$

$$+ \iiint\{y(x)-t\}\,\xi\,b\;p(t|x)\,p(x)\,p(\xi)\,dx\,dt\,d\xi \qquad \frac{d\left(y(s(x,\xi)) - t\right)}{d\xi}$$

$$+ \frac{1}{2}\iiint \xi^T\left(\{y(x)-t\}B + bb^T\right)\xi\,p(\xi)\,p(t|x)\,p(x)\,d\xi\,dx\,dt$$

Which come from:

$$\tilde{E} = \frac{1}{2}\iiint(y(x)-t)^2 p(t|x)p(x)p(\xi)\,dx\,dt\,d\xi$$

$$+ \frac{1}{2}\iiint \xi\frac{\partial}{\partial x}(y(x)-t)^2\frac{\partial x}{\partial \xi}\,p(t|x)p(x)p(\xi)\,dx\,dt\,d\xi$$

$$+ \frac{1}{2}\iiint \frac{\xi^2}{2}\frac{\partial}{\partial x}\left(\left[\frac{\partial^2 x}{\partial^2 \xi}\nabla y(x) + \frac{\partial x}{\partial \xi}\nabla\nabla y(x)\frac{\partial x}{\partial \xi}\right] - t\right)^2 \xi\,p(\xi)\,p(t|x)\,p(x)\,d\xi\,dx\,dt$$

$$= \frac{1}{2}\iiint(y(x)-t)^2 p(t|x)p(x)p(\xi)\,dx\,dt\,d\xi$$

$$+ \iiint \xi\,(y(x)-t)\,\nabla y(x)\,b^T p(t|x)\,p(x)\,p(\xi)\,dx\,dt\,d\xi$$

$$+ \iiint \xi^2\,(y(x)-t)\left(\frac{\partial^2 x}{\partial^2 \xi} + \frac{\partial^2 y}{\partial \xi^2}\right)\xi\,p(\xi)\,p(t|x)\,p(x)\,d\xi\,dx\,dt$$

$$(a+b-t)(a'b+ab')$$

5.28 $a < \pi_i < b$; $E = \frac{1}{2}\iint \{y(x) - t\}^2 p(t|x)p(x)\,dx\,dt$

Backpropagation Algorithm: $\frac{\partial E}{\partial w_i} = \sum_j \frac{\partial E_n}{\partial a_j}\frac{\partial a_j}{\partial w_i} = \sum \delta_j^{(n)} \cdot z_{ji}^{(n)}$

5.29. $\frac{\partial \tilde{E}}{\partial w_i} = \frac{\partial E}{\partial w_i} + \lambda \sum_j \gamma_j(w_i) \frac{(w_i - \mu_j)}{\sigma_j^2}$; verify ; $\tilde{E}(w) = E(w) + \lambda \Omega(w)$

$$p(w) = \prod_i p(w_i)$$

$$p(w_i) = \sum_{j=1}^{M} \pi_j N(w_i|\mu_j, \sigma_j^2)$$

$$\Omega(w) = -\sum_i \ln\left(\sum_{j=1}^{M} \pi_j N(w_i|\mu_j, \sigma_j^2)\right)$$

$$\gamma_j(w) = \frac{\pi_j N(w|\mu_j, \sigma_j^2)}{\sum_k \pi_k N(w|\mu_k, \sigma_k^2)}$$

$$\frac{\partial \tilde{E}}{\partial w_i} = \frac{\partial E}{\partial w_i} + \lambda \frac{\partial \Omega}{\partial w_i} = \frac{\partial E}{\partial w_i} + \lambda \frac{\partial}{\partial w}\left[-\sum \ln\left(\sum_{j=1}^{M} \pi_j N(w_i|\mu_j, \sigma_j^2)\right)\right]$$

$$= \frac{\partial E}{\partial w_i} + \lambda \sum_j \frac{\pi_j N(w_i|\mu_j, \sigma_j^2)}{\sum \pi_j N(w_i|\mu_j, \sigma_j^2)} \frac{(w_i - \mu_j)}{\sigma_j^2} = \boxed{\frac{\partial E}{\partial w_i} + \lambda \sum_j \gamma_j(w_i) \frac{(w_i - \mu_j)}{\sigma_j^2}}$$

5.30 Prove $\frac{\partial \tilde{E}}{\partial \mu_j} = \lambda \sum_i \gamma_j(w_i) \frac{(\mu_i - w_j)}{\sigma_j^2}$ $\boxed{\frac{\partial \tilde{E}}{\partial \mu_j} = \cancel{\frac{\partial E}{\partial \mu_j}}^{0} + \lambda \sum \frac{\pi_j N(w_j|\mu_j, \sigma_j^2)}{\sum \pi_j N(w_i|\mu_j, \sigma_j^2)} \frac{(\mu_i - w_j)}{\sigma_j^2}}$

5.31 Prove $\frac{\partial \tilde{E}}{\partial \sigma_j} = \lambda \sum_i \gamma_j(w_i)\left(\frac{1}{\sigma_j} - \frac{(w_i - \mu_j)^2}{\sigma_j^3}\right) = \cancel{\frac{\partial E}{\partial \sigma_j}}^{0} + \lambda \sum \gamma_j(w_j) \frac{\partial}{\partial \sigma_j}\left(\frac{(w_j - \mu_j)^2}{2\sigma^2}\right)$

$$= \lambda \sum_j \gamma_j(w_i)\left(-\frac{(w_j - \mu_j)}{\sigma^3}\right)$$

5.32 $\pi_j = \dfrac{\exp(\eta_j)}{\sum_{k=1}^{m} \exp(\eta_k)}$; $\dfrac{\partial \pi_k}{\partial \eta_j} = \dfrac{\exp(\eta_j)'(\sum_{k=1}^{m} \exp(\eta_k)) - \exp(\eta_j)(\sum_{k=1}^{m} \exp(\eta_k))'}{(\sum_{k=1}^{m} \exp(\eta_k))^2}$

$$\boxed{= \delta_{jk} \cdot \pi_k - \pi_j \pi_k}$$

5.33. $\boxed{\begin{aligned} x_1 &= L_1 \cos(\theta_1) \\ x_2 &= L_2 \cos(\theta_2) \end{aligned}}$

5.34 $\dfrac{\partial E_n}{\partial a_k^{\pi}} = \pi_k - \gamma_k$ Derive ; $E(w) = -\sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k(x_n, w) N(t_n \mid \mu_k(x_n, w), \sigma_k^2(x_n, w)) \right\}$

$$\frac{\partial E(w)}{\partial a_k^{\pi}} = \frac{-\pi_k(x_n, w) N(t_n \mid \mu_k(x_n,w), \sigma_k^2(x_n, w))}{\sum \pi_k(x_n, w) N(t_n \mid \mu_k(x_n, w), \sigma_k^2(x_n, w))} - \left(-\frac{\pi_n(x_n, w)}{\sum \pi_k(x_n, w)}\right)$$

$$= \pi_k - \gamma_k$$

$\dfrac{\pi_k N_k}{\sum \pi_l N_l} - \dfrac{\pi}{\pi_l}$

5.35. Derive $\dfrac{\partial E_n}{\partial a_{kl}^{\mu}} = \gamma_k \left\{ \dfrac{\mu_{kl} - t_l}{\sigma_k^2} \right\}$; $\dfrac{\partial E}{\partial a_{kl}^{\mu}} = -\dfrac{\pi_k(x_n, w) N(t_n \mid \mu_k(x_n, w), \sigma_k^2(x_n, w))}{\sum_{l=1}^{K} \pi_l N(t_n \mid \mu_k(x_n, w), \sigma_k^2(x_n, w))} \left(\dfrac{t_n - \mu_k}{\sigma^2} \right)$

5.36. Derive $\dfrac{\partial E_n}{\partial a_k^{\sigma}} = \dfrac{\partial}{\partial a_k^{\sigma}} - \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k(x_n, w) N(t_n \mid \mu_k(x_n, w), \sigma_k^2(x_n, w)) \right\}$

$$= \frac{\pi_k(x_n, w) N(t_n \mid \mu_k(x_n, w), \sigma_k^2(x_n, w))}{\sum \pi_k(x, w) N(t_n \mid \mu_k(x_n, w), \sigma_k^2(x_n, w))} \left\{ \frac{(t_n - \mu)^2}{\sigma^3} - \frac{1}{\sigma^1} \right\}$$

5.37. $E[t|x] = \int t p(t|x) dt = \sum_{k=1}^{K} \pi_k(x) \mu_k(x)$ $\quad u = \dfrac{(t_n - \mu_k)^2}{2\sigma^2}$

$$= \int t \sum_{k=1}^{K} \pi_k(x) N(t \mid \mu_k(x), \sigma_k^2(x)) dt \qquad \frac{du}{d\sigma} = \frac{(t_n - \mu_k)}{\sigma^3} \cdot e$$

$$= \sum_{k=1}^{K} \pi_k(x) \int N(t \mid \mu_k(x), \sigma_k^2(x)) \cdot t \, dt \qquad -\sigma^{-2}$$

$$\frac{1}{\sigma^3} N$$

$$\boxed{= \sum_{k=1}^{K} \pi_k(x) \mu_k(x)}$$

$$s^2(x) = E[\| t - E[t|x] \|^2 \mid x] = \int p((t - E[t|x])^2 \mid x) \, x \, d(t - E[t|x])$$

$$= \int \sum_{k=1}^{K} \pi_k(x) [N(t - E[t|x]) \mid \mu_k(x), \sigma_k^2(x)] dt + \int \sum_{k=1}^{K} \pi_k(x) N(t \mid \mu_k(x), \sigma_k^2(x)) dt$$

$$- \int \sum \pi_k(x) N(t \mid \mu_k(x), \sigma_k^2(x)) dt$$

$$= \sum_{k=1}^{K} \pi_k(x) \left[\sigma^2(x) + \mu_k(x) - \mu_k(x) \right]$$

4.26 $\int \phi(\lambda a) N(a|\mu,\sigma^2)\,da = \phi\left(\frac{1}{(\lambda^{-2}+\sigma^2)^{1/2}}\right)$; $a = \mu + \sigma z$

$$\frac{d}{d\mu}\left[\int \phi(\lambda a) N(a|\mu,\sigma^2)\,da\right] = \frac{d}{d\mu}\left[\int \phi(\lambda(\mu+\sigma z)) \cdot N(\mu+\sigma z|\mu,\sigma^2)\,d\mu\right]$$

$$= \frac{d}{d\mu}\left[\int_{-\infty}^{\infty} (1+e^{-\lambda(\mu+\sigma z)})^{-1} \cdot e^{-\frac{1}{2}(\mu+\sigma z-\mu)^2/\sigma}\,d\mu\right] = \frac{d}{d\mu}\left[\int_{-\infty}^{\infty} (1+e^{-\lambda(\mu+\sigma z)})^{-1} \cdot e^{-\frac{1}{2}(\sigma z)^2/\sigma}\,d\mu\right]$$

$$= \frac{d}{d\mu}\left[e^{-\frac{1}{2}\sigma z^2} \cdot \int_{-\infty}^{\infty} (1+e^{-\lambda(\mu+\sigma z)})^{-1}\,d\mu\right] = \frac{d}{d\mu}\left[e^{-\frac{1}{2}\sigma z^2} \cdot \frac{-1}{\lambda}\left(-\lambda(\mu+\sigma z) - \ln|1+e^{-\lambda(\mu+\sigma z)}|\right)\right]$$

$$= \left[-\frac{1}{\lambda}\left(-\lambda - \frac{-\lambda(e^{-\lambda(\mu+\sigma z)})}{1+e^{-\lambda(\mu+\sigma z)}}\right)\right]$$

$$= 1 + \frac{e^{-\lambda(\mu+\sigma z)}}{1+e^{-\lambda(\mu+\sigma z)}}$$

$$\frac{d}{d\mu}\phi\left(\frac{\mu}{(\lambda^{-2}+\sigma^2)^{1/2}}\right) = \frac{d}{d\mu}\left[\frac{\mu}{1+e^{\sqrt{\lambda^{-2}+\sigma^2}}}\right] \frac{d}{d\mu}\left[(1+e^{\sqrt{\lambda^2+\sigma^2}})^{-1}\right]$$ $\boxed{\text{unsure}}$

Review or write in back of book

$\cancel{5.20}$ 5.21 $H_N = \sum_{n=1}^{N} b_n b_n^T$; $b_n \equiv \nabla_w a_n$; Prove $H_{L+1} = H_L + b_{L+1} b_{L+1}^T$

$$H_L = \sum_{n=1}^{L} b_n b_n^T ; \quad H_{L+1} = \sum_{n=1}^{L+1} b_n b_n^T = \sum_{n=1}^{L} b_n b_n^T + b_{L+1} b_{L+1}^T$$

$$(M+vv^T)^{-1} = M^{-1} - \frac{(M^{-1}v)(v^T M^{-1})}{1+v^T M^{-1} v} ; \quad (H_{L+1})^{-1} = (H_L + b_{L+1} b_{L+1}^T)^{-1} = H_L^{-1} - \frac{(H_L^{-1} b_{L+1})(b_{L+1}^T H_L^{-1})}{1+b_{L+1}^T H_L^{-1} b_{L+1}}$$

5.22. $$\frac{\partial^2 E_n}{\partial w_{kj}^{(2)} \partial w_{k'j'}^{(2)}} = z_j z_{j'} M_{kk'} ; \quad \frac{\partial^2 E_n}{\partial w_{ji}^{(1)} \partial w_{j'i'}^{(1)}} = x_i x_{i'} h''(a_{j'}) I_{jj'} \sum_k w_{kj}^{(2)} \delta_k + x_i x_{i'} h'(a_{j'}) h'(a_j) \sum_k \sum_{k'} w_{k'j'}^{(2)} w_{kj}^{(2)} M_{kk'}$$

$$\frac{\partial^2 E_n}{\partial w_{ji}^{(1)} \partial w_{kj'}^{(2)}} = x_i h'(a_{j'}) \left\{ \delta_k I_{jj'} + z_j \sum_{k'} w_{k'j'} H_{kk'} \right\}$$

6.1 Dual Representation: $J(a) = \frac{1}{2} a^T \Phi \Phi^T \Phi \Phi^T a - a^T \Phi \Phi^T t + \frac{1}{2} t^T t + \frac{\lambda}{2} a^T \Phi \Phi^T a$

$$a_n = -\frac{1}{\lambda}\{w^T \phi(x_n) - t_n\}$$

$$= \frac{1}{2}\left(-\frac{1}{\lambda}\{w^T \phi(x_n) - t_n\}\Phi\Phi^T\right)^2 - \frac{1}{\lambda}\{w^T \phi(x_n) - t_n\}\Phi\Phi^T t$$

$$+ \frac{1}{2} t^T t + \frac{\lambda}{2}\left\{-\frac{1}{\lambda}\{w^T \Phi(x_n) - t_n\}\right\}^2$$

$$\boxed{= \frac{\lambda}{2} w^T w + \frac{1}{2} \sum_n \{w^T \phi(x) - t_n\}^2}$$

6.2 $w^{(\tau+1)} = w^{(\tau)} - \eta \nabla E_p(w) = w^{(\tau)} + \eta \phi_n t_n$; Prove w is a linear comb. of $t_n \phi(x_n)$ $t \in \{-1, +1\}$

$$w = -\frac{1}{\lambda} \sum_{n=1}^{N} \{w^{(\tau)} \phi(x_n) - t_n\} \phi(x_n)$$

$$= -\frac{1}{\lambda} \sum_{n=1}^{N} \left\{ \sum_{k=1}^{n} \alpha_n \{w_n + \eta \phi_n t_n\} \phi(x_n) - t_n \right\} \phi(x_n)$$

$$\boxed{= \frac{-1}{\lambda} \sum_{n=1}^{N} \left\{ \sum_{k=1}^{n} \alpha_n \{w_n + \eta t_n K(x_n, x_n)\} - t_n \right\} \phi(x_n)}$$

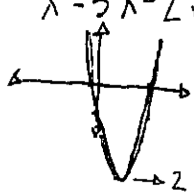
Possible error.

6.3. $p(x|C_k) = \frac{K_k}{N_k V}$; $\sum_k N_k = N$; $K = \sum_{n=1}^{N} k\left(\frac{x - x_n}{h}\right)$; $\|x - x_n\|^2$

$$\boxed{= \sum_{n=1}^{N} \frac{K \sqrt{K(\|x - x_n\|)}}{N_k V}}$$

6.4 $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$; $\lambda_1 = 5.37$, $\lambda_2 = -0.37$

$\lambda^2 - 5\lambda - 2$; $(\lambda-2)^2 - 5(\lambda-2) - 2$
$= \lambda^2 - 9\lambda - 4$;
$=$



1:9
2:9 -4
3:6 $-4 = 3.6 + x$
4:5 $-4 = 4.5 + x$

$\begin{pmatrix} -(\lambda+2)+1 & 2 \\ 3 & -(\lambda-2)+4 \end{pmatrix} = \boxed{\begin{pmatrix} 3 & 2 \\ 4 & 6 \end{pmatrix}}$ Guess, may be likely $\lambda^2 - \cancel{9}\lambda + 5 = 0$

$-\lambda$

$-$or

$\cancel{\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}}$ $\begin{pmatrix} 5 & 1/2 \\ 1/2 & 5 \end{pmatrix}$

6.5. verify $K(x, x') = c k_1(x, x') = c\, \phi(x)^T \phi(x') = c \sum_{i=1}^{M} \phi_i(x) \phi_i(x') = \boxed{\sum \phi_i(cx) \phi_i(x')}$

$K(x, x') = f(x) K_1(x, x') f(x') = f(x) \sum_{i=1}^{M} \phi_i(x) \phi_i(x') f(x') = \boxed{\sum_{i=1}^{M} \phi_i(f(x)x) \cdot \phi_i(f(x')x)}$

6.6 $K(x,x') = q(K_1(x,x')) = a(K_1(x_1,x'))^2 + b(K_1(x_1,x'))^2 + C$

$$\boxed{= a\left[\sum_{i=1}^{M} \phi_i(x)\phi_i(x')\right]^2 + b\left[\sum_{i=1}^{M} \phi_i(x)\phi_i(x')\right] + C}$$

$$= \cancel{\sum_{i=1}^{M} \phi_i^2(x)\phi_i^2(x)} + \ldots$$

$$K(x,x') = \exp(K_1(x,x')) = \boxed{e^{\sum_i^M \phi_i(x)\phi_i(x)}}$$

6.7. $K(x,x') = K_1(x,x') + K_2(x,x') = \sum_{i=1}^{M} \phi_i(x)\phi_i(x') + \sum_{i=1}^{n} \phi_i(x)\phi_i(x')$

$$= (x^T x')^2 + (x^T x')^2 = (x_1x_1' + x_2x_2')^2 + (x_1x_1' + x_2x_2')^2$$

$$= (x_1^2 x_1'^2 + 2x_1x_1'x_2x_2' + x_2^2x_2'^2 + x_1^2x_1'^2 + 2x_1x_1'x_2x_2' + x_2^2x_2'^2)$$

$$= (2(x_1^2, \sqrt{2}x_1x_2, x_2^2)(x_1'^2, \sqrt{2}x_1'x_2', x_2'^2)^T)$$

$$\boxed{= 2\sum_{i=1}^{2} \phi(x_i)\phi(x_i')}$$

$$K(x,x') = K_1(x,x')K_2(x,x') = \phi(x)^T\phi(x') \cdot \phi(x)^T\phi(x') = \sum_{i=1}^{M} \phi(x_i)\phi(x')\sum_{j=1}^{M} \phi(x_j)\phi(x_j')$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{M} \phi(x_i)\phi(x')\phi(x_j)\phi(x_j')$$

$$\boxed{= \sum_{l=1}^{K} \phi(x_l)\,\phi(x_l')}$$

6.8 $K(x,x') = K_3(\phi(x), \phi(x'))$

$$= \sum_{i=1}^{M} \phi_i[\phi(x)]\phi_i[\phi(x')] \quad \psi_i(x) = \phi_i[\phi(x)]$$

$$\boxed{= \sum_{i=1}^{M} \psi_i(x)\psi_i(x')}$$

$$K(x,x') = x^T A x' = \sum_{i=1}^{M}\sum^{N} x_i^T \phi_i\phi_i^T x_j \quad x_i\phi_i = \psi(x)$$

$$\boxed{= \psi_i(x)\psi_i(x)}$$

- ~~Check time~~
- Bayes N.K.v.
- Graph chapter 5 papers
- Equals

6.9. $K(x,x') = K_a(x_a,x_a') + K_b(x_b,x_b') = \sum \phi_i(x_a)\phi_i(x_a') + \sum_i \psi(x_b)\psi(x_b')$

$= (x_a^T x_a')^2 + (x_b^T x_b')^2 = (x_{1a}^2 x_{1a}'^2 + 2x_{1a}x_{1a}'x_{2a}x_{2a}' + x_{2a}^2 x_{2a}'^2)$
$+ (x_{1b}^2 x_{1b}'^2 + 2x_{1b}x_{1b}'x_{2b}x_{2b}' + x_{2b}^2 x_{2b}'^2)$

$= (x_{1a}^2, \sqrt{2}x_{1a}x_{2a}, x_{2a}^2)(x_{1a}'^2, \sqrt{2}x_{1a}'x_{2a}', x_{2a}'^2) + (x_{1b}^2, \sqrt{2}x_{1b}x_{2b}, x_{2b}^2)(x_{1b}'^2, \sqrt{2}x_{1b}'x_{2b}', x_{2b}'^2)$

$\boxed{= \sum_{A} \gamma(x')\gamma(x')}$

$K(x,x') = K_a(x_a,x_a')K_b(x_b,x_b')$

$= \sum_i \phi_i(x_a)\phi_i(x_a') \sum_i \psi_i(x_b)\psi(x_b')$ $\qquad \phi_i(x_a)\psi(x_b) = \gamma(x)$

$\boxed{= \sum_i \gamma(x)\gamma(x')}$

6.10 $K(x,x') = f(x)f(x')$ $\boxed{y(x) = k(x)^T(K+\lambda I_N)^{-1}t = f(x)f(x')(K+\lambda I_N)^{-1}t}$

$\boxed{y(x) \propto f(x)}$

6.11 $K(x,x') = \exp(-x^Tx/2\sigma^2)\exp(x^Tx'/\sigma^2)\exp(-(x')^Tx'/2\sigma^2)$

Prove inner product of $K(x,x') = \exp(-\|x-x'\|^2/2\sigma^2)$

$= \exp(-(x^Tx - 2x^Tx' + x'^Tx')/2\sigma^2) = \exp^{-x^Tx/2\sigma^2} \cdot \exp^{-x^Tx'/\sigma^2} \exp^{-x'^Tx'/2\sigma^2}$

$= K(x,x)\exp^{-x^Tx'/2\sigma^2} K(x',x')$

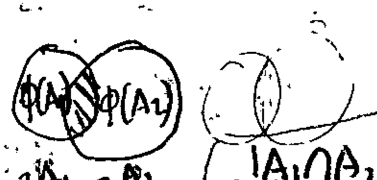
$= K(x,x)\left[1 - \frac{(x^Tx)^2}{2!2^2\sigma^4} + \frac{(x^Tx)^3}{3!2^3\sigma^6} - \frac{(x^Tx)^4}{4!2^4\sigma^8}\right] K(x',x')$

$= K(x,x)[1]K(x',x) - K(x,x)\frac{(x^Tx)^2}{2!2^2\sigma^4}K(x',x) + \cdots$

$\boxed{= \sum K(x,x)\frac{(-1)^n(x^Tx)^n}{n!2^n\sigma^{2n}}K(x',x)}$

6.12 $K(A_1,A_2) = 2^{|A_1 \cap A_2|}$; $\phi(A)$; $A \in D$, $\phi_U(A) = \begin{cases} 1, & \text{if } U \subseteq A \\ 0, & \text{otherwise} \end{cases}$

$= \phi(A_1)^T\phi(A_2) = \phi$



$2^{\phi_1} 2^{\phi_2} = \boxed{2^{|A_1 \cap A_2|}}$

6.13. $K(x,x') = g(\theta,x)^T F^{-1} g(\theta,x')$ $\quad \theta \to \psi(\theta)$; $\psi(\cdot)^{-1} = \psi(\cdot)$; $\psi'(\cdot) = \psi(\cdot)$

$$F = E_x\left[g(\theta,x) g(\theta,x)^T\right]$$

$$\Rightarrow g(\psi(\theta),x) F^{-1} g(\psi(\theta),x) = \frac{g(\psi(\theta),x) \cdot g(\psi(\theta),x)}{E_x[g(\theta,x), g(\theta,x)]}$$

$$\simeq \frac{g(\psi(\theta),x) \cdot g(\psi^{-1}(\theta),x)}{E_x[g(\theta,x), g(\theta,x)]}$$

$$\simeq \frac{g(\psi'(\theta),x) g(\psi'(\theta),x)}{E_x[g(\theta,x), g(\theta,x)]}$$

6.14 $K(x,x') = g(\mu,x)^T F^{-1} g(\mu,x') = \nabla_\mu \ln p(x|\mu)\, F^{-1}\, \nabla_\mu \ln p(x'|\mu)$

$$= \nabla_\mu \ln N(x|\mu,S) F^{-1} \nabla_\mu \ln N(x'|\mu,s)$$

$$= \left[\nabla \frac{(x-\mu)^2}{2S} - \frac{1}{2}\cancel{\nabla \ln 2\pi S}\right] F^{-1} \left[\nabla \frac{(x'-\mu)^2}{2S} - \frac{1}{2}\cancel{\nabla \ln 2\pi S}\right]$$

$$= \left[\frac{\nabla(x-\mu)^2}{2S}\right] \frac{1}{E_x\left[\frac{\nabla(x-\mu)}{2S}\right]^2} \left[\frac{\nabla(x'-\mu)^2}{2S}\right]$$

$$= (x-\mu)^T \cdot \frac{1}{S} \cdot (x'-\mu)$$

6.15. $K_{22} = k(x_{21}, x'_{21}) = \frac{1}{\alpha}\phi(x_n)^T\phi(x'_n)$; Cauchy-Schwartz inequality

$$K(x,x_2)^2 \le K(x_1,x_1)K(x_2,x_2)$$

$$\frac{1}{\alpha^2}K(x,x')^2 = \frac{1}{\cancel{\alpha^2}}\left[\phi(x)^T\phi(x')\right]^2 \le \frac{\phi(x)^T\phi(x)\,\phi(x')\phi(x')}{\cancel{\alpha^2}}$$

6.16 W_N ; x_n ; $\phi(x)$; $J(w) = f(w^T\phi(x_1), \cdots, w^T\phi(x_n)) + g(w^Tw)$

$g(\cdot)$ is increasing

$$w = \sum_{n=1}^{N} \alpha_n \phi(x_n) + w_\perp$$

$$\frac{\partial J(w)}{\partial w} = f'(w^T\phi(x_1), \cdots, w^T\phi(x_n)) \cdot \phi(x_1)\phi(x_2)\cdots\phi(x_n) + 2g'(w^Tw)w^T$$

$$\boxed{w = -\frac{f'(w^T\phi(x_1)\cdots w^T\phi(x_n))}{2g'(w^Tw)} \cdot \prod_{i=1}^{N}\phi(x_i)}$$

6.17 $E = \frac{1}{2}\sum_{n=1}^{N}\int\{y(x_n+\xi) - t_n\}^2 v(\xi)d\xi$

$$E[y(x)+\epsilon\eta(x)] = \frac{1}{2}\sum_{n=1}^{N}\int\{y(x+\xi)+\epsilon\eta(x+\xi)-t\}^2 v(\xi)d\xi$$

$$\frac{\partial E}{\partial y(x)} = \sum_{n=1}^{N}\int(y(x+\xi)-t)\eta(x_n+\xi)v(\xi)d\xi = 0$$

$$= \sum_{n=1}^{N}\int\{y(x+\xi)-t_n\}\delta(x_n+\xi-z)v(\xi)d\xi \quad z = x+\xi;$$

$$= \sum_{n=1}^{N}\int\{y(z)-t_n\}v(z-x_n)\eta(z)dz = 0$$

$$\cancel{= \sum\int y(z)v(z-x_n)\eta(z) - \sum\int t_n}$$

$$= \sum\int y(z)v(\cancel{z}-x_n)\eta(\cancel{z})d\cancel{z} = \sum\int t_n v(z-x_n)\cancel{\eta(z)}d\cancel{z}$$

$$\boxed{y(z) = \frac{\sum t_n}{\sum\int y(z)}}$$

x, t ; $N(x|t|\sigma^2 I)$; $p(t|x)$ for $E[t|x]$ and $\mathrm{var}[t|x]$ for $K(x, x_n)$

6.18. $K(x, x_n) = \dfrac{g(x-x_n)}{\sum_m g(x-x_m)}$; $g(x) = \int_{-\infty}^{\infty} f(x,t)\,dt$

$$= \frac{\int f(x-x_n, t)}{\sum \int f(x-x_n, t)} = \frac{\int N(x-x_n|t, \sigma^2 I)}{\sum \int N(x-x_n|t, \sigma^2 I)} = \frac{\sqrt{\pi}\cdot\sigma^2}{m\sqrt{\pi}\,\sigma^2} \boxed{\frac{1}{m}}$$

$\boxed{6.19}$. ~~$t_n = y(z_n) + N(t|\sigma^2)$~~

$x_n = z_n + \xi_n = z_n + N(z|\sigma^2) = z_n + g(\xi)$; consider $\{x_n, t_n\}$

$E = \frac{1}{2}\sum_{n=1}^{N} \int \{y(x_n - \xi_n) - t_n\}^2 g(\xi_n)\, d\xi_n$; Nadaraya-Watson ;

$\dfrac{\partial E}{\partial y(z)} = \sum_{n=1}^{N} \int \{y(x_n - \xi_n) - t_n\} \cdot g(\xi_n)\, d\xi_n$;

$$K(x, x') = \frac{g(x - x_n)}{\sum g(x - x_n)}$$

$$y(x) = \frac{\sum g(x - x_n) t_n}{\sum g(x - x_n)} = \sum K(x, x') t_n$$
$$= \sum \phi_n(x) \phi(x_n') t_n$$

$$F[y(x) + \epsilon\eta(x)] = F[y(x)] + \epsilon \int \frac{\partial F}{\partial y(x)} \eta(x)\, dx + O(\epsilon^2)$$

$$F[y(z) + \epsilon\eta(z)] = \frac{1}{2} \sum \int \{y(x_n - \xi_n) + \epsilon\eta(x_n - \xi) - t_n\}^2 g(\xi_n)\, d\xi_n$$
$$+ \epsilon \int \sum_{n=1}^{N} \{y(x_n - \xi_n) + \epsilon\eta(x_n - \xi) - t_n\} g(\xi_n)\, d\xi_n$$

$\boxed{6.20}$ $m(x_{n+1}) = K^T C_N^{-1} t$; $\sigma^2(x_{n+1}) = C - K^T C_N^{-1} K$; $C_{N+1} = \begin{pmatrix} C_N & K \\ K^T & c \end{pmatrix}$

$c = k(x_{N+1}, x_{N+1}) + \beta^{-1}$

$p(t_{N+1}) = N(t_{N+1} | 0, C_{N+1})$

$= \frac{1}{\sqrt{2\pi C_{N+1}}} e^{-(t)^2/C_{N+1}} = \frac{1}{\sqrt{2\pi C_{N+1}}} e^{-t^2/(C_N \cdot c - K^T K)}$

$-t^2/C_N(c - K^T C^{-1} K)$

$= \frac{1}{\sqrt{2\pi C_{N+1}}} e$

$= \frac{1}{\sqrt{2\pi C_{N+1}}} \frac{(t-m)^2}{2\sigma^2}$

$\frac{(t-m)^2}{2} = \frac{t^2}{C_N}$; $\frac{t^2 - tm + m^2}{2} = \frac{t^2}{C_N}$

$\mu_{a|b} = \mu_a + \Sigma_{aa} \Sigma_{bb}^{-1} (x_b - \mu_b)$

$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba}$

$\Sigma_{aa} = C_N, \Sigma_{ab} = k, \Sigma_{ba} = k^T; \Sigma_{bb} = c$

$\boxed{\sigma^2(x_{n+1}) = C_N - k C_N^{-1} k^T}$

$\boxed{m(x_{n+1}) = 0 + C_N K^T (X - 0)}$

6.21 $K(x, x') = \Sigma \phi(x)\phi(x')$; $p(t | x, t, \kappa, \beta) = N(t | m_N^T \phi(x), \sigma_N^2(x))$

$\boxed{\text{Woodbury Identity}}$

C.6 $(I + AB)^{-1} A = A(I + BA)^{-1}$; $[AB]^2 = B^T A^T$

C.7 $(A + BD^{-1}C)^{-1} = A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1}$: $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$

$t_n = y_n + \epsilon_n$

$p(t_n | y_n) = N(t_n | y_n, \beta^{-1})$: $p(t|y) = N(t | y, \beta^{-1} I_N)$; $\boxed{p(t_{n+1} | x_{n+1}) = N(t_{n+1} | x_{n+1}, \beta^{-1})}$

$C_N = \underbrace{\Sigma \phi(x)^T \phi(x)}_{\alpha} + \beta^{-1} I_N$; $m(x_{n+1}) = k^T C_N^{-1} t = \Sigma a_n k(x_n, x_{n+1})$ 6.66

$= \sum_{n=1}^{N} C_N^{-1} t K(x_n, x_{n+1})$

$\boxed{\sigma^2(x_{n+1})}$

$\boxed{= \sum_{n=1}^{N} \frac{t\, k(x_n, x_{n+1})}{\underbrace{\Sigma \phi(x)^T \phi(x)}_{\alpha} + \beta^{-1} I_n}}$

6.22 $x = \langle x_1, \dots, x_n \rangle$

$\cancel{L = \langle x_{n+1}, \dots, x_{n+L} \rangle}$

$\boxed{t(x) = N(t \mid m(x_{N+1}), \sigma^2(x_{n+1}))}$ $\rightarrow p(t) = \dfrac{p(t_n)}{\sum p(t_n)}$

$= N(t(x) \mid k^T C_N^{-1} t, c - k^T C_N^{-1} k)$

$p(t_j) = \dfrac{p(t_j)}{\sum_{j=N+1}^{j=N+L} p(t_j)}$

$\boxed{= N(t_j \mid m(x_j), \sigma^2(x_j)) \cdot N(t(x), m(x_{N+1}), \sigma^2)_{j=N+1}}$

6.23. $p(t_D \mid m(x_D), \sigma^2(x_D)) = N(x \mid \mu, \Sigma) = \dfrac{1}{(2\pi)^{D/2}} \dfrac{1}{|\Sigma|^{1/2}} \exp^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$

$p(t_{N+1}(x) \mid X_{N+1})$

$= N(x_{D+1} \mid \mu_{D+1}, \Sigma_{D+1}) = \cancel{\dfrac{1}{(2\pi)^{(D+1)/2}} \dfrac{1}{|\Sigma^{-1} k|} e^{-\frac{1}{2}(x-}}$

N

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$\boxed{= \dfrac{1}{(2\pi)^{(D+1)/2}} \dfrac{1}{|c - \mu^T \Sigma^{-1} \mu|} e^{-\frac{1}{2}(x - \mu^T \Sigma \mu)^T (c - \mu^T \Sigma^{-1} \mu)^{-1} (x - \mu^T \Sigma \mu)}}$

6.24 W; $0 < W_{ii} < 1$ $\boxed{W_{TOTAL} = \sum_i^D \sum_j^D [W'_{ij} + W_{ij}]}$ if $(W'_{ij} > 0 \wedge W_{ij} > 0)$

then $W_{ij} \geq 0$

$W_{ij} \cdot W_{ij} \geq 0$

6.25 $W^{(new)} = W^{(old)} - H^{-1} \nabla E(W)$: Newton - Raphson

$f(x) = \sum_{n=1}^{N} w_n h(\|x - x_n\|)$; $a_N^{new} = C_N(I + W_N C_N)^{-1} \{t_n - \sigma_N + W_N a_N\}$

$\nabla \Psi(a_n) = t_n - \sigma_n - C_N^{-1} a_N = -E(w)$

$-\nabla\nabla \Psi(a_n) = +W_N + C_N^{-1} = H$

$a_N^{new} = (I + W_N C_N)^{-1} C_N \{t_n - \sigma_n\} + \{I + W_N C_N\}^{-1} C_N W_N a_N$

$= (\cancel{I + W_N C_N})^{-1} a_N^{*} +$

6.25 $a_N^{(new)} = a_N^{(old)} + (W_N + C_N)^{-1}\{t_n - \sigma_n - C_N^{-1} a_N\} = \underbrace{C_N(t_N - \sigma_N)}_{a_1} + (\underbrace{W_N}_{W_n} + C_N^{-1})^{-1}\{t_n - \sigma_n - \underbrace{C_N^{-1} a_n}_{W_n}\}$

$$= (W_{n+1} + C_{N+1}^{-1})^{-1}\{t_{n+1} - \sigma_{n+1} - C_{N+1}^{-1} a_{n+1}\}$$

$$= \frac{1}{W_{N+1} + \frac{1}{C_{N+1}}}\{t_{n+1} - \sigma_{n+1} - C_{N+1}^{-1} a_{n+1}\}$$

$$\boxed{= \frac{C_N}{W_{n+1} C_N + I}\{t_{n+1} - \sigma_{n+1} - C_{N+1}^{-1} a_{n+1}\}}$$

$\boxed{6.26}$ $p(y) = N(y \mid A\mu + b, L^{-1} + A\cancel{\Lambda} A^T)$ $\to 0$

$p(x|y) = N(x \mid \Sigma\{A^T L(y-b) + \Lambda\mu\}, \Sigma)$ $\quad \Sigma = (\Lambda + A^T L A)^{-1}$

$E[a_{N+1} | t_N] = K^T(t_N - \sigma_N)$ $\quad p(a_{N+1}|t_N) = N(a_{N+1} \mid K^T C_N^{-1} a_N, C - K^T C_N^{-1} K)$

$\mathrm{var}[a_{N+1} | t_n] = C - K^T(W_N^{-1} + C_N)^{-1} K$

$E[a_{N+1} | t_n] = \int p(a_{N+1}|a_N) p(a_N|t_N) da_N = \int N(a_{N+1} \mid K^T C_N^{-1} a_N, C - K^T C_N^{-1} K) N(a_N \mid a_N^*, H^{-1}) da_N$

$= \int N(a_{N+1} \mid K^T C_N^{-1} a_N, C - K^T C_N^{-1} K) N(a_N \mid C_N(t - \sigma_N), W_N + C_N^{-1}) da_N$

$E[x|y] = (\Lambda + A^T L A)^{-1}\{A^T L(y-b) + A\mu\}$

$\cancel{= (C_N + K^T C^{-1} K)\{K^T C^{-1} C_N^{-1}(t - \sigma) + K}$

$\mathrm{cov}[x|y] = (\Lambda + A^T L A)^{-1}$: $L = (W_N + C_N^{-1})^{-1}$, $A = K$, $\Lambda = C$

$\boxed{\mathrm{Cov}[a_{n+1} | t_n] = (C - K^T(W_N + C_N^{-1})^{-1} K)^{-1}}$

$E[a_{N+1} | t_N] = (C + K^T(W_N + C_N^{-1})^{-1} K)^{-1}\{K^T(W_N + C_N^{-1})(t - \sigma) + K \cdot 0\}$

$= \boxed{}$

6.22 $X = \langle x_1, \dots, x_n \rangle$

$L = \langle x_{n+1}, \dots, x_{n+L} \rangle$

$$p(t) = \frac{p(t_n)}{\sum p(t_n)}$$

$$\boxed{t(x) = N(t \mid m(x_{N+1}), \sigma^2(x_{n+1}))}$$

$$p(t_j) = \frac{p(t_j)}{\sum_{j=N+1}^{N+L} p(t_j)}$$

$$= N(t(x) \mid k^T C_N^{-1} t, c - k^T C_N^{-1} k)$$

$$\boxed{= N(t_j \mid m(x_j), \sigma^2(x_j)) \cdot N(t(x), m(x_{N+1}), \sigma^{-2})_{j=N+1}}$$

6.23. $$p(t_D \mid m(x_D), \sigma^{-2}(x_D)) = N(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

$$p(t_{N+1}(x) \mid X_{N+1})$$

$$= N(x_{D+1} \mid \mu_{D+1}, \Sigma_{D+1}) = \cancel{\frac{1}{(2\pi)^{(D+1)/2}} \frac{1}{|C - k^T \Sigma^{-1} k|}} e^{\cancel{-\frac{1}{2}(x-}}$$

N

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\boxed{= \frac{1}{(2\pi)^{(D+1)/2}} \frac{1}{|c - \mu^T \Sigma^{-1} \mu|} e^{-\frac{1}{2}(X - \mu^T \Sigma \mu)^T (c - \mu^T \Sigma^{-1} \mu)^{-1} (X - \mu^T \Sigma \mu)}}$$

6.24 W ; $0 < W_{ii} < 1$

$$\boxed{W_{TotAL} = \sum_{i}^{D} \sum_{j}^{D} [W'_{ij} + W_{ij}]}$$

if $(W'_{ij} > 0 \wedge W_{ij} > 0)$

then $W_{ij} \geq 0$

$W_{ij} \cdot W_{ij} \geq 0$

$=$

6.25 $W^{(new)} = W^{(old)} - H^{-1} \nabla E(W)$: Newton-Raphson

$$f(x) = \sum_{n=1}^{N} w_n h(\|x - x_n\|) \; ; \; a_N^{new} = C_N (I + W_N C_N)^{-1} \{t_n - \sigma_N + W_N a_N\}$$

$$\nabla \Psi(a_n) = t_n - \sigma_n - C_N^{-1} a_N = -E(w)$$

$$-\nabla\nabla \Psi(a_n) = +W_N + C_N^{-1} = H$$

$$a_N^{new} = (I + W_N C_N)^{-1} C_N \{t_n - \sigma_n\} + \{I + W_N C_N\}^{-1} C_N W_N a_N$$

$$= (I \cancel{+ W_N C_N})^{-1} a_N^* +$$

7.1 $x_n, t_n \in \{-1,1\}$, Parzen Kernel = Parzen Window = $K(u) = \begin{cases} 1 & |u_i| \leq 1/2 \\ 0, & \text{otherwise} \end{cases}$ $i=1,\ldots,D$

$$K(x,x') = \prod^{N} v\, N(x_n|\mu_1,\Sigma)^{t_N} \, v\, N(x'_n|\mu_2,\Sigma)^{1-t_N}$$

$$= \prod^{N} v^2\, p(x_n|\mu_1,\Sigma)^{t_N} \cdot N(x_n|\mu_2,\Sigma)^{1-t_N}$$

if $K(x,x') = x^T x'$

$p(t|x) \propto p(x|t)p(t)$

$\hookrightarrow p(x|t) = \frac{1}{N_t}\sum_{t_n=1}^{N} \frac{1}{Z_k} K(x,x')\delta(t,t_n)$

$K(u) = \begin{cases} +1 & \text{iff } \frac{1}{N_+}\sum_{t_i=1} K(\tilde{x},x) \geq \frac{1}{N_-}\sum_{t_i=1}^{N} K(\tilde{x},x) \\ -1 & \text{otherwise} \end{cases}$

$\boxed{K(u) = \operatorname{sgn}\left(\frac{1}{N}\sum t \cdot K(\tilde{x},x)\right)}$ $\boxed{K(u) = \operatorname{sgn}\left(\begin{smallmatrix}+1\\-1\end{smallmatrix}\; x^T x \geq x^T x\right)}$

$\boxed{7.2}$ $t_n(w^T\phi(x_n)+b) \geq 1 \quad n=1,\ldots,N$ if $1 = \gamma > 0$

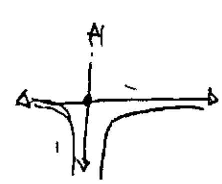
$t_n(w^T\phi(x_n)+b) \geq \gamma > 0$

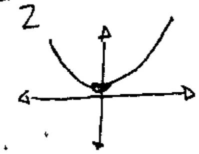
$$\frac{t_n y(x_n)}{\|w\|} = \frac{t(w^T\phi(x_n)+b)}{\|w\|} \geq \gamma > 0 \; ; \; \arg\max\left\{\frac{1}{\|w\|}\arg\min t(w^T\phi(x_n)+b)\right\} \geq \gamma > 0$$

Normalization $\neq 1$

$$\frac{d}{dw}\|w\|^{-1} = -\|w\|^2 \qquad \frac{d}{dw}\frac{\|w^2\|}{2} = \|w\|$$

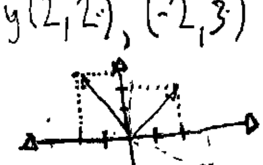
$$\boxed{\left[\frac{+}{(+)}\right] \geq \gamma > 0}$$





$\cancel{7.3}$ Prove irrespective of D, a dataset of two datapoints, 2 classes, is sufficient to determine location of max margin hyperplane!

$y(2,2), (-2,3)$



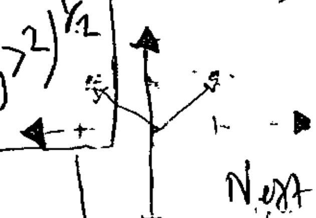
$\frac{3-2}{-2-2}$

$(-4,1)$ $= \frac{1}{-4}$

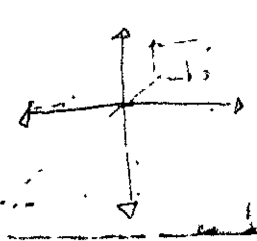
$$y = w^T\phi(x) + b$$

$$= \frac{[\langle\bar{x}\rangle - \langle\bar{y}\rangle]}{2} K(x,y) + (\langle\bar{x}\rangle^2 + \langle y\rangle^2)^{1/2}$$

$$= \frac{\left[\frac{1}{N}\sum_i^N x_i - \frac{1}{K}\sum_{i}^{K} y_i\right]}{N+K} K(\phi(x),\phi(y)) + \sqrt{\left[\frac{1}{N}\sum_0^N x_i\right]^2 - \left[\frac{1}{K}\sum_{0}^{K} y_i\right]^2}$$



Next page



$\boxed{7.4}$: Show the value of ρ for maximum-margin hyperplane

$$\frac{1}{\rho^2} = \sum_{n=1}^{N} a_n \; ; \; \tilde{L}(a) = \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{M} a_n a_m t_n t_m K(x_n, x_m)$$

$$a_n \geq 0$$

$$\sum_{n=1}^{N} a_n t_n = 0$$

$$\cancel{\|w\|^2 = \sqrt{w^T w} = \sum_{n=1}^{N} a_n = \frac{1}{\rho^2} \; ; \; \rho = \frac{1}{\|w\|}}$$

$$\cancel{\tilde{L}(a) = \frac{1}{2}\|w\|^2 = \frac{1}{2}\frac{1}{\rho^2} \; , \; 2\tilde{L}(a) = \frac{1}{\rho^2}}$$

$$\tilde{L}(a) = \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m t_n t_m K(x_n, x_m)$$

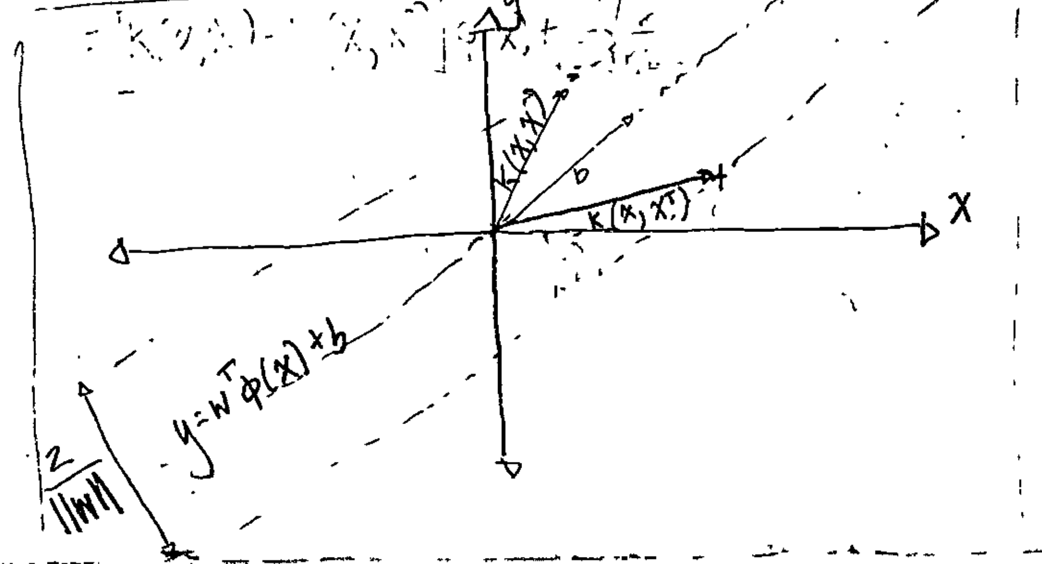
$$\boxed{\frac{2}{\|w\|^2} = \sum_{n=1}^{N} a_n \; ; \; \rho = \frac{\sqrt{2}}{\|w\|} \; ; \; \frac{1}{\rho^2} = \sum^{N} a_n}$$

7.3: Prove irrespective of D, a dataset of two datapoints, two classes, is sufficient to determine location of max margin hyper-plane.

$$y = w^T\phi(x) + b \; ; \; \sum_{n=1}^{N} a_n = w^T \; ; \; K(x,\cdot) = \phi(x) \; ; \; b = \frac{1}{2}(\|c_-\|^2 - \|c_+\|^2)$$

$$= \frac{1}{n_+}\sum K(x, x_+) - \frac{1}{n_-}\sum K(x, x_-)$$

$$\boxed{y = \left[\frac{1}{n_+}\sum K(x,x^+) - \frac{1}{n_-}\sum K(x,x^-)\right]\phi(x) + \frac{1}{2}\left[\sqrt{\frac{1}{n_+}\sum K(x,x^+)}^{\,2} + \sqrt{\frac{1}{n_-}\sum K(x,x^-)}^{\,2}\right]}$$



7.5. $\rho = \frac{\sqrt{2}}{\|W\|} = \sum_{n=1}^{N} a_n$; $2\cancel{L(w,b,a) = 2\sum_{n=1}^{N} a_n} = 2 \cdot W = \frac{2 \cdot \sqrt{2}}{\rho}$

$$\boxed{\frac{1}{\rho^2} = \frac{2}{\|W\|^2} = 2 \cdot \sum_{n=1}^{N} a_n = 2\hat{L}(a)}$$

7.6. $t \in \{-1, 1\}$; if $p(t=1|y) = \sigma(y)$; where $y(x) = W^T\phi(x) + b$

Prove $-\log(\text{likelihood} + \text{quadratic reg term})$

$$= \sum_{n=1}^{N} E_{LR}(y_n t_n) + \lambda \|W\|^2$$

$$\boxed{-\log(p(t=1|y)) + \lambda\|W\|^2 = -\log\left(\frac{1}{1+e^{-W^T\phi(x)+b}}\right) + \lambda\|W\|^2}$$

7.7. $L = C\sum_{n=1}^{N}(\xi_n + \hat{\xi}_n) + \frac{1}{2}\|W\|^2 - \sum_{n=1}^{N}(\mu_n \xi_n + \hat{\mu}_n \hat{\xi}_n) - \sum_{n=1}^{N} a_n(\epsilon + \xi_n + y_n - t_n) - \sum_{n=1}^{N} \hat{a}_n(\epsilon + \hat{\xi}_n - y_n + t_n)$

$\frac{\partial L}{\partial W} = \|W\|$; $0 = \|W\| = \sum_{n=1}^{N}(a_n - \hat{a}_N)\phi(x_n)$

$\frac{\partial L}{\partial b} = 0$; $0 = b$

$\frac{\partial L}{\partial \xi_n} = C - \sum_{n=1}^{N}\mu_N - \sum_{n=1}^{N} a_n$; $0 = C - \sum_{n=1}^{N}\mu_N - \sum_{n=1}^{N} a_n$; $C = \mu_N + a_N$

$\frac{\partial L}{\partial \hat{\xi}_n} = C - \sum_{n=1}^{N}\hat{\mu}_N - \sum_{n=1}^{N}\hat{a}_N$; $0 = C - \sum_{n=1}^{N}\hat{\mu}_N - \sum_{n=1}^{N}\hat{a}_N$; $C = \hat{\mu}_N + \hat{a}_N$

Backsubstitution:

$$\tilde{L}(a, a) = (\mu_N + a_N)\sum_{n=1}^{N}(\xi_N + \hat{\xi}) - \sum_{n=1}^{N}(\mu_N \xi_N + \hat{\mu}_N \hat{\xi}_N)$$

$$- \sum_{n=1}^{N} a_n(\epsilon + \xi_N + y_N - t_N) - \sum_{n=1}^{N}\hat{a}_N(\epsilon + \hat{\xi}_N - y_N + t_N)$$

$$= \mu_N \cancel{\sum} \xi_N + \mu_N \sum \hat{\xi}_N + a_N \sum \xi_N + a_N \sum \hat{\xi}_N - \sum \cancel{\mu_N} \xi_N - \sum \hat{\mu}_N \hat{\xi}_N - \sum_{n=1}^{N}$$

7.4 Show that the value ρ of the margin for maximum-margin hyperplane

$$\frac{1}{\rho^2} = \sum_{n=1}^{N} a_n \;;\; a_n \quad \tilde{L}(a) = \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m t_n t_m k(x_n, x_m)$$

$$a_n \geq 0 \quad n=1 \ldots N$$

$$\sum_{n=1}^{N} a_n t_n = 0$$

$$\tilde{L}(a) = \frac{1}{\rho^2} - \sum_{n=1}^{N}\sum_{m=1}^{N} \frac{1}{\rho_n^2}\frac{1}{\rho_m^2} t_n t_m K(X_n, X_m)$$

if $\frac{1}{\rho^2} t_n = 0$; then $t_n = 0$

$$\tilde{L}(a) = \frac{1}{\rho^2}$$

$\boxed{\frac{1}{\rho^2} = \|W\|^2 \text{ because } \rho = \frac{1}{\|W\|}.}$

7.7, 7.8, 7.10

7.5 $\frac{1}{\rho^2} = 2\tilde{L}(a)$; where $\frac{1}{\rho^2} = \|W\|^2$

7.8 Prove when $\xi > 0$; $a_n = C$ $\quad (C - a_n)\xi = 0$; $C\xi = a_n\xi$; $\boxed{C = a_n}$

and $\hat{\xi}_n > 0$; $\hat{a}_n = C$ $\quad (C - \hat{a}_n)\xi = 0$; $C\xi = \hat{a}_n\xi$; $\boxed{C = \hat{a}_n}$

7.9. $m = \beta \Sigma \Phi^T t \qquad p(W|t, X, \alpha, \beta) = N(W|m, \Sigma)$

$\Sigma = (A + \beta\phi^T\phi)^{-1} \qquad p(W|t) = N(W|m_n, S_n)$

$$m_n = S_N(S_0^{-1}m_0 + \beta\phi^T t)$$

$p(W|t, X, \alpha, \beta) = N(W|m, \Sigma) \qquad S_N^{-1} = S_0^{-1} + \beta\phi^T\phi$

$= p(\beta|t, X, \alpha) \cdot p(\alpha|X, t) \cdot p(t|X)\, p(X)$

Example: $y(x, W) = \sum_{j=0}^{M-1} W_j \phi_j(x) = W^T\phi(x)$; $p(t|X, W, \beta) = N(t|y(x, W), \beta^{-1})$

$$p(W|t) = N(W|m_N, S_N)$$

$$p(t|t, \alpha, \beta) = \int p(t|W, \beta)\, p(W|t, \alpha, \beta)\, dW$$

$$\therefore p(t|X, t, \alpha, \beta) = \int N(t|\phi(x)^T W, \beta^{-1})\, N(W|m_N, S_N)\, dW$$

Since $p(y|x) = N(y|Ax+b, L^{-1})$ & $p(x) = N(x|\mu, \Lambda^{-1})$

We have $p(y) = N(y|A\mu + b, L^{-1} + A\Lambda^{-1}A^T)$

$\therefore p(t|X, t, \alpha, \beta) = \int p(t|X, w, \beta) \cdot p(w|m_N, S_N) \cdot p(w|m_{N-1}, S_{N-1}) \cdots p(w|\alpha^{-1}I)\,dw$

$$p(t|t, X, \alpha, \beta) = \prod_{n=1}^{N} p(t_n|x_n, w, \beta^{-1}) \prod^{N} N(w_i|0, \alpha_i^{-1})$$

$$= \prod_{n=1}^{N} p(t_n|\phi_n, \beta^{-1}) \prod_{n=1}^{N} N(w_i|0, \alpha_i^{-1})$$

$$= \prod_{n=1}^{N} N(t|\phi_n, \beta^{-1}) \prod_{n=1}^{N} N(w_i|0, \alpha_i^{-1})$$

$$\cancel{= N(\textstyle\sum t_n | \sum \phi_n, \beta^{-1}) N(\sum w_i | 0, \alpha_i^{-1})}$$

$$= \frac{\beta^{1/n}}{(2\pi)^{1/n}} e^{-\frac{\beta}{2}(t_n - \phi_n)^2} \cdot \frac{\alpha^{1/n}}{(2\pi)^{1/n}} e^{-\frac{\alpha}{2} w_i^2}$$

$$\cancel{= \frac{\beta^{1/n} \cdot \alpha^{1/n}}{(2\pi)} e^{-\frac{\beta}{2}(t_n - \phi_n)^2 - \frac{\alpha}{2} w_i^2}} =$$

$$= \prod_{n=1}^{N} N(t|\phi_n, \beta^{-1}) \prod_{n=1}^{N} N(w_i|0, \alpha^{-1}) = \left(\frac{\beta}{2\pi}\right)^{N/2} \exp^{-\frac{\beta}{2}\sum (t-\phi_n)^2} \cdot \left(\frac{\alpha}{2\pi}\right)^{N/2} \exp^{-\frac{\alpha}{2}\sum w_i^2}$$

$$= \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} \left(\frac{\alpha}{2\pi}\right)^{N/2} \exp^{-\frac{\beta}{2}\sum(t-\phi_n)^2 - \frac{\alpha}{2}\sum w_i^2}$$

$\cancel{= (}$

7.9 $p(w|t, X, \alpha, \beta) = p(w|\phi^T t, \alpha, \beta) = p(w|\phi^T t, \beta^{-1}) \cdot p(w|\alpha) = N(w|\phi^T t, \beta^{-1}) N(w|0, \alpha_i^{-1})$

$$= \left(\frac{\beta}{2\pi}\right)^{N/2} \cdot \left(\frac{\alpha}{2\pi}\right)^{N/2} \cdot \exp^{-\frac{\beta}{2}\{w - \phi^T t\}^2} \cdot \exp^{-\frac{\alpha}{2} w^T w}$$

$$= \left(\frac{\beta}{2\pi}\right)^{N/2} \left(\frac{\alpha}{2\pi}\right)^{N/2} \exp^{-\frac{\beta}{2}\{w - \phi^T t\}^2 - \frac{\alpha}{2} w^T w} = \left(\frac{\beta}{2\pi}\right)^{N/2} \cdot \left(\frac{\alpha}{2\pi}\right)^{N/2} \exp^{-\frac{\beta}{2}(w^T w - 2w\phi^T t + (\phi^T t)^2) - \frac{\alpha}{2} w^T w}$$

$$= \left(\frac{\beta}{2\pi}\right)^{N/2} \cdot \left(\frac{\alpha}{2\pi}\right)^{N/2} \cdot \exp^{-\frac{\beta}{2}(w^T w) - \frac{\alpha}{2} w^T w + \beta w \phi^T t - \frac{\beta}{2}(\phi^T t)^2} = \left(\frac{\beta}{2\pi}\right)^{N/2} \cdot \left(\frac{\alpha}{2\pi}\right)^{N/2} \cdot e^{u}$$

$$u = -\frac{\beta}{2}(w^T w) - \frac{\alpha}{2} w^T w + \beta w \phi^T t - \frac{\beta}{2}(\phi^T t)^2$$

$$= -\frac{1}{2}(\alpha + \beta) w^T w + \beta w \phi^T t - \frac{\beta}{2}(\phi^T t)^2 = -\frac{1}{2}(\alpha+\beta)\left[w^T w - \frac{2\beta w \phi^T t}{(\alpha+\beta)}\right] - \frac{\beta}{2}(\phi^T t)^2$$

$$\frac{\phi^T\left[A^{-1}\frac{\partial A}{\partial\alpha}A^{-1}\right]\phi\Sigma}{(\beta\alpha)} - \frac{t^T\Sigma\phi\left[-A^{-1}\frac{\partial A}{\partial\alpha}A^{-1}\right]\phi\Sigma t}{(\beta A)^2} = 0$$

$$\phi^T\left[A^{-1}\frac{\partial A}{\partial\alpha}A^{-1}\right]\phi\Sigma(\beta A) - t^T\Sigma\phi\left[-A^{-1}\frac{\partial A}{\partial\alpha}A^{-1}\right]\phi\Sigma t = 0 \qquad m = \beta\Sigma\phi^T t$$

$$\phi^T\left[A^{-1}\cancel{\frac{\partial A}{\partial\alpha}}A^{-1}\right]\phi\Sigma(\beta A) - \frac{m\left[-A^{-1}\cancel{\frac{\partial A}{\partial\alpha}}A^{-1}\right]m}{\beta^2} = 0 \qquad \gamma = 1-\alpha\Sigma_{ii}$$

$$\ln p(t|X,\alpha,\beta) = \frac{N}{2}\ln\beta + \frac{1}{2}\Sigma\ln\alpha_i - E(t) - \frac{1}{2}\ln|\Sigma| - \frac{N}{2}\ln(2\pi)$$

$$\frac{\partial\ln p(t|X,\alpha,\beta)}{\partial\alpha} = \frac{1}{2\alpha} - \frac{1}{2}\Sigma_{ii} - \frac{1}{2}m_i^T m \;; \quad 0 = 1-\alpha\Sigma_{ii} - \alpha m^T m$$

$$\boxed{\alpha = \frac{1-\alpha\Sigma_{ii}}{m^T m}}$$

$$\frac{\partial\ln p(t|X,\alpha,\beta)}{\partial\beta} = \frac{N}{2\beta} - \frac{\phi^T\phi}{2}\ln\alpha_i + \frac{t^T t}{2} - \frac{\phi^T\phi}{2\Sigma} + \frac{m^T m}{\Sigma^2} \quad \phi^T\phi$$

$$= \frac{N}{2\beta} - \frac{\phi^T\phi}{2}\cancel{\ln\alpha_i + \frac{t^T t}{2}} \qquad m$$

$$= \left(\frac{N}{\beta} - \|t-m\phi\|^2 - \mathrm{Tr}\left[\Sigma\phi^T\phi\right]\right)\phi\Sigma^{-1}t \quad 2\phi\beta\Sigma\phi\, t\, t$$

$$\Sigma\phi^T\phi = \Sigma\phi^T\phi + \frac{\Sigma A}{\beta} - \frac{\Sigma A}{\beta}$$

$$= \Sigma\left(\beta\phi^T\phi + A\right)\beta^{-1} - \frac{\Sigma A}{\beta}$$

$$= \left(A+\beta\phi^T\phi\right)^{-1}\left(\phi^T\phi\beta + A\right)\beta^{-1} - \beta^{-1}\Sigma A$$

$$= \left(I - A\Sigma\right)\beta^{-1}$$

$$= (\gamma_i)\beta$$

$$0 = \frac{1}{2}\left(\frac{N}{\beta} - \|t-m\phi\|^2 - \mathrm{Tr}\left[\gamma_i\beta\right]\right) \Rightarrow \frac{\|t-m\phi\|^2}{N-\mathrm{Tr}\,\gamma\beta} = \beta$$

$$\Sigma\phi^T\alpha$$

$$p(t|x,X,t,\alpha^*,\beta) = \int p(t|x,w,\beta^*)\,p(w|X,t,\alpha^*,\beta^*)\,dw$$

$$= \int N(t_n|x_n,w,\beta)\circ N(w|m,\Sigma^{-1})\,dw$$

$$= \int \left(\frac{\beta}{2\pi\Sigma^{-1}}\right)^{N/2} e^{-\frac{1}{2}\left[(t_n-W\phi)^T\beta(t_n-W\phi)+(W-m)^T\Sigma^{-1}(W-m)\right]}\,dW$$

$$= \int \left(\frac{\beta}{2\pi\Sigma^{-1}}\right)^{N/2} e^{-\frac{1}{2}\left(\left[t_n^2-2W\phi t+(W\phi)^2\right]\beta+\left[W^2-2Wm+m^2\right]\Sigma^{-1}\right)}$$

$$= \int \left(\frac{\beta}{2\pi\Sigma^{-1}}\right)^{N/2} e^{-\frac{1}{2}\left(\left[t_n^2-2W\phi t+(W\phi)^2\right]\beta+\left[W^2-2W\beta\Sigma^{-1}\phi^T t+(\beta\Sigma^{-1}\phi^T t)^2\right]\Sigma^{-1}\right)}$$

$$= \int \left(\frac{\beta}{2\pi\Sigma^{-1}}\right)^{N/2} e^{-\frac{1}{2}\left[(t_n^2-2Wt)\beta+(-2W\beta\Sigma^{-1}\phi^T t+(\beta\Sigma^{-1}\phi^T t)^2)\Sigma^{-1}\right]}\cdot e^{-\frac{1}{2}\left[(W\phi)^2\beta+W^2\Sigma^{-1}\right]}\,dW$$

$$= \int \left(\frac{\beta}{2\pi\Sigma^{-1}}\right)^{N/2}\circ e^{-\frac{1}{2}\left[(\cancel{\beta}+\beta\Sigma^{-3}\phi^T\phi)t^2-2(1+\phi^T\Sigma^{-1})Wt\beta\right]}\circ e^{-\frac{1}{2}W^2\left[\Sigma^{-1}+\beta\phi^T\phi\right]}$$

$\boxed{3.10}$ ~~3.59~~
$$p(t|x,X,t,\alpha^*,\beta) = \int p(t|x,w,\beta^*)\,p(w|X,t,\alpha^*,\beta^*)\,dW$$

$$= \int N(t_n|x_n,W,\beta^{-1})\cdot N(W|m,\Sigma^{-1})\,dW$$

$$= \int \underbrace{N(t_n|W\phi(x),\beta^{-1})}\circ N(W|m,\Sigma^{-1})\,dW$$

$$\cdot N(y|\cancel{XA},L^{-1})\circ N(X|\mu,\Lambda^{-1})$$

$$= \int N(y|A\mu+b,\,L^{-1}+A\Lambda^{-1}A^T) = \boxed{\int N(t_n|\cancel{\phi(x)m},\beta^{-1}+\phi(x)\Sigma^{-1}\phi(x))\,dW}$$

$$= -\frac{1}{2}(\alpha+\beta)w^T w + \beta w \phi^T t - \frac{\beta}{2}(\phi^T t)^2 = -\frac{1}{2}\left[(\alpha+\beta)w^T w - 2\beta w \phi^T t + \beta(\phi^T t)^2\right]$$

$$p(W \mid t, X, \alpha, \beta) = p(t \mid W, \beta)\, p(W \mid \alpha) \qquad W\phi - t$$

$$= N(t \mid W\phi, \beta^{-1}) \cdot N(W \mid \alpha^{-1})$$

$$= \left(\frac{\beta}{2\pi}\right)^{N/2} \left(\frac{\alpha}{2\pi}\right)^{N/2} \exp^{-\frac{\beta}{2}(t - W\phi)^2} \cdot \exp^{-\frac{\alpha}{2}(W^T W)}$$

$$= \left(\frac{\beta}{2\pi}\right)^{N/2} \left(\frac{\alpha}{2\pi}\right)^{N/2} \exp^{-\frac{\beta}{2}(t^2 - 2W\phi t + (W\phi)^2) - \frac{\alpha}{2} W^T W}$$

$$\eta = -\frac{\beta}{2}(t^2 - 2W\phi t + (W\phi)^2) - \frac{\alpha}{2} W^T W$$

$$= -\frac{1}{2}\left[\beta t^2 - 2\beta W\phi t + \beta(W\phi)^2 + \alpha W^T W\right]$$

$$= -\frac{1}{2}\left[\beta t^2 - 2\beta W\phi t + (\alpha + \beta\phi^T\phi) W^T W\right]$$

$$= -\frac{1}{2}\left[(\alpha+\beta\phi^T\phi)\left[W^T W - \frac{2\beta W\phi t}{(\alpha+\beta\phi^T\phi)}\right] + \beta t^2\right]$$

$$= -\frac{1}{2}\left[\Sigma^{-1}\left[W^T W - 2 m_n W\right] + \beta t^2\right]$$

Leftover

$$\boxed{-\frac{1}{2}\left[\Sigma^{-1}[W - m_n]^2 - m_n^T \Sigma^{-1} m_n + \beta t^2\right]}$$

$$-\beta\Sigma^{-1}\phi t \,\Sigma^{-1} \beta \Sigma^{-1} \phi t + \beta t^2$$

$$\beta^2 \Sigma^{-1} \phi^T \phi t^2 + \beta t^2$$

$$(\beta^2 \Sigma^{-1} \phi^T \phi + \beta) t^2$$

17.10 $$p(t \mid X, \alpha, \beta) = \int p(t \mid X, w, \beta)\, p(w \mid \alpha)\, dw$$

$$= \int \left(\frac{\beta}{2\pi}\right)^{N/2} \exp^{-\frac{\beta}{2}(t - W\phi)^2} \cdot \left(\frac{\alpha}{2\pi}\right)^{N/2} \exp^{-\frac{\alpha}{2} W^T W}\, dw$$

$$= \left(\frac{\beta}{2\pi}\right)^{N/2} \cdot \left(\frac{\alpha}{2\pi}\right)^{N/2} \cdot \int \exp^{-\frac{1}{2}(W - m_n)^T \Sigma^{-1} (W - m_n)} \cdot \exp^{-\frac{m^T \Sigma m}{2}} \cdot \exp^{-\frac{\beta}{2} t^T t}\, dW$$

$$= \frac{\sqrt{\Sigma / \beta\alpha}}{\sqrt{2\pi}} \cdot \exp^{-\frac{m^T \Sigma m}{2}} \cdot \exp^{-\frac{\beta}{2} t^T t}$$

$$= \frac{\sqrt{\Sigma/\beta\kappa}}{\sqrt{2\pi}} \circ \exp^{\frac{-\beta\Sigma\phi^T t \circ \Sigma^{-1} \circ \beta\Sigma\phi^T t}{2}} \circ \exp^{-\frac{\beta t^T t}{2}}$$

$$= \frac{\sqrt{\Sigma/\beta\kappa}}{\sqrt{2\pi}} \circ \exp^{-\frac{1}{2}[\beta^2\Sigma\phi^T\phi - \beta] t^T t}$$

$$\frac{\beta^2\phi^T\phi}{A+\beta\phi^T\phi} - \beta = \frac{\beta^2\phi^T\phi}{A+\beta\phi^T\phi} - \frac{\beta(A+\beta\phi^T\phi)}{(A+\beta\phi^T\phi)}$$

$$= \frac{\beta^2\phi^T\phi - \beta A - \beta^2\phi^T\phi}{(A+\beta\phi^T\phi)}$$

$$= \frac{-\beta A}{(A+\beta\phi^T\phi)} = C$$

$$= \frac{\sqrt{C}}{\sqrt{2\pi}} \exp^{-\frac{1}{2} t^T C t}$$

$$\boxed{\ln \frac{\sqrt{C}}{\sqrt{2\pi}} \exp^{-\frac{1}{2} t^T C t} = \frac{1}{2}\left[\ln C - \ln 2\pi - \frac{1}{2} t^T C^{-1} t\right]}$$

7.11 $\left| \frac{C}{2\pi} e^{-\frac{1}{2} t^T C^{-1} t} = N(t|0, C^{-1}) \right. \quad p(y) = N(y|A\mu + b, L^{-1} + A\Lambda^{-1}A^T)$

7.12 $\ln p(t|X,\alpha,\beta) = \ln N(t|0,C) = -\frac{1}{2}\{N\ln(2\pi) + \ln|C| + t^T C^{-1} t\}$; $C = \beta^{-1}I + \phi A^{-1}\phi^T$

$$\frac{d\ln p(t|X,\alpha,\beta)}{d\alpha} = \frac{-1}{2}\{\cancel{N\ln(2\pi)} + \mathrm{Tr}\left(C^{-1}\frac{\partial C}{\partial\alpha}\right) - t^T C^{-1}\frac{\partial C}{\partial\alpha} C^{-1} t\}$$

$$\frac{d\ln p(t|X,\alpha,\beta)}{d\beta} = \{\mathrm{Tr}\left(C^{-1}\frac{\partial C}{\partial\beta}\right) - t^T C^{-1}\frac{\partial C}{\partial\beta} C^{-1} t\} \frac{-1}{2}$$

$$\gamma_i := 1 - \alpha_i \Sigma_{ii}$$

$$\mathrm{Tr}\left(C^{-1}\phi^T[-A^{-1}\frac{\partial A}{\partial\alpha}A^{-1}]\phi\right) - t^T C^{-1}\phi^T[-A^{-1}\frac{\partial A}{\partial\alpha}A^{-1}]\phi C^{-1} t = 0$$

$$C^{-1}\phi^T[-A^{-1}\frac{\partial A}{\partial\alpha}A^{-1}]\phi[1 - t^T t] = 0$$

$$\frac{\phi^T[-A^{-1}\frac{\partial A}{\partial\alpha}A^{-1}]\phi(A+\beta\phi^T\phi)}{\beta A}[1 - t^T t] = 0$$

$$\frac{\phi[-A^{-1}\frac{\partial A}{\partial\alpha}A^{-1}]\phi\Sigma}{\beta A}[1 - t^T t] = 0$$

$$C^{-1}\phi^T[-A^{-1}\frac{\partial A}{\partial\alpha}A^{-1}]\phi - t^T C^{-1}\phi[-A^{-1}\frac{\partial A}{\partial\alpha}A^{-1}]\phi C^{-1} t = 0$$

$$\frac{\phi^T[-A^{-1}\frac{\partial A}{\partial\alpha}A^{-1}]\phi\Sigma^{-1}}{\beta A} - \frac{t^T\Sigma\phi[-A^{-1}\frac{\partial A}{\partial\alpha}A^{-1}]\phi\Sigma^{-1}t}{(\beta A)^2} = 0$$

7.13 $\alpha_i^{New} = \frac{\gamma_i}{m_i^2}$; $(\beta^{New})^{-1} = \frac{\|t - \Phi m_N\|^2}{N - \sum_i \gamma_i}$; $\ln p(t|X, \alpha, \beta) = \ln N(t|0, C)$

$$= -\frac{1}{2}\{N \ln(2\pi) + \ln|C| + t^T C^{-1} t\}$$

$$\mathrm{Gam}(\tau|a, b) = \frac{1}{\Gamma(a)} b^a \tau^{a-1} e^{-b\tau}$$

Maximize α ~~and~~ β of $p(t, \alpha, \beta|X)$

$$\cancel{p(t|X)p(\alpha|X)p(\beta|X)} = \cancel{\frac{1}{\Gamma(t)} \cdot 0 \cdot t^{x-1} \cdot \frac{1}{\Gamma(\alpha)} \alpha^{x-1} \cdot \frac{1}{\Gamma(\beta)} \beta^{x-1}}$$

$$\cancel{= \frac{1}{\Gamma(t)\Gamma(\alpha)\Gamma(\beta)} t^{x-1} \cdot \alpha^{x-1} \cdot \beta^{x-1}}$$

$$\cancel{\frac{d\,p(t|X)p(\alpha|X)p(\beta|X)}{d\alpha} = \frac{(x-1)\alpha^{x-2} \cdot \Gamma(\alpha) - \alpha^{x-1} \cdot \Gamma'(\alpha)}{\Gamma^2(\alpha)} \cdot \frac{t^{x-1} \cdot \beta^{x-1}}{\Gamma(t)\Gamma(\beta)}}$$

$$p(t, \cancel{\alpha}, \beta|X) = p(t|X)\,p(\cancel{\alpha, \beta|X})$$

$$= \cancel{p(t|X)p(\alpha|X)}p(\beta|X)$$

$$\mathrm{Gam}(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)} \beta^\alpha x^{\alpha-1} e^{-\beta x}$$

$$L = \prod_i \Gamma(\alpha, \beta)_x = \frac{1}{\Gamma^i(\alpha)} \beta^\alpha x_1^{\alpha-1} \cdots x_i^{\alpha-1} \cdot e^{-\beta x}$$

$$\log L = (\alpha - 1)\sum_i \log x_i - \cancel{\beta x} + \alpha \log\beta - \log\Gamma(\alpha)$$

$$\frac{\partial \log L}{\partial \alpha} = \sum \log x_i + \log\beta = 0 ; \boxed{\beta = -\prod x_i}$$

$$\frac{\partial \log L}{\partial \beta} = -x + \frac{\alpha}{\beta} = 0 ; \boxed{\alpha = \beta x}$$

7.14 $p(t|x, X, t, \alpha^*, \beta^*) = \int p(t|x, w, \beta^*)\, p(w|X, t, \alpha^*, \beta)\, dw = N(t|m^T\phi(x), \sigma^2(x))$

$$= \int N(t|y(x), \beta^{-1})\, N(w|m, \Sigma)\, dw = \int N(t|y(x), \beta^{-1}) \cdot N(t|x_n|w, \beta^{-1})\, N(w|0, \alpha^{-1})$$

Prove $\sigma^2(x) = (\beta^*)^{-1} + \phi(x)^T \Sigma\, \phi(x)$; where $\Sigma = (A + \beta\Phi^T\Phi)^{-1}$

$$= \int N(t|y(x), \beta^{-1})\, N(\ \ , \Sigma)\, dw = \left(\frac{\beta^*}{2\pi\Sigma}\right)^{N/2} \int e^{[(t - w\phi)^T \beta^{*-1}(t - w\phi) + (t - w\phi)^T \Sigma^{-1}(t - w\phi)]\frac{-1}{2}}\, dw$$

$t\Sigma_i + 2w\phi\Sigma_i + w\phi^2\Sigma_i$

$$= \left(\frac{\beta^*}{2\pi\Sigma}\right)^{N/2} \int e^{-\frac{1}{2}}$$

7.15 $|C| = |C_{-i}|\,|1+\alpha_i^{-1}\varphi_i^T C_{-i}^{-1}\varphi_i|$; where $C^{-1} = C_{-i}^{-1} - \dfrac{C_{-i}^{-1}\varphi_i\varphi_i^T C_{-i}^{-1}}{\alpha_i + \varphi_i^T C_{-i}^{-1}\varphi_i}$

Prove $\ln p(t|X,\alpha,\beta) = \ln N(t|0,C) = -\frac{1}{2}\{N\ln(2\pi) + \ln|C| + t^T C^{-1} t\}$

could be $L(\alpha) = L(\alpha_{-i}) + \lambda(\alpha_i)$

$$= -\frac{1}{2}\left\{N\ln(2\pi) + \ln|C_{-i}|\,|1+\alpha_i^{-1}\varphi_i^T C_{-i}^{-1}\varphi_i| + t^T\left(C_{-i}^{-1} - \frac{C_{-i}^{-1}\varphi_i\varphi_i^T C_{-i}^{-1}}{\alpha_i+\varphi_i^T C_{-i}^{-1}\varphi_i}\right)t\right\}$$

$$= -\frac{1}{2}\left\{N\ln(2\pi) + \ln|C_{-i}|\,|1+\alpha_i^{-1}s_i| + t^T\left(C_{-i}^{-1}\right)t - \frac{q_i^2}{\alpha_i+s_i}\right)$$

$$= -\frac{1}{2}\left\{N\ln(2\pi) + \ln|C_{-i}| - \ln|\alpha_i+s_i| + \ln\alpha_i + t^T C_{-i}^{-1} t - \frac{q_i^2}{\alpha_i+s_i}\right\}$$

$$\boxed{= -\frac{1}{2}\left\{N\ln(2\pi) + \ln|C_{-i}| + t^T C_{-i}^{-1} t\right\} - \frac{1}{2}\left\{\ln\alpha_i - \ln|\alpha_i+s_i| - \frac{q_i^2}{\alpha_i+s_i}\right\}}$$

$$\boxed{= L(\alpha_{-i}) + \lambda(\alpha_i)}$$

7.16. $\dfrac{\partial^2\lambda(\alpha_i)}{\partial^2\alpha} = \dfrac{\partial}{\partial\alpha}\cdot\dfrac{\partial\lambda(\alpha_i)}{\partial\alpha_i} = \dfrac{\partial}{\partial\alpha}\cdot\dfrac{1}{2}\left[\dfrac{1}{\alpha_i} - \dfrac{1}{\alpha_i+s_i} - \dfrac{q_i^2}{(\alpha_i+s_i)^2}\right]$

$$~~= \frac{1}{2}\left[\frac{-1}{\alpha_i^2} + \frac{1}{(\alpha_i+s_i)^2} + \frac{2q_i^2}{(\alpha_i+s_i)^3}\right] = 0;~~ \quad \frac{1}{(\alpha_i+s_i)^2} + \frac{2q_i^2}{(\alpha_i+s_i)^3} = \frac{1}{\alpha_i^2}$$

$$~~\frac{(\alpha_i+s_i)+2q_i^2}{(\alpha_i+s_i)^3} = \frac{1}{\alpha_i^2}~~$$

$$= \partial_\alpha\cdot\frac{1}{2}\left[\frac{\alpha_i^{-1}s_i^2 - (q_i^2-s_i)}{(\alpha_i+s_i)^2}\right]$$

$$= \frac{-\alpha_i^{-2}s_i^2(\alpha_i+s_i)^2 - 2\alpha_i^{-1}s_i^2(\alpha_i+s_i)^3}{(\alpha_i+s_i)^4} + \frac{2(q_i^2-s_i)}{(\alpha_i+s_i)^3} = 0;$$

$$= \frac{-\alpha_i^{-2}s_i^2(\alpha_i+s_i)^{\cancel{2}}}{(\alpha_i\cancel{+s_i})} - \frac{2\alpha_i^{-1}s_i^2(\alpha_i+s_i)^{\cancel{3}2}}{(\alpha_i+\cancel{s_i})} + 2(q_i^2-s_i) = 0$$

$$= -\alpha_i^{-2}(\alpha_i+s_i) + 2\alpha_i^{-1}\cdot(\alpha_i+s_i) + \frac{2(q_i^2-s_i)}{s_i^2} = 0$$

$$= \frac{-(\alpha_i+s_i)~~+2\alpha~~(\alpha_i+s_i)}{\alpha_i^2} + \frac{2(q_i^2-s_i)}{s_i^2} = 0$$

$$\frac{(\alpha_i+s_i)}{\alpha_i^2} = -\frac{2(q_i^2-s_i)}{s_i^2}; \quad \boxed{\frac{-s_i^2}{2(q_i^2-s_i)} = \frac{\alpha_i^2}{(\cancel{\alpha_i}+s_i)}}$$

7.17. $\Sigma = (A + \beta\phi^T\phi)^{-1}$; $C = \beta^{-1} I + \phi A^{-1}\phi^T$; $(A + BD^{-1}C)^{-1} = A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1}$

$$s_i = \varphi_i^T C^{-1} \varphi_i = \varphi_i^T [\beta - \beta^2 \phi^T \Sigma \phi] \varphi_i = \boxed{\varphi_i^T \beta \varphi_i - \varphi_i^T \beta^2 \varphi_i \phi^T \Sigma \phi}$$

$$Q_i = \varphi_i^T C^{-1} t = \varphi_i^T [\beta - \beta\phi(A^{-1} + \phi\beta\phi)^{-1}\phi\beta] t = \varphi_i^T [\beta - \beta^2 \phi^T \Sigma \phi] t$$

$$\boxed{= \varphi_i^T \beta t - \beta^2 \varphi_i^T \phi^T \Sigma \phi t}$$

7.18. $\ln p(w|t,\alpha) = \ln\{p(t|w)p(w|\alpha)\} - \ln p(t|\alpha)$

$$= \sum_{n=1}^{N} \{t_n \ln y_n + (1-t_n)\ln(1-y_n)\} - \frac{1}{2} w^T A w + \text{const}$$

$$\nabla \ln p(w|t,\alpha) = \frac{t_n}{y_n} \cancel{\sigma(1-\sigma)\phi(x)} \cancel{+ \frac{1-t_n}{1-y_n} \sigma(1-\sigma)\phi(x) - \frac{1}{2} w^T A w}$$

$$\cancel{= \frac{t_n}{y_n} w^T\phi(x)(1 - w^T\phi(x)) - \frac{1-t_n}{1-y_n}}$$

$$= \frac{t_n - \cancel{y_n t_n} - y_n + \cancel{y_n t_n}}{y_n(1-y_n)} - Aw = \boxed{\left(\frac{t_n - y_n}{b}\right)\phi - Aw}$$

$$\nabla\nabla \ln p(w|t,\alpha) = \boxed{-\phi^T b^{-1} \phi + A}$$

7.19. $p(t|\alpha) = \int p(t|w)p(w|\alpha)dw$

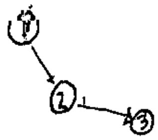
$$\simeq p(t|w^*)p(w^*|\alpha)(2\pi)^{M/2}|\Sigma|^{1/2} \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(t-w^*)^2} \cdot e^{-\frac{1}{2}(w^*-\alpha)^2} \cdot (2\pi)^{M/2}|\Sigma|^{1/2}$$

$$\ln p(t|\alpha) = \ln p(t|w^*)p(w^*|\alpha)(2\pi)^{M/2}|\Sigma|^{1/2} = -\frac{1}{2}(t-w^*)^2 - \frac{1}{2}(w^*-\alpha)^2 - \cancel{\ln(2\pi)} + \frac{M}{2}\cancel{\ln(2\pi)} + \frac{1}{2}\ln|\Sigma|$$

$\nabla\nabla \alpha^* - \alpha$

8.1 $p(x)=\prod_{k=1}^{K} p(x_k|pa_k) = p(x_1,\dots,x_K) = p(x_K|x_1,\dots,x_{K-1})\,p(x_{K-1}|x_1,\dots,x_{K-2})\cdots p(x_2|x_1)\,p(x_1)$

if $\tilde{p}(x) = \dfrac{p(x)}{\sum p(x_n)} = \dfrac{p(x_1\dots x_K)}{\sum p(x_n)} = \boxed{\dfrac{p(x_K|x_1\dots x_{K-1})\,p(x_{K-1}|x_1\dots x_{K-2})\cdots p(x_2|x_1)\cdot p(x_1)}{\sum p(x_n)}}$

8.2 
$\boxed{p(x_1,x_2,x_3) = p(x_3|x_1,x_2)\,p(x_2|x_1)\,p(x_1) = \prod_{n=1}^{3} p(x_n|pa_n)}$

8.3 $p(a,b|c) = \dfrac{p(a,b,c)}{p(c)} = \dfrac{p(a)p(c|a)p(b|c)}{p(c)} = p(a|c)p(b|c)$

$p(a,b) = p(a)\sum_c p(c|a)p(b|c) = p(a)p(b|a)$ $\boxed{\text{Independent by Marginalization}}$

Proof $p(a,b) \neq p(a)p(b)$

8.4. $p(a,b,c) = p(a)p(c|a)p(b|c) = 0.192$ $\boxed{p(a=1)=\frac{2}{5},\ p(a=0)=\frac{3}{5}}$

$p(b=0|c=0)\,p(a=0) = p(a,b,c) = 0.192$

$p(b=0|c=0)\,p(a=1) = p(a,b,c) = 0.192$

$p(b=1|c=1)\,p(a=0) = p(a,b,c) = 0.216$

$p(b=1|c=1)\,p(a=1) = p(a,b,c) = 0.096$

$p(b=0|c=0) = 0.192\cdot\frac{5}{3} = 0.32$
$p(b=0|c=0) = 0.192\cdot\frac{5}{2} = 0.48$ $\Big\}\ 0.32+0.48 = 0.80$

$p(b=1|c=1) = 0.216\cdot\frac{5}{3} = 0.36$
$p(b=1|c=1) = 0.096\cdot\frac{5}{2} = 0.24$ $\Big\}\ 0.36+0.24 = 0.60$

$p(c=0|a=0) = p(a,b,c)/p(b=0) = 0.192/0.592 = 0.324$

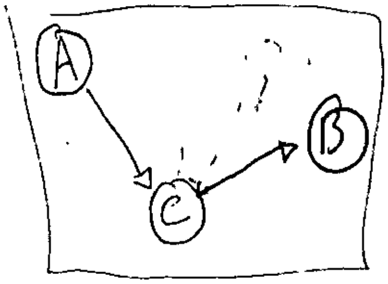
$p(c=0|a=0) = p(a,b,c)/p(b=1) = 0.048/0.408 = 0.118$

$p(c=1|a=1) = p(a,b,c)/p(b=0) = 0.064/0.592 = 0.108$

$p(c=1|a=1) = p(a,b,c)/p(b=1) = 0.096/0.408 = 0.235$

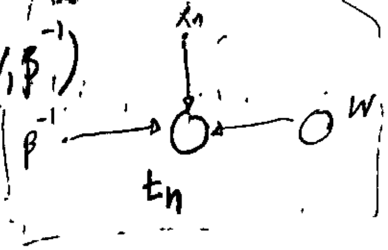
$\boxed{p(c=0|a=0) = 221/500}$ $\boxed{p(c=1|a=1) = 343/1000}$

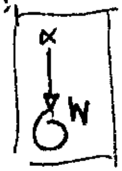
← Off by 15% $p(a,b,c) = p(a)p(c|a)p(b|c)$

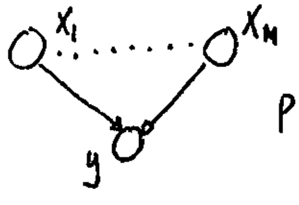


8.5 Graph of $p(t|X, \alpha, \beta) = \prod_{n=1}^{N} p(t_n | x_n, w, \beta^{-1})$

$p(w|\alpha) = \prod_{i=1}^{M} N(w_i | 0, \alpha_i^{-1})$





8.6. 

$p(y|x_1, \dots, x_M)$ $x_i \in \{0,1\}$ $2^M \longrightarrow M+1$

by $p(y=1|x_1, \dots, x_M) = \sigma\left(w_0 + \sum_{i=1}^{M} w_i x_i\right) = \sigma(w^T x)$

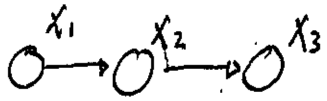
Alternatively, $p(y=1|x_1, \dots, x_M) = 1 - (1-\mu_0) \prod_{i=1}^{M} (1-\mu_i)^{x_i}$

μ_0 is the initial conditions average.

8.7. $E[x_i] = \sum_{j \in pa_j} w_{ij} E[x_j] + b_i$ $\quad cov[x_i, x_j] = E[(x_i - E[x_i])(x_j - E[x_j])]$

$$= E\left[(x_i - E[x_i])\left\{\sum_{k \in pa_j} w_{jk}(x_k - E[x_k]) + \sqrt{v_j}\,\epsilon_j\right\}\right]$$

$$= \sum_{k \in pa_j} w_{jk}\, cov[x_i, x_k] + I_{ij} v_j$$

Prove 

are $\mu = (b_1, b_2 + w_{21} b_1, b_3 + w_{32} b_2 + w_{32} w_{21} b_1)^T$

and

$$\Sigma = \begin{pmatrix} v_1 & w_{21} v_1 & w_{32} w_{21} v_1 \\ w_{21} v_1 & v_2 + w_{21}^2 v_1 & w_{32}(v_2 + w_{21}^2 v_1) \\ w_{32} w_{21} v_1 & w_{32}(v_2 + w_{21}^2 v_1) & v_3 + w_{32}^2 (v_2 + w_{21}^2 v_1) \end{pmatrix}$$

$E[x_3] = \mu_3 = \sum_{j \in pa_j}^{3} w_{ij} E[x_j] + b_i = w_{i1} \cdot E[x_1] + b_1$

$= w_{i1}[w_{2i} \cdot E[x_2] + b_2] + b_1$

$= w_{1j}[w_{2i}[w_{3j} E[x_3] + b_3] + b_2] + b_1$ if $w_{ij} = 1$

$\boxed{\mu = [b_1, b_2 + w_{21} b_1, b_3 + w_{21} w_{32} b_1 + w_{32} b_2]^T}$

$$\operatorname{cov}[X_i, X_j] = \sum_{k \in pa_j} W_{jk} \operatorname{cov}[X_i, X_k] + I_{ij} V_j \quad \text{if } I = 1;\ k = 1,2,3;\ j = 1,2,3$$

$$= W_{11} \operatorname{cov}[X_i, X_1] + V_{11} \qquad \text{if } W_{11} = 1$$

$$= \left[\sum_{2}^{3} W_{2i} \operatorname{cov}[X_i, X_2] + V_2\right] + V_1$$

$$= \left[\sum_{2}^{3} W_{2i}\left[\sum_i W_{3i} \operatorname{cov}[X_i, X_3] + V_3\right] + V_2\right] + V_1$$

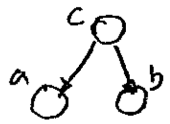
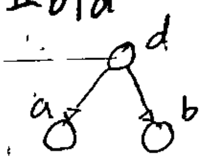
$$= \quad V_1, \quad V_1 V_2$$

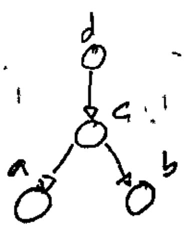
$$\operatorname{cov}[X_i, X_j] = \sum_K W_{jK}\left[\sum_i W_{iK}(1) + V_i\right] + I_{ij} V_j \quad ; \text{ if } I_{ij} = 1$$

$$= W_{j1}\left[W_{j2}\left[W_{j3}\left[\sum_i W_{iK}(1) + V_i\right] + V_2\right] + V_3\right]$$

$$= \boxed{\quad\quad\quad\quad} \ ?$$

8.8 Show $a \perp\!\!\!\perp b, c \mid d$ implies $a \perp\!\!\!\perp b \mid d$

 $= p(a,b|c)$ 

 $= \boxed{p(a,b|c)\, p(c|d) = \ \uparrow \ = p(a,b|d)}$

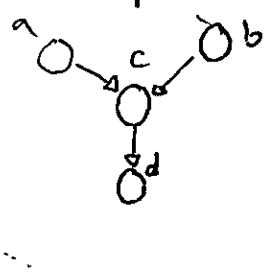
8.9 D-separation

(a) arrows meet head-to-tail or tail-to-tail at node, and is set C

(b) arrows meet head-to-head at node, and neither the node, nor descendants is set C.

$$\text{(A)} \longrightarrow \text{(X)} \longrightarrow \text{(B)}$$

$$p(a|x) \cdot p(x|b) = p(a|b)$$

8.10  $p(a,b,c) = p(a)p(b)p(c|a,b)$

$$\frac{p(a,b,c)}{p(c|a,b)} = \boxed{p(a,b) = p(a)p(b)}$$

8.11.

	G = 1		G = 0	
	B = 1	B = 0	B = 1	B = 0
F = 1	0.8	0.2	0.2	0.8
F = 0	0.2	0.1	0.8	0.9

$p(F=1)=0.9$

$p(F=0)=0.1$

$p(B=1)=0.9 \quad p(B=0)=0.1$

$p(G=1|F=1,B=1)=0.8 \qquad p(G=0|F=1,B=1)=0.2$

$p(G=1|F=0,B=1)=0.2 \qquad p(G=0|F=0,B=1)=0.8$

$p(G=1|F=1,B=0)=0.2 \qquad p(G=0|F=1,B=0)=0.8$

$p(G=1|F=0,B=0)=0.1 \qquad p(G=0|F=0,B=0)=0.9$

$$P(G=0) = \sum_{B\in 0,1}\sum_{F\in 0,1} p(G=0|B,F)p(B)p(F)$$

$$= p(G=0|F=1,B=1)p(B=1)P(F=1) + p(G=0|F=0,B=1)p(B=1)p(F=0)$$
$$+ p(G=0|F=1,B=0)p(B=0)p(F=1) + p(G=0|F=0,B=0)p(B=0)p(F=0)$$

$$= 0.2\times0.9\times0.9 + 0.8\times0.9\times0.1 + 0.8\times0.1\times0.9 + 0.9\times0.1\times0.1 = 0.315$$

$p(D=1|G=1)=0.9 \;;\; p(D=0|G=0)=0.9$

$$p(D=0|F=0) = p(D=0|G=0)\cdot p(G=0|F=0) = 0.9\times\sum_{B\in 0,1} p(G=0|B,F=0)p(B)$$

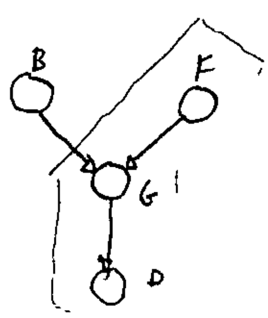
$$= 0.9\times\left(p(G=0|B=0,F=0)\cdot p(B=0) + p(G=0|B=1,F=0)p(B=1)\right)$$

$$= 0.9\times(0.9\times0.1 + 0.8\times0.9) = \boxed{0.729}$$

$$p(D=0|F=0,B=0) = p(D=0|G=0)\,p(G=0|F=0,B=0)\cdot p(B=0)$$

$$= 0.9\times0.9\times0.1 = \boxed{0.081}$$

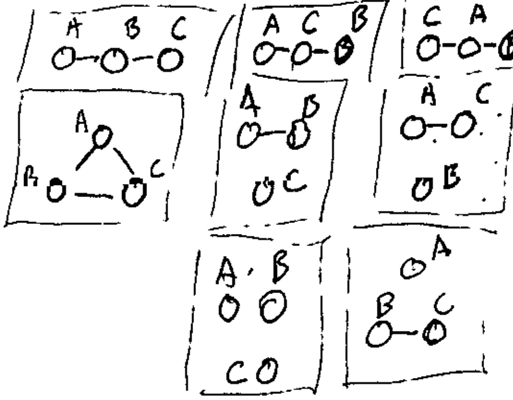
72.9%



0.081 < 0.729 because
We are choosing specific
conditions of joint
Probability

8.12 $2^{M(M-1)/2}$ Triangular Matrix $= D(D+1)/2$; if $\{1 \text{ or } 0\}$, then 2-cases with n comb. $2^{D(D+1)/2}$

$M=3$



$\begin{bmatrix}1&1&0\\0&1&1\\0&0&1\end{bmatrix}$ $\begin{bmatrix}1&0&1\\0&1&1\\0&0&1\end{bmatrix}$ $\begin{bmatrix}1&1&1\\0&1&0\\0&0&1\end{bmatrix}$

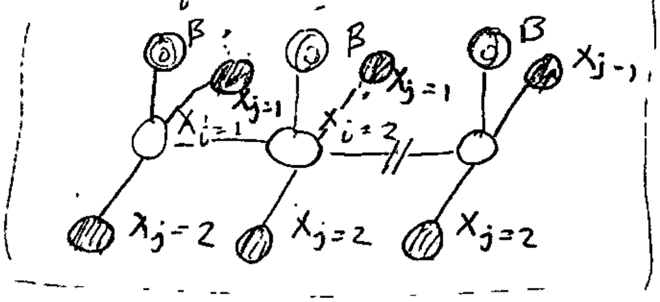
$\begin{bmatrix}1&1&1\\0&1&1\\0&0&1\end{bmatrix}$ $\begin{bmatrix}1&1&0\\0&1&0\\0&0&1\end{bmatrix}$ $\begin{bmatrix}1&0&1\\0&1&0\\0&0&1\end{bmatrix}$

$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$ $\begin{bmatrix}1&0&0\\0&1&1\\0&0&1\end{bmatrix}$

8.13 $E(x,y) = h\sum_i x_i - \beta\sum_{\{i,j\}} x_i x_j - \eta\sum x_i y_i$

which defines $p(x,y) = \frac{1}{Z}\exp\{-E(x,y)\}$

$\Delta E = E(x_{j=2}, y) - E(x_{j=1}, y) = \beta\sum_i x_i (x_2 - x_1)$; x



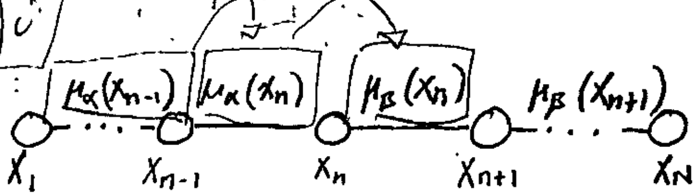
8.14 $\beta = h = 0$

$$\frac{dp(x,y)}{dE(x,y)} = 0 = -\frac{E(x,y)}{Z} \cdot e^{\{-E(x,y)\}}$$

$$0 = \left[0\cdot\sum x_i - 0\cdot\sum \cancel{x_i} x_j - \eta \sum \cancel{x_i} y_i\right] e^{-\eta\sum x_i y_i}$$

$$\boxed{0 = \sum x_i y_i}$$

8.15 Prove $p(x_{n-1}, x_n)$



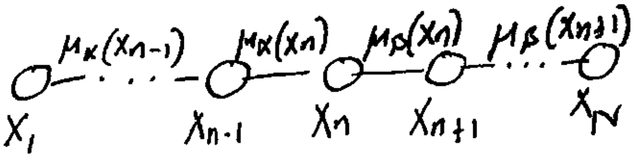
$= \frac{1}{Z}\mu_\alpha(x_{n-1})\psi_{n-1,n}(x_{n-1}, x_n)\mu_\beta(x_n)$

otherwise, $p(x_n) = \frac{1}{Z}\mu_\alpha(x_n)\mu_\beta(x_n)$

$= \frac{1}{Z}\sum \psi_{n-1,n}(x_{n-1}, x_n)\mu_\alpha(x_{n-1}) \cdot \mu_\beta(x_n)$

$$\boxed{= \frac{1}{Z}\mu_\alpha(x_{n-1})\sum \psi_{n-1,n}(x_{n-1}, x_n)\mu_\beta(x_n)}$$

8.16. $p(x_n|x_N)$



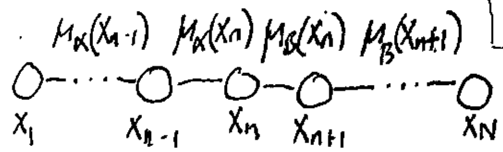
Message Passing Algorithm: $p(x) = \frac{1}{Z} \psi_{1,2}(x_1, x_2)\psi_{2,3}(x_2, x_3)\cdots\psi_{N-1,N}(x_{N-1}, x_N)$

Marginal - $p(x_n) = \sum_{x_1}\cdots\sum_{x_{n-1}}\sum_{x_{n+1}}\cdots\sum_{x_N} p(x)$

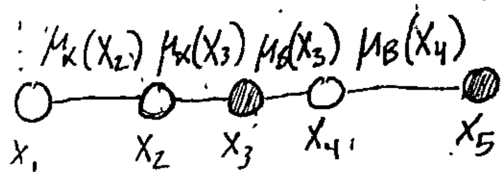
$$= \frac{1}{Z}\underbrace{\left[\sum_{x_{n-1}}\psi_{n-1,n}(x_{n-1},x_n)\cdots\left[\sum_{x_2}\psi_{2,3}(x_2,x_3)\left[\sum_{x_1}\psi_{1,2}(x_1,x_2)\right]\right]\right]}_{\mu_\alpha(x_n)}$$

$$\underbrace{\left[\sum_{x_{n+1}}\psi_{n,n+1}(x_n,x_{n+1})\cdots\left[\sum \psi_{N-1,N}(x_{N-1},x_N)\right]\cdots\right]}_{\mu_\beta(x_n)}$$

8.17.



N=5



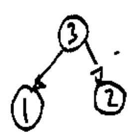
Prove $x_2 \perp\!\!\!\perp x_5 | x_3 = p(x_2, x_5|x_3) = p(x_2|x_3) \cdot p(x_5|x_3)$

$$\boxed{= \mu_\alpha(x_3)\cdot\mu_\beta(x_3)\cdot\mu_\beta(x_4)}$$

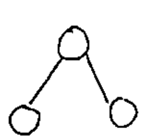
Show $p(x_2|x_3, x_5) = p(x_2|x_3) \cdot p(x_5)/Z$

$$= \frac{\mu_\alpha(x_3)\cdot\mu_\beta(x_3)\cdot\cancel{\mu_\beta(x_4)}}{[\mu_\alpha(x_3)+\mu_\beta(x_3)]\cancel{\mu_\beta(x_4)}} = \boxed{\frac{\mu_\alpha(x_3)\cdot\mu_\beta(x_3)}{[\mu_\alpha(x_3)+\mu_\beta(x_3)]}}$$

8.18



$p(1,2|3) = p(1|3)p(2|3)$ $\left\|\ \frac{p(1,2|3)}{(1+3)(2+3)} = \frac{p(1|3)p(2|3)}{(1+3)(2+3)}\ \right\|$ $D(D+1)/2$



$1 \perp\!\!\!\perp 2|3 = (1 \perp\!\!\!\perp 3)(2 \perp\!\!\!\perp 3)$ $\left\|\ \frac{(1 \perp\!\!\!\perp 2|3)}{(1+3)(2+3)} = \frac{(1 \perp\!\!\!\perp 3)(2 \perp\!\!\!\perp 3)}{(1+3)(2+3)}\ \right\|$

8.19 Sum-product Algorithm : $p(x) = \sum_{x \setminus x} p(X) = \prod_{s \in ne(x)} F_s(x, X_s)$

$$(8.54)\quad p(X_n) = \sum_{X_1} \cdots \sum_{X_{n-1}} \sum_{X_{n+1}} \cdots \sum_{X_N} p(X)$$

$$= \sum_{X_1} \cdots \sum_{X_{n-1}} \sum_{X_{n+1}} \cdots \sum_{X_n} \frac{\psi(X_1, X_N)}{Z}$$

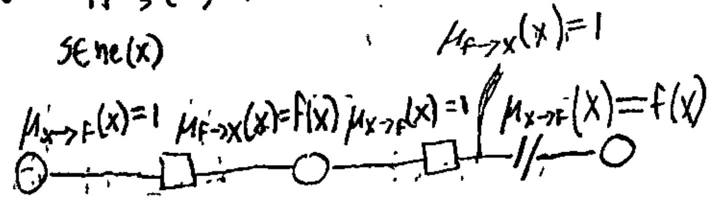
$$= \sum_{X_1} \cdots \sum_{X_{n-1}} \sum_{X_{n+1}} \cdots \sum_{X_n} \frac{\psi(X_1, X_2)\psi(X_2, X_3) \cdots \psi(X_{n-1}, X_n)}{Z}$$

$$(8.55)\quad \boxed{= \left[\sum_{X_{n-1}} \psi(X_{n-1}, X_n) \cdots \sum_{X_2} \psi_{2,3}(X_2, X_3) \sum_{X_1} \psi_{1,2}(X_1, X_2)\right]}$$

$$(8.56)\quad \boxed{\left[\sum_{X_{n+1}} \psi_{n,n+1}(X_n, X_{n+1}) \cdots \sum_{X_N} \psi_{N-1,N}(X_{N-1}, X_n)\right]}$$

$$(8.57)\quad \boxed{= \frac{1}{Z} \mu_\alpha(X_n) \mu_\beta(X_n)}$$

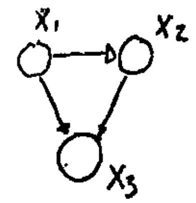
8.20 Sum-product Algorithm : $p(X) = \sum_{x \setminus x} p(X) = \prod_{s \in ne(x)} F_s(X, X_s)$

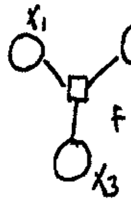


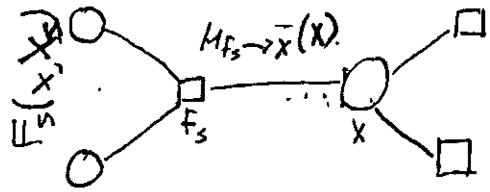
8.21 Sum - Product Algorithm: $P(X) = \sum_{x \setminus x} P(X) = \prod_{s \in ne(x)} F_s(X, X_s)$

$$= \prod_{s \in ne(x)} F_s(X_s) \left[\sum_{x} \cdots \sum_{X_{n-1}} \sum_{X_{n+1}} \cdots \sum_{X_n} p(x)\right]$$

$$\boxed{= F_s(X_s) \prod_{s \in ne(X)} \mu_{X_i \to F_s}(X_i)}$$

8.23 





(8.61) $p(x) = \sum_{x \setminus x} p(x)$

(8.62) $p(x) = \prod_{s \in ne(x)} F_s(x, X_s)$

$p(x) = \sum_{x \setminus x} \prod_{s \in ne(x)} F_s(x, X_s) = \prod_{s \in ne(x)} \sum_{x \setminus x} F_s(x, X_s)$

$= \prod_{s \in ne(x)} \mu_{f_s \to x}(x)$

8.24 $p(x) = \prod_{s \in ne(x)} \left[\sum_{X_s} F_s(x, X_s) \right] = \prod_{s \in ne(x)} \left[\sum_{x_1} \cdots \sum_{x_M} f_s(x, x_1, \ldots, x_M) \cdot \mu_{x \to f_s}(x_m) \right]$

$$\boxed{= f_s(X_s) \prod_{s \in ne(x)} \mu_{x_i \to f_s}(x_i)}$$

8.25 (8.86) $\tilde{p}(x_2) = \mu_{f_a \to x_2}(x_2) \cdot \mu_{f_b \to x_2}(x_2) \cdot \mu_{f_c \to x_2}(x_2)$

$= \left[\sum_{x_1} f_a(x_1, x_2) \right] \left[\sum_{x_3} f_b(x_2, x_3) \right] \left[\sum_{x_4} f_c(x_2, x_4) \right]$

$= \sum_{x_1} \sum_{x_2} \sum_{x_4} f_a(x_1, x_2) f_b(x_2, x_3) f_c(x_2, x_4)$

$= \sum_{x_1} \sum_{x_3} \sum_{x_4} \tilde{p}(x)$

$$\boxed{\tilde{p}(x_1) = \sum_{\cancel{x_1}} f_a(\cancel{x_2}, x_1) \cdot \sum_{x_2} f_b(x_3, x_2) \sum_{x_2} f_c(x_4, x_2)}$$

$$\boxed{\tilde{p}(x_3) = \sum_{x_3} f_a(x_2, x_3) \sum_{x_2} f_b(x_1, x_2) \sum_{x_2} f_c(x_4, x_2)}$$

$$\boxed{\tilde{p}(x_1, x_2) = \sum_{x_2} f_b(x_3, x_2) \sum_{x_4} f_c(x_4, x_2)}$$

$$f(\mu_K) = E[x|\mu_K] = \sum_{n=1}^{N}\sum_{K=1}^{K}\left[\int \mu_K x_n^2 \, d\mu_K - 2\int x_n \mu_K^2 \, d\mu_K + \int \mu_K^3 \, d\mu_K\right]$$

$$= \sum_{n=1}^{N}\sum_{K=1}^{K}\left[\frac{x_n \mu_K^2}{2} - \frac{2}{3} x_n \mu_K^8 + \frac{\mu_K^4}{4}\right]$$

$$= \sum_{n=1}^{N}\sum_{K=1}^{K}\left[\frac{x_n}{2} - \frac{2}{3}\mu_K\right]\mu_K^2 + \mu_K^4$$

$$\frac{dJ}{d\mu_K} = -2\sum_{n=1}^{N}\sum_{K=1}^{K} r_{nK} \|x_n - \mu_K\| = 0$$

$$\cancel{-\sum_{K=1}^{K-1} r_{nK} \|x_n - \mu_K\| - \mu_K}$$

$$\sum\sum \|x_n - \mu_K\| = 0 \; ; \; -\sum\sum \|x_n - \mu_K\| - \mu_K = 0$$

$$\cancel{\mu_K = \sum_{K=1}^{K-1} r_{nK} \|x_n - \mu_K\|}$$

$$\boxed{\mu_K' = \mu_K - \sum\sum \|x_n - \mu_K\|}$$

$$\boxed{\cancel{\mu_K = \mu_K + r_{nK}\|x_n - \mu_K\|}}$$

9.3. $p(z) = \prod_{K=1}^{K} \pi_K^{z_K}$ $\quad p(x|z) = \prod_{K=1}^{K} N(x|\mu_K, \Sigma_K)^{z_K}$ $\quad p(x) = \sum p(z)p(x|z) = \sum \pi_K N(x|\mu_K, \Sigma_K)$

$$p(x) = \sum \prod_{K=1}^{K} \pi_K^{z_K} \cdot \prod_{K=1}^{K} N(x|\mu_K, \Sigma_K)^{z_K} = \sum\left[\pi_K \cdot N(x|\mu_K, \Sigma_K)\right]^{\sum z_K}$$

$$\boxed{= \sum_{K} \pi_K \cdot N(x|\mu_K, \Sigma_K)}$$

9.4. EM For Gaussian Mixtures:

1. Initialize means (μ_K), covariances (Σ_K), and mixing coefficients π_K, and evaluate log likelihood.
2. E Step: Evaluate the responsibilities using current parameters

$$\gamma(z_{nK}) = \frac{\pi_K N(x_n|\mu_K, \Sigma)}{\sum \pi_j N(x_n|\mu_j, \Sigma_j)}$$

3. M Step: Re-estimate the parameters using current responsibilities:

$$\mu_K^{new} = \frac{1}{N_K}\sum_{n=1}^{N}\gamma(z_{nK})x_n \; ; \; \Sigma_K^{new} = \frac{1}{N_K}\sum_{n=1}^{N}\gamma(z_{nK})(x_n - \mu_K^{new})(x_n - \mu_K)^T$$

$\pi_k^{new} = \frac{N_k}{N}$ where $N_k = \sum_{n=1}^{N} \gamma(z_{nk})$

4. Evaluate the log likelihood

$$\ln p(X|\mu,\Sigma,\pi) = \sum_{n=1}^{N} \ln\left\{\sum_{k=1}^{K} \pi_k N(x_n|\mu_k,\Sigma)\right\}$$

The General EM Algorithm

1. Choose an initial setting for the parameters θ^{old}.
2. E Step: Evaluate $p(Z|X,\theta^{old})$
3. M Step: Evaluate θ^{new} given by

$$\theta^{new} = \arg\max_{\theta} Q(\theta,\theta^{old})$$

Where

$$Q(\theta,\theta^{old}) = \sum_{Z} p(Z|X,\theta^{old}) \ln p(X,Z|\theta)$$

4. Check for convergence criterion is not satisfied; then let $\theta^{old} \leftarrow \theta^{new}$

9.4 ① Initializes ② E Step: $p(Z|X,\theta^{old}) = p(Z|X)p(\theta^{old})$

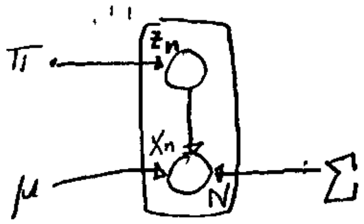
③ M Step $\ln p(Z|X,\theta^{old}) = \ln p(Z|X) \cdot p(\theta^{old})$

$$= \ln p(Z|X) + \ln p(\theta^{old})$$

$$Q(\theta,\theta^{old}) = \sum_{Z} p(Z|X,\theta^{old})[\ln p(Z|X) + \ln p(\theta^{old})]$$

$$\boxed{= \sum_{z=1} p(Z|X,\theta^{old}) \ln p(Z|X) + \ln p(\theta^{old})}$$

9.5



$$p(x) = \prod_{k=1}^{K} p(x_k|pa_k)$$

$$\boxed{p(Z|X,\mu,\Sigma,\pi) = \prod_{k=1}^{K} p(z_n|x_n,\mu,\Sigma,\pi)}$$

$$\boxed{p(x_a,x_b) = \sum_{x\backslash x} p(x_a) \sum_{x\backslash x} p(x_b)}$$

$(x \wedge y) \in \{0,1,2\}$ $p(\hat{x},\hat{y}) = e^{-xy}$

$p(x,\hat{y}) = -xe^{-xy} = 0; x=0$ | $p(\hat{x},y) = -ye^{-xy} = 0$

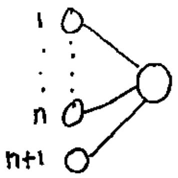
$\boxed{p(\hat{x},\hat{y}) = +xye^{-xy}}$

$p(x) = \mu_{1\to2}\mu_{2\to3}\mu_{3\to4}\mu_{4\to1}$

$= \sum_{x\backslash x} f_a \sum_{x\backslash x} f_b \sum_{x\backslash x} f_c \sum_{x\backslash x} f_d$:

$\boxed{= \sum_{x\backslash x_1} f_a \sum_{x\backslash x_1} f_b \sum_{x\backslash x_1} f_c \sum_{x\backslash x_1} f_d \left[\sum_{x} f_a \sum_{x} f_b \sum_{x} f_c \sum_{x} f_d\right]}$

1.



$p(x) = \prod_{x_n} \sum_{n} \mu_{x\to f(x)}(x)$; $p(x_n) + p(x_{n+1}) = \prod_{x_n} \sum \mu_{x\to f(x)}$

$\boxed{= \prod_{x_{n+1}} \sum_{x_{n+1}} \mu_{x\to}}$

for 9.

$r_{nk} = \begin{cases} 1 & \text{if } k = \arg\min_j \|x_n - \mu_j\|^2 \\ 0 & \text{otherwise} \end{cases}$

$\mu_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}}$; $\frac{dJ}{dr_{nk}} = \sum_{n=1}^{N} \sum_{k=1}^{K} \|x_n - \mu_k\|^2 + 2r_{nk}\|x_n - \mu_k\|$

$\boxed{= \sum_{n=1}^{N} \sum_{k=1}^{K} \|x_n - \mu_k\|^2 = 0}$; $x_n = \mu_k$

2 Robbins-Monro:

$$f(\theta) \equiv E[z|\theta] = \int z p(z|\theta) dz$$

Distortion Measure:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|x_n - \mu_k\|^2$$

$$f(\mu_k) = E[x_n|\mu_k] = \int \mu_k \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|x_n$$

9.6 1. Initialize μ_K and Σ_K and π_K

2. E Step:

$$\gamma(z_{nK}) = \frac{\pi_K N(x_n \mid \mu_K, \Sigma_K)}{\sum_{j=1}^{K} \pi_j N(x_n \mid \mu_j, \Sigma_j)}$$

3. M Step:

$$\Sigma_K^{New} = \frac{1}{N_K} \sum_{n=1}^{N} \gamma(z_{nK})(x-\mu_K)(x-\mu_K)^T$$

$$= \sum_{n=1}^{N} \gamma(z_{nK}) \mathrm{cov}(x_i, x_K)$$

$$\boxed{= \sum_{n=1}^{N} \gamma(z_{nK}) \Sigma_{iK}}$$

9.7

$$\ln p(X, Z \mid \mu, \Sigma, \pi) = \sum_{n=1}^{N} \sum_{K=1}^{K} z_{nK} \{\ln \pi_K + \ln N(x_n \mid \mu_K, \Sigma_K)\}$$

Group of Data Points:

$$\mu_K: \frac{dp(X, Z \mid \mu, \Sigma, \pi)}{d\mu} = \frac{d}{d\mu} \sum_{n=1}^{N} \sum_{K=1}^{K} z_{nK} \{\ln \pi + \ln N(x_n \mid \mu_K, \Sigma)\}$$

$$= \sum_{n=1}^{N} \sum_{K=1}^{K} z_{nK} \cdot \frac{\cancel{\#}(x-\mu)/\Sigma}{N(x_n \mid \mu_K, \Sigma)} = 0$$

$$\boxed{x_{nK} = \mu_{nK}}$$

$$\Sigma: \frac{dp(X, Z \mid \mu, \Sigma, \pi)}{d\Sigma} = \frac{d}{d\Sigma} \sum_{n=1}^{N} \sum_{K=1}^{K} z_{nK} \{\ln \pi + \ln N(x_n \mid \mu_K, \Sigma)\}$$

$$= \sum_{n=1}^{N} \sum_{K=1}^{K} z_{nK} \frac{\frac{1}{2}(x-\mu)^2/\Sigma_K^2}{N(x_n \mid \mu_K, \Sigma)} = 0$$

$$\boxed{\sum_{n=1}^{N} \left((x-\mu)^2/\Sigma_1^2 + (x-\mu)^2/\Sigma_2 + \cdots + (x-\mu)^2/\Sigma_N \right) = 0}$$

Mixing Coefficients: z_{nK}: $\frac{dp(x, Z \mid \mu, \Sigma, \pi)}{dz_{nK}} = \boxed{\sum_{n=1}^{N} \sum_{K=1}^{K} \frac{z_{nK}}{\pi}}$

9.8 $$E_z[\ln p(X,Z|\mu,\Sigma,\pi)] = \sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk})\{\ln\pi_k + \ln N(x_n|\mu_k,\Sigma_k)\}$$

$$\frac{dE_z[\ln p(X,Z|\mu,\Sigma,\pi)]}{d\mu} = \sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk})\frac{(x-\mu)/\Sigma}{N(x|\mu,\Sigma)} = 0$$

$$\mu_k = \frac{1}{\sum_{n=1}^{N}\gamma(z_{nk})}\sum_{n=1}^{N}\gamma(z_{nk})x$$

$$\boxed{\mu = \frac{1}{N}\sum_{n=1}^{N}\gamma(z_{nk})x}$$

9.9 $$\frac{dE_z[\ln p(X,Z|\mu,\Sigma,\pi)]}{d\Sigma} = \sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk})\left[\ln\pi_k + \ln|2\pi\Sigma|^{-\frac{1}{2}} - \tfrac{1}{2}(x-\mu)(x-\mu)^T/\Sigma\right]$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk})\left[\frac{1}{2\Sigma} + \frac{(x-\mu)(x-\mu)^T/\Sigma^2}{2}\right] = 0$$

$$\boxed{\Sigma = \frac{1}{N}\sum_{k=1}^{N}(x-\mu)(x-\mu)^T}$$

$$\frac{dE_z[\ln p(X,Z|\mu,\Sigma,\pi)]}{d\pi} = \frac{d}{d\pi}\left[\sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk})\{\ln\pi_k + \ln N(x_n|\mu_k,\Sigma)\} + \lambda\left(\sum\pi_k - 1\right)\right]$$

$$= \frac{d}{d\pi}\left[\sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk})[\ln\pi_k N(x_n|\mu_k,\Sigma)] + \lambda\left(\sum\pi_k - 1\right)\right]$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk})\cdot\frac{N(x_n\cancel{|\mu_k},\Sigma)}{\pi N(x_n\cancel{|\mu_k},\Sigma)} + \lambda = 0$$

$$\boxed{\lambda = -\sum\gamma(z_{nk}) = -N}$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K}\frac{\gamma(z_{nk})}{\pi} - N = 0 \;;\; \boxed{\pi = \frac{N_k}{N}}$$

9.10 $p(x) = \sum_{k=1}^{K} \pi_k \, p(X|k) \qquad x = (X_a, X_b)$

Show that the conditional density: $p(X_b|X_a)$ is a gauss-mixture

$$p(X) = p(X_a, X_b) = \sum_{k=1}^{K} \pi_k \, p(X_a, X_b | k)$$

Baysian Theorem: $P(Y|X) = \dfrac{P(X|Y)\,p(Y)}{p(X)}$

$$p(X_b|X_a) = \frac{p(X_a|X_b)\,p(X_b)}{p(X_a)}$$

$$\boxed{= \frac{\sum \pi_k \, p(X_a|X_b, k) \sum \pi_k \, p(X_b|k)}{\sum \pi_k \, p(X_a|k)}}$$

$$\boxed{\pi_{k(b|a)} = \frac{\pi_{k(a|b)} \cdot \pi_{k(a)}}{\pi_{k(b)}}}$$

9.11. $E_z[\ln p(X, Z | \mu, \Sigma, \pi)] = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \{\ln \pi_k + \ln N(x_n | \mu_k, \Sigma_k)\}$

$$p(X|\mu_k, \Sigma_k) = \frac{1}{(2\pi\epsilon)^{1/2}} \exp\left\{-\frac{1}{2\epsilon} \|X - \mu\|^2\right\}$$

$$\gamma(z_{nk}) = \frac{\pi_k \exp\{-\|x_n - \mu_k\|^2 / 2\epsilon\}}{\sum_j \pi_j \exp\{-\|x_n - \mu_j\|^2 / 2\epsilon\}}$$

$$\lim_{\epsilon \to 0} E_z[\ln p(X, Z | \mu, \Sigma, \pi)] = \lim_{\epsilon \to 0} \sum \sum \gamma(z_{nk}) \{\ln \pi_k + \ln N(x_n | \mu_k, \Sigma_k)\}$$

$$\boxed{= -\frac{1}{2\epsilon} \sum \sum r_{nk} \|x_n - \mu_n\|^2 + \sum \sum r_{nk} \ln \pi}$$

Huh

$$\boxed{\tilde{=} -\frac{1}{2} \sum \sum r_{nk} \|x_n - \mu_n\|^2}$$

111

66%

9.12 $p(x) = \sum_{k=1}^{K} \pi_k p(x|k)$ Denote μ_k and Σ_k

$E[x] = \int \pi_k p(x|k)\, x\, dx = \pi_k \int \prod_{i=1}^{D} \mu_i^{x_i} (1-\mu_i)^{(1-x_i)} x\, dx$

$\cancel{= \pi_k \prod_{i=1}^{D} \int \mu^{x_i} \mu (1-\mu_i)^{(1-x_i)} \mu^{-1} x\, dx}$

$\cancel{= \pi_k \prod_{i=1}^{D} \int \mu_i^{x_i+1} (1-\mu_i)^{1-(x_i+1)} x\, dx}$

$\cancel{= \pi_k \prod_{i=1}^{D} \int \mu_i^{x_i+1} (1-\mu_i)^{1} (1-\mu_i)^{-(x_i+1)} x\, dx}$ $\quad b = x_i + 1$

$\cancel{= \pi_k \prod_{i=1}^{D} \int (1-\mu_i) \left(\frac{\mu}{1-\mu_i}\right)^{x_i+1} x\, dx}$ $\quad u = x \quad dv = \left(\frac{\mu}{1-\mu}\right)^{x_i+1}$

$du = 1 \quad v = \frac{\left(\frac{\mu}{1-\mu}\right)^{x_i+2}}{x_i+2}$

$\cancel{= \pi_k \prod_{i=1}^{D}}$

$= \pi_k \prod_{i=1}^{D} \int \mu_i^{x_i-1} \mu_i (1-\mu_i)^{(1-x_i)} \mu$ $\quad \frac{1-\mu_i+\mu}{(1-\mu_i)^2}$ $\quad uv - \int v\, du$

$\int_a^b (x-a)^{m-1} (b-x)^{n-1} dx = (b-a)^{m+n-1} \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}$

$= \pi_k \prod_{i=1}^{D} \int \mu_i^{x_i} (1-\mu_i)^{(1-x_i)} x\, dx = \pi_k \prod \sum \mu_i^{x_i} (1-\mu_i)^{1-x_i}$ $\quad 1 - x_i = (x+1)$

$= \pi_k \prod_{}^{D} \left[\mu^{1} (1-\mu)^{1-1} + \mu^{2} (1-\mu)^{1-2}\right] = \pi_k \prod^{D} \left[\mu(0) + \frac{\mu^2}{1-\mu}\right]$ $\quad 1 - x_i = n - 1$

$= \pi_k \left[\mu^0 (1-\mu)^{1-0} + \mu^1 (1-\mu)^{1-1}\right] = \pi_k \left[(1-\mu) + \mu\right] = \pi_k$

9.13 $E[X] = \frac{1}{N}\sum_{n=1}^{N} X_n \equiv \overline{X}$; $\mu_K = \hat{\mu}$ for $K = 1, \dots, K$.

The EM Algorithm:

1. Initialize μ_K, Σ_K, and π_K
2. E Step: Evaluate responsibilities. $\gamma(z_{nK}) = \frac{\sum z_{nK} [\pi_K p(X_n|\mu_K)]^{z_{nK}}}{\sum [\pi_j p(X_n|\mu_j)]^{z_{nj}}}$

$$= \frac{\pi_K \cdot p(X_n|\mu_K)}{\sum_{j=1}^{K} \pi_j p(X_n|\mu_j)}$$

3. M Step: Re-estimate parameters $\mu_K = \frac{1}{N_K}\sum_{n=1}^{N} \gamma(z_{nK}) X_n$

$\Sigma_K = /$

$\pi_K = \frac{N_K}{N}$

Prove one iteration converges means of Bernoulli Distribution:

1. Initialize μ_K, and π_K
2. Evaluate Responsibilities $\gamma(z_{nK}) = \frac{\pi_K p(X_n|\mu_K)}{\sum_{j=1}^{K} \pi_j p(X_n|\mu_j)}$
3. $\mu_1 = E[X] = \frac{1}{N}\sum_{n=1}^{N} X_n \equiv \overline{X}$; $\mu_2 = \frac{1}{N}\sum_{n=1}^{N} \frac{\cancel{\pi_K p(X_n|\mu_2)}}{\cancel{\sum_{j=1}^{K} \pi_j p(X_n|\mu_2)}} X_n$

1 iteration

$$\boxed{= \frac{1}{N}\sum_{n=1}^{N} \overline{X_n}}$$

9.14 $p(x|z,\mu) = \prod_{k=1}^{K} p(x|\mu_k)^{z_k}$

$p(z|\pi) = \prod_{k=1}^{K} \pi_k^{z_k}$

$$p(x|\mu,\pi) = p(x|z,\mu)p(z|\pi) = \prod_{k=1}^{K} p(x|\mu_k)^{z_k} \cdot \prod_{k=1}^{K} \pi_k^{z_k}$$

$$= \prod_{k=1}^{K} [\pi_k \cdot p(x|\mu_k)]^{z_k}$$

$$\boxed{= \sum \pi_k p(x|\mu_k)}$$

9.15 $$E_z[\ln p(X,Z|\mu,\pi)] = \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \left\{ \ln \pi_k + \sum_{i=1}^{D} [x_{ni} \ln \mu_{ki} + (1-x_{ni}) \ln(1-\mu_{ki})] \right\}$$

$$\frac{dE_z}{d\mu} = \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \left\{ \cancel{\ln \pi_k} + \sum_{i=1}^{D} \frac{x_{ni}}{\mu_{ki}} - \frac{(1-x_{ni})}{1-\mu_{ki}} \right\} = 0$$

$$\sum_{i=1}^{D} \frac{x_{ni}(1-\mu_{ki})}{\mu_{ki}(1-\mu_{ki})} - \frac{(1-x_{ni})\mu_{ki}}{\mu_{ki}(1-\mu_{ki})} = 0$$

$$x_{ni} - x_{ni}\mu_{ki} - (\mu_{ki} - x_{ni}\mu_{ki}) = 0$$

$$\boxed{x_{ni} = \mu_{ki}}$$

9.16
$$E_z[\ln p(X,Z|\mu,\pi)] = \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \left\{ \ln \pi + \sum_{i=1}^{D} [x_{ni} \ln \mu_{ki} + (1-x_{ni}) \ln(1-\mu_{ki})] \right\} + \lambda\left\{ \sum \pi_k - 1 \right\}$$

$$\frac{dE_z}{d\pi_k} = \sum_{n=1}^{N}\sum_{k=1}^{K} \frac{\gamma(z_{nk})}{\pi_k} + \lambda = 0$$

$$\frac{N_k}{\pi_k} + \lambda = 0 \; ; \; \lambda = -N \; ; \; \boxed{\frac{N_k}{N} = \lambda}$$

9.17. $0 \le p(x_n|\mu_k) \le 1$

$$\lim_{x^+\to 0} \ln p(X,Z|\mu,\pi) = \lim_{x^+\to 0} \sum\sum \gamma(z_{nk}) \{ \ln \pi_k + \sum x_{ni} \ln \mu + (1-x_{ni}) \ln(1-\mu_{ki}) \}$$

$$\lim_{x^-\to 0} \ln p(X,Z|\mu,\pi) = \lim_{x^-\to 0} \sum\sum \gamma(z_{nk}) \{ \ln \pi_k + \sum x_{ni} \ln \mu + (1-x_{ni}) \ln(1-\mu) \}$$

$$\lim_{x^+\to 1} \ln p(X,Z|\mu,\pi) = \lim_{x^+\to 1} \sum\sum \gamma(z_{nk}) \{ \ln \pi_k + \sum x_{ni} \ln \mu + (1-x_{ni}) \ln(1-\mu_i) \}$$

$$\lim_{x^-\to 1} \ln p(X,Z|\mu,\pi) = \lim_{x^-\to 1} \sum\sum \gamma(z_{nk}) \{ \ln \pi_k + \sum x_{ni} \ln \mu + (1-x_{ni}) \ln(1-\mu_i) \}$$

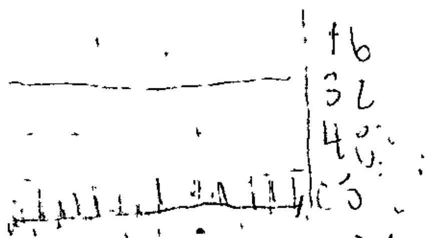
Flag: Label

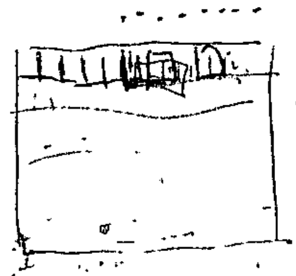
m_Pos

buffer

bytesToWrite

row = (m_Pos % 128) / 8





cnt?

row = ((m_Pos % 128) % 8 / 2)

col = ((m_Pos % 128) % 16) / 2

col = ((m_Pos % 128) % 16) / 2

Line?

$$= \sum \sum \gamma(z_{nk}) [\ln \pi + \ln(1-\mu_k)]$$
$$= \sum \sum \gamma(z_{nk}) [\ln \pi + \ln(1-\mu_k)]$$
$$= \sum \sum \gamma(z_{nk}) [\ln \pi \mu]$$
$$= \sum \sum \gamma(z_{nk}) [\ln \pi \mu]$$

9.18. $p(x|\mu) = \prod_{i=1}^{D} \mu_i^{x_i} (1-\mu_i)^{(1-x_i)}$; $p(\mu_k | a_k, b_k)$; $\mathrm{Beta}(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1} (1-\mu)^{b-1}$

$\mathrm{Dir}(\pi|\alpha) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_K)} \prod_{k=1}^{K} \pi_k^{\alpha_k - 1}$

$p(\mu_k | a_k, b_k) = p(\mu | a, b) \cdot p(\pi | \alpha) =$

EM Algorithm:

1. Initialize μ_k, a_k, b_k and π_k, α
2. Expectation $\gamma(z_{nk}) = \frac{\pi_k \, p(\mu_k | a_k, b_k)}{\sum_{j=1}^{K} \pi_j \, p(\mu_k | a_k, b_k)}$
3. Maximization : $\mu_k = \frac{1}{N_k} \sum \gamma(z_{nk}) x_n$; $a_k = \frac{1}{N} (a_k + 1)$; $b_k = \frac{1}{N} (b_k - 1)$
4. R Check convergence

Organic Chemistry:

9.19 $\sum_{ij} X_{ij} = 1$; $p(x) = \sum_{k=1}^{K} \pi_k \, p(x|\mu_k)$ where $p(X|\mu_k) = \prod_{i=1}^{D} \prod_{j=1}^{M} \mu_{kij}^{x_{ij}}$

$0 \le \mu_{kij} \le 1$; $\sum_j \mu_{kij} = 1$

Given $\{X_n\}$ where $n = 1, \ldots, N$

Derive the E and M Step of the EM

Algorithm for optimizing π_k and μ_{kij}

EM Algorithm:

1. Initialize μ_{kij}
2. Expectation Step: $\gamma(z_{nk}) = E[z_{nk}] = \dfrac{\pi_k \, p(x_n|\mu_k)}{\sum_{j=1}^{k} \pi_j \, p(x_n|\mu_j)}$
3. M Step: Re-estimate $N_k = \sum_{n=1}^{N} \gamma(z_{nk})$

$$\bar{X}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \cdot X$$

4. Estimation converges after one-step.

9.20 Maximize $E[\ln p(t, w|\alpha, \beta)] = \frac{M}{2} \ln\left(\frac{\alpha}{2\pi}\right) - \frac{\alpha}{2} E[w^T w] + \frac{N}{2} \ln\left(\frac{\beta}{2\pi}\right) - \frac{\beta}{2} \sum_{n=1}^{N} E[(t_n - w^T \phi_n)^2]$

$$\frac{dE[\ln p(t, w|\alpha, \beta)]}{d\alpha} = \frac{M \cdot \cancel{2\pi}}{\cancel{2} \cdot \alpha \cdot \cancel{2\pi}} - \frac{E[w^T w]}{\cancel{2}} = 0 \; ; \; \boxed{\alpha = \frac{M}{E[w^T w]}}$$

9.21 $$\frac{dE[\ln p(t, w|\alpha, \beta)]}{d\beta} = \frac{N \, \cancel{2\pi}}{2\beta \cdot \cancel{2\pi}} - \frac{1}{2} \sum_{n=1}^{N} E[(t_n - w^T \phi_n)^2] = 0$$

$$\frac{N}{\sum_{n=1}^{N} E[(t_n - w^T \phi_n)^2]} = \beta$$

9.22 $E_w[\ln p(t|X,w,\beta)p(w|\alpha)] = E\left[\ln \prod_{n=1}^{N} p(t_n|x_n,w,\beta^{-1}) \prod_{i=1}^{M} N(w_i|0,\alpha_i^{-1})\right]$

$= E\left[\ln \prod_{n=1}^{N}\left(\frac{\beta}{2\pi}\right)^{M/2} e^{-\frac{\beta}{2}(t-wx)^2}\left(\frac{\alpha}{2\pi}\right)^{N/2} e^{-\frac{\alpha}{2}(w^Tw)}\right]$

$= \frac{M}{2}\ln\left(\frac{\beta}{2\pi}\right) - \frac{\beta}{2}\sum_{i=1}^{N} E[(t-wx)^2] + \frac{N}{2}\ln\left(\frac{\alpha}{2\pi}\right) - \frac{\alpha}{2}\sum_{i=1}^{M} E[w^Tw]$

$$\frac{dE[\ln p(t|X,w,\beta)p(w|\alpha)]}{d\alpha} = \frac{N\cdot\cancel{2\pi}}{\cancel{2}\cdot\alpha\cdot\cancel{2\pi}} - \frac{1}{\cancel{2}}\sum_{i=1}^{M} E[w^Tw] = 0$$

$$\alpha = \frac{N}{\sum_{i=1}^{M} E[w^Tw]} = \frac{N}{m_N^T m_N + Tr(S_n)}$$

$$\alpha_i^{new} \boxed{= \frac{1}{m_i^2 + \Sigma_{ii}}}$$

$$\frac{dE[\ln p(t|X,w,\beta)p(w|\alpha)]}{d\beta} = \frac{M\cdot\cancel{2\pi}}{\cancel{2}\cdot\beta\cdot\cancel{2\pi}} - \frac{1}{2}\sum_{i=1}^{N} E[(t-wx)^2] = 0$$

$$\beta^{-1} = \frac{\sum E[(t-wx)^2]}{M}$$

$$\boxed{(\beta^{new})^{-1} = \frac{\|t-wx\|^2 + \beta^{-1}\Sigma_{ii}}{M}}$$

9.23 $\alpha_i^{new} = \frac{\gamma_i}{m_i^2}$ $\quad (\beta^{new})^{-1} = \frac{\|t-\phi m\|^2}{N-\sum_i \gamma_i} \longrightarrow N = \|t-\phi m\|^2 + \beta^{-1}\sum_i \gamma_i$

$\alpha(m_N^T m_N + Tr(S_n)) = M$

$\alpha_i^{new} = \frac{1}{m_i^2 + \Sigma_{ii}}$ $\quad (\beta^{new})^{-1} = \frac{\|t-\phi m_N\|^2 + \beta^{-1}\sum_i \gamma_i}{N}$

9.24 $\ln p(X|\theta) = L(q,\theta) + KL(q\|p) = \sum q(z)\ln\frac{p(X,z|\theta)}{q(z)} - \sum q(z)\ln\frac{p(z|X,\theta)}{q(z)}$

$= \sum q(z)\cdot\ln\frac{p(X,z|\theta)}{\cancel{q(z)}}\cdot\frac{\cancel{q(z)}}{p(z|X,\theta)} = \boxed{\ln\frac{p(X|\theta)\,p(z|\theta)}{p(z|X,\theta)}}$

Study Hall

Year Book Signing

(11:45 - 12:[illegible] am

9.25 Prove lower bound $L(q,\theta)$ is $L(q,\theta)=\sum q(z)\ln\frac{p(X,z|\theta)}{q(z)}$ with $q(z)=p(z|X,\theta^{old})$

$$L(q,\theta)=\sum_z q(z)\ln\frac{p(X,z|\theta)}{q(z)}$$

$$=\sum_z p(z|X,\theta^{old})\cdot\ln\frac{p(X,z|\theta)}{p(z|X,\theta^{old})}$$

$$=\sum_z p(z|X,\theta^{old})\left[\ln p(X,z|\theta)-\ln p(z|X,\theta^{old})\right]$$

$$\boxed{=\sum_z p(z|X,\theta^{old})\ln p(X,z|\theta)-\sum_z p(z|X,\theta^{old})\ln p(z|X,\theta^{old})}$$

$$\ln p(X|\theta)=L(q,\theta)+KL(q\|p)=\sum_z q(z)\ln\frac{p(X,z|\theta)}{q(z)}-\sum q(z)\ln\frac{p(z|X,\theta)}{q(z)}$$

$$=\sum q(z)\left[\ln p(X,z|\theta)-\cancel{\ln q(z)}\right]-\sum q(z)\left[\ln p(z|X,\theta)-\cancel{\ln q(z)}\right]$$

$$=\sum q(z)\left[\ln p(X,z|\theta)-\ln p(z|X,\theta)\right]$$

$$=\sum p(z|X,\theta^{old})\left[\ln p(X,z|\theta)-\ln p(z|X,\theta)\right]$$

$$\boxed{=\sum p(z|X,\theta^{old})\ln p(X,z|\theta)-\sum p(z|X,\theta^{old})\ln p(z|X,\theta^{old})}$$

9.26 $\mu_K=\frac{1}{N_K}\sum_{n=1}^{N}\gamma(z_{nk})\cdot x_n$; $N_K=\sum_{n=1}^{N}\gamma(z_{nk})$

$$N_K^{new}-N_K^{old}=\gamma^{new}(z_{nk})-\gamma^{old}(z_{nk})$$

$$\mu_K^{new}-\mu_K^{old}=\frac{1}{N_K^{new}}\gamma(z_{nk})^{new}x_n^{new}-\frac{1}{N_K^{old}}\gamma(z_{nk})^{old}x_{old}\quad \boxed{N_K^{new}=N_K^{old}+\gamma^{new}(z_{nk})-\gamma^{old}(z_{nk})}$$

$$\boxed{\mu_K^{new}=\mu_K^{old}+\frac{(\gamma(z_{nk})^{new}-\gamma^{old})}{N_K^{new}}(x_m-\mu_K)}$$

9.27 Restimate Σ_{ii} and π_k

$$\boxed{\Sigma_{ii} = \frac{1}{N_K} \sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_K^{new})(x_n - \mu_K^{new})^T}$$

$$\boxed{\pi_K^{new} = \frac{N_k}{N}}$$

Chapter 10:

10.1 Verify, ~~$\ln p(X, Z) = \ln p(X)p(Z) = \ln p(X) + \ln p(Z)$~~

~~$\ln p(Z|X)$~~ $= \ln p(Z, X)/p(X) = \ln p(Z,X) - \ln p(X)$

~~$\stackrel{?}{=} L(q, \theta) + KL(q \| p)$~~

$$\ln p(Z|X) = \sum_i \ln p(Z, X)/p(X) = \sum \ln p(Z, X) - \sum \ln p(X)$$

$$= \int \ln p(Z, X) - \int \ln p(X) + \int \ln q(z) + \int \ln q(z)$$

$$\ln p(X) = \int \ln \frac{p(Z, X)}{q(Z)} - \int \ln \frac{p(Z|X)}{q(Z)}$$

$$\boxed{L(q, \theta) \cdot q(z) = \frac{\ln p(Z, X)}{q(z)}} \qquad KL(q \| p) \cdot q(z) = \frac{\ln p(z|x)}{q(z)}$$

$$= \int q(z) \ln \frac{p(Z, X)}{q(z)} + \int q(z) \ln \frac{p(Z|X)}{q(z)}$$

$$\boxed{\ln p(X) = L(q) + KL(q \| p)}$$

10.2 $E[z_1] = m_1$; $E[z_2] = m_2$; $q^*(z_1) = N(z_1 | m_1, \Lambda_{11}^{-1}) = \left(\frac{\Lambda}{(2\pi)}\right)^{1/2} \exp\left(\frac{-\Lambda}{2}(z_1 - m_1)^T (z_1 - m_1)\right)$

$$q^*(z_2) = N(z_2 | m_2, \Lambda_{22}^{-1}) = \left(\frac{\Lambda}{2\pi}\right)^{1/2} \exp\left(\frac{-\Lambda}{2}(z_2 - m_2)^T (z_2 - m_2)\right)$$

$$q(z) = q^*(z_1) q^*(z_2)$$

$$q(z) = \frac{(\Lambda_{11}\Lambda_{22})^{1/2}}{2\pi} \exp^{-\frac{\Lambda_{11}}{2}(z_1-m_1)^T(z_1-m_1)} \exp^{-\frac{\Lambda_{22}}{2}(z_2-m_2)^T(z_2-m_2)}$$

$$\boxed{\Rightarrow E[z_1] \cdot E[z_2] = m_1 \cdot m_2 = [\mu_1 - \Lambda_{11}^{-1}\Lambda_{12}(E[z_2]-\mu_2)][\mu_2 - \Lambda_{22}^{-1}\Lambda_{21}(E[z_1]-\mu_1)]}$$

$$E[z_2] = \int \left(\frac{\Lambda_{22}}{2\pi}\right)^{1/2} \exp^{\frac{-\Lambda_{22}}{2\pi}(z_2-m_2)^T(z_2-m_2)} \cdot z_2 \cdot d(z_2-m_2) \cdots$$

$$= z_2 = \mu_2$$

10.3 $q(z) = \prod_{i=1}^{M} q_i(z_i)$; Kullback-Leibler Divergence:

$$KL(q\|p) = -\int p(z)\left[\sum_{i=1}^{M} \ln q_i(z_i)\right] dz + \text{const}$$

$$= -\int \left[p(z)\ln q_j(z_{ij}) + p(z)\sum_{i\neq j}\ln q_i(z_i)\right] dz + \text{const}$$

$$= -\int p(z)\ln q_j(z_j)\, dz + \text{const} = -\int \ln q_j(z_j)\left[\int p(z)\prod_{i\neq j} dz\right] dz + \text{const}$$

$$= -\int F_j(z_j)\ln q_j(z_{ij})\, dz + \text{const} = -\int F_j(z_j)\ln q_j(z)\, dz + \lambda\left(\int q_j(z_j)\, dz - 1\right)$$

$$d/dq = 0 \rightarrow = \frac{-F_j(z_j)}{q_j(z_{ij})} + \lambda = 0; \quad \lambda = \frac{\int \ln q(z) \int p(z) \prod dz\, dz}{q_j(z_{ij})} = 1; \quad q_j = (F_j)$$

$$q_j^*(z_j) = \int p(z) \prod_{i\neq j} dz_i = p(z)$$

10.4. $q(x) = N(x|\mu, \Sigma)$

$$KL(q\|p) = -\int q(z) \ln\left\{\frac{p(z)}{q(z)}\right\} dz$$

$$= -\int N(x|\mu,\Sigma) \ln \frac{p(z)}{N(x|\mu,\Sigma)}\, dz$$

$$\mu: \quad \frac{dKL(q\|p)}{d\mu} = \frac{d}{d\mu}\left[-\int N(x|\mu,\Sigma)\ln\frac{p(z)}{N(x|\mu,\Sigma)}\, dz\right]$$

$$= \frac{d}{d\mu}\left[-\int N(x|\mu,\Sigma)\ln p(z)\, dz + \int N(x|\mu,\Sigma)\ln N(x|\mu,\Sigma)\, dz\right]$$

$$= \frac{+1}{(2\pi\Sigma)^{1/2}} \frac{(x-\mu)}{\Sigma} e^{-\frac{1}{2\Sigma}(x-\mu)^2} \cdot \int \ln p(z)\, dz$$

$$+ \frac{d}{d\mu}\left[\int N(x|\mu,\Sigma) \cdot -\frac{1}{2\Sigma}(x-\mu)^2 dz + \int N(x|\mu,\Sigma)\ln\frac{1}{2\pi\Sigma}\, dz\right]$$

$$= \frac{-1}{(2\pi\Sigma)^{1/2}} \frac{(x-\mu)}{\Sigma} e^{-\frac{1}{2\Sigma}(x-\mu)^2} \int p(z)dz$$

$$+ \left[\int N(x|\mu,\Sigma)\cdot\frac{-1}{2\Sigma}(x-\mu)^2 dz + \int N(x|\mu,\Sigma)\cdot \ln\frac{1}{2\pi\Sigma}dz\right]'$$

$$= \frac{-1}{(2\pi\Sigma)^{1/2}} \cdot \frac{(x-\mu)}{\Sigma} e^{-\frac{1}{2\Sigma}(x-\mu)^2} \int p(z)dz$$

$$+ \left[\frac{1}{(2\pi\Sigma)^{1/2}}\cdot\frac{1}{2\Sigma}\int N(x|\mu,\Sigma)(x-\mu)^2 dz + \frac{1}{2\pi\Sigma}\int N(x|\mu,\Sigma)dz\right]'$$

$$= \frac{-(x-\mu)}{(2\pi)^{1/2}\cdot\Sigma} e^{-\frac{1}{2\pi}(x-\mu)^2} \int p(z)dz + \left[\frac{1}{2\Sigma\sqrt{\pi}}\int \cancel{\exp}^{-\frac{1}{2\Sigma}(x-\mu)^2}\cdot(x-\mu)^2 dz + \frac{1}{2\pi\Sigma}\int \cancel{\exp}^{-\frac{1}{2\Sigma}(x-\mu)^2} dz\right]' = 0$$

$$\cancel{= \frac{-(x-\mu)}{(2\pi)^{1/2}\cdot\Sigma}\int p(z)dz + \left[\frac{1}{2\Sigma\sqrt{\pi}}\int (x-\mu)^2 dz + \frac{1}{2\pi\Sigma}\int dz\right]' = 0}$$

$$= \frac{-1}{(2\pi\Sigma)^{1/2}} \cdot \frac{(x-\mu)}{\Sigma}\cdot e^{-\frac{1}{2\Sigma}(x-\mu)^2} \int p(z)dz + \frac{1}{(2\pi\Sigma)^{1/2}}\cdot\frac{1}{2\Sigma}\int e^{-\frac{1}{2\Sigma}(x-\mu)^2}\cdot(x-\mu)^2 dz$$

$$+ \frac{1}{2\pi\Sigma}\int e^{-\frac{1}{2\Sigma}(x-\mu)^2} dz = 0$$

$$= \frac{-1}{\cancel{(2\pi\Sigma)^{1/2}}} \frac{\cancel{(x-\mu)}}{\cancel{\Sigma}} \cancel{e}^{-\frac{1}{2\Sigma}(x-\mu)^2} \int p(z)dz + \frac{1}{\cancel{(2\pi\Sigma)^{1/2}}}\cdot\frac{1}{2\cancel{\Sigma}}\left[\frac{+1}{\Sigma}\cancel{(x-\mu)}\, \cancel{e}^{-\frac{1}{2\Sigma}(x-\mu)^2}\cdot(x-\mu)^2 - 2\,\cancel{e}^{-\frac{1}{2\Sigma}(x-\mu)^2}\cdot\cancel{(x-\mu)}\right]$$

$$+ \frac{\ln(2\pi/\Sigma)\,\cancel{(x-\mu)}}{(2\pi\cancel{\Sigma})^{1/2}\, 2\cancel{\Sigma}}\int \cancel{e}^{-\frac{1}{2\Sigma}(x-\mu)^2} dz = 0$$

$$\cancel{\int p(z)dz + \frac{1}{2}\left[\frac{(x-\mu)^2}{\Sigma} - 2\right] + \ln(2\pi/\Sigma) = 0}$$

$$= \frac{+1}{(2\pi\Sigma)^{1/2}}\frac{\ }{\Sigma}\cdot\int p(z)dz - \frac{(x-\mu)}{\Sigma} = 0 \;;\; \frac{\int p(z)dz}{\cancel{(2\pi\Sigma)^{1/2}}} = (x-\mu)\;;$$

$\boxed{\text{Back of Book}}$

10.5 $q(z,\theta) = q_z(z)q_\theta(\theta)$; $q_\theta(\theta) = \delta(\theta-\theta_0)$; Show the EM algorithm optimizes $q_z(z)$ and M maximizes log likelihood of θ respect to θ_0.

$$KL(q\|p) := -\iint q(\theta)q(z)\ln\frac{p(z,\theta|X)}{q(\theta)q(z)}\,dz\,d\theta = -\int q(\theta)q(z)\ln\frac{p(z,\theta|X)}{q(\theta)}dz + \text{const}$$

$$= -\int q(\theta)\ln\frac{p(z|\theta,X)p(\theta|X)}{q(z)}dz + \text{const} = \boxed{-\int q(z)\ln\frac{p(z|\theta,X)}{q(z)}dz + \text{const}}$$

If mimized, then Expectation step.

$$\text{const} := +\int q(z)\int q(\theta)\ln\frac{p(X,\theta,z)}{q(\theta)q(z)}\,dz\,d\theta$$

$$= \int q(\theta)E_{q,z}[\ln p(X,\theta,z)]\,d\theta - \int q(\theta)\ln q(\theta) + \text{const}$$

$$\boxed{\text{Maximation Step: } E_{q,z}[\ln p(X,\theta,z)]}$$

10.6. $D_\alpha(p\|q) = \frac{4}{1-\alpha^2}\left(1-\int p(x)^{(1+\alpha)/2}\cdot q(x)^{(1-\alpha)/2}dx\right)$; $-\infty < \alpha < \infty$

$p^\epsilon = \exp(\epsilon\ln p) = 1 + \epsilon\ln p + O(\epsilon^2)$. then $\lim_{\epsilon\to 0} D_\alpha(p\|q)$

$$\lim_{\epsilon\to 0} D_\alpha(p\|q) = \frac{4}{1-\alpha^2}\left(1-\int p^\epsilon\cdot q(x)^{(1-\alpha)/2}dx\right) \quad \epsilon = (1+\alpha)/2 ; \; 2\epsilon - 1 = \alpha$$

$$= \frac{4}{1-\alpha^2}\left(1-\int q(x)^{(1-\alpha)/2}dx - \epsilon\int \ln p\cdot q(x)^{(1-\alpha)/4}dx - \text{const}\, O(\epsilon^2)\right)$$

$$= \frac{4}{1-\alpha^2}\left(1-\int q(x)^{(1-\alpha)/2}dx - 0 - 0\right)$$

$$= \frac{4}{1-\alpha^2}\left(1-\int q(x)^{(1-\alpha)/2}dx\right) = \frac{4}{1-\alpha^2}\left(1-\int q(x)^{(1-2\epsilon+1)/2}dx\right)$$

$$\boxed{= \frac{4}{1-\alpha^2}\left(1-\int q(x)dx\right) ; \text{ if } \alpha = 1 \; D_\alpha(p\|q) = 0}$$

$$\lim_{\epsilon\to 0} D_\alpha(p\|q) = \frac{4}{1-\alpha^2}\left(1-\int q(x)^{(1-\alpha)/2}dx - \epsilon\int \ln p\cdot q(x)^{(1-\alpha)/2}dx - \text{const}\, O(\epsilon^2)\right)$$

$$\boxed{= \frac{4}{1-\alpha^2}\left(1-\int q(x)dx\right) ; \text{ if } \alpha = -1 ; \text{ then } D_\alpha(q\|p) = 0}$$

10.7 $p(D|\mu,\tau) = \left(\frac{\tau}{2\pi}\right)^{N/2} \exp\left\{-\frac{\tau}{2}\sum_{n=1}^{N}(x_n-\mu)^2\right\}$

$p(\mu|\tau) = N(\mu|\mu_0, (\lambda_0\tau)^{-1})$; $p(\tau) = Gam(\tau|a_0, b_0)$

Prove $q_\mu(\mu)$ is a Gaussian of the form $N(\mu|\mu_N, \lambda_N^{-1})$

with $\mu_N = \frac{\lambda_0\mu_0 + N\bar{x}}{\lambda_0 + N}$ and $\lambda_N = (\lambda_0 + N)E[\tau]$

$$\ln q_\mu^*(\mu) = E_\tau[\ln p(D|\mu,\tau) + \ln p(\mu|\tau)] + const$$

$$= -\frac{E[\tau]}{2}\left\{\lambda_0(\mu-\mu_0)^2 + \sum_{n=1}^{N}(x-\mu)^2\right\} + const$$

$$= -\frac{E[\tau]}{2}\left\{\lambda_0(\mu-\mu_0)^2 + \sum_{n=1}^{N}(x^2-\mu x+\mu^2)\right\} + const.$$

$$= -\frac{E[\tau]}{2}\left\{\lambda(\mu^2-\mu\mu_0+\mu_0^2) + \sum_{n=1}^{N}(x-\mu x+\mu^2)\right\} + const$$

$$= -\frac{E[\tau]}{2}\left\{\lambda\mu^2 - \lambda\mu\mu_0 + N\mu + Nx + N\mu^2\right\} + const$$

$$= -\frac{E[\tau]}{2}\left\{\lambda\mu^2 - (\lambda\mu_0 + N)\mu + N\mu^2 + Nx\right\} + const$$

$$= -\frac{E[\tau]}{2}\left\{(\lambda+N)\mu^2 - (\lambda\mu_0+N)\mu + Nx\right\} + const$$

$$= -\frac{E[\tau]}{2}\left\{(\lambda+N)\left[\mu^2 - \frac{(\lambda\mu_0+N)}{(\lambda+N)}\mu\right] + Nx\right\} + const$$

$$= -\frac{E[\tau]}{2}\left\{(\lambda+N)\left(\mu - \frac{(\lambda\mu_0+N)}{2(\lambda+N)}\right)^2 - \left[\frac{(\lambda\mu_0+N)}{2(\lambda+N)}\right]^2 + \frac{Nx}{(\lambda+N)}\right\} + const$$

$$= \underbrace{-\frac{E[\tau](\lambda+N)}{2}}_{\boxed{\lambda_N}}\left[\mu - \underbrace{\frac{(\lambda\mu_0+N)}{2(\lambda+N)}}_{\boxed{\mu_n}}\right]^2 + const$$

$$\ln q_\tau^*(\tau) = E_\mu[\ln p(D|\mu,\tau) + \ln p(\mu|\tau)] + \ln p(\tau) + const.$$

$$= (a_0-1)\ln\tau - b_0\tau + \frac{N}{2}\ln\tau - \frac{\tau}{2}E_\mu\left[\sum_{n=1}^{N}(x_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right] + const$$

$$Gam(\tau|a_n,b_n) = \frac{1}{\Gamma(a)} b^a \tau^{a-1} \exp(-b\tau)$$

$$\ln q^*(\pi,\mu,\Lambda) = \ln p(\pi) + \sum_{k=1}^{K}\ln p(\mu_k,\Lambda_k) + E_z[\ln p(z|\pi)] + \sum_{k=1}^{K}\sum_{n=1}^{N}\mathbb{E}[z_{nk}]\ln N(x_n|\mu_k,\Lambda_k^{-1}) + \text{const}$$

$$\cong q(\pi)q(\mu,\Lambda_k) \cong q(\pi)\prod_{k=1}^{K} q(\mu_k,\Lambda_k)$$

$$Dir(\mu|\alpha) = C(\alpha)\prod_{k=1}^{K}\mu_k^{\alpha_k-1} \quad ; \sum_{k=1}^{K}\mu_k = 1 \quad ; \quad 0<\mu_k<1$$

if $\ln \rho_{nk} = \mathbb{E}[\ln \pi_k] + \frac{1}{2}\mathbb{E}[\ln|\Lambda_k|] - \frac{D}{2}\ln(2\pi) - \frac{1}{2}\mathbb{E}_{\mu,\Lambda}\left[(x-\mu)^T\Lambda_k(x-\mu)\right]$

and $\quad r_{nk} = \dfrac{\rho_{nk}}{\sum_{j=1}^{K}\rho_{nj}} \quad ; \quad \mathbb{E}[z_{nk}] = r_{nk}$

then $\ln q^*(\pi) = (\alpha_0 - 1)\sum_{k=1}^{K}\ln \pi_k + \sum_{k=1}^{K}\sum_{n=1}^{N} r_{nk}\ln \pi_k + \text{const}$

where $\alpha_k = \alpha_0 + N_k$

Practice : $\ln q^*(\pi) = \cancel{\alpha_0 \sum \ln \pi_k} - \cancel{\sum \ln \pi_k} + \sum \sum \mathbb{E}[z_{nk}] \cdot \ln \pi_k$

$$\boxed{= (\alpha - 1)\ln \pi_k + \sum\sum r_{nk}\ln \pi_k}$$

Practice: $q^*(\mu_k,\Lambda_k) = N(\mu_k|m_k,(\beta_k\Lambda_k)^{-1})\,W(\Lambda_k|W_k,\nu_k)$

Wishart Distribution

$$= \left(\frac{\beta\Lambda_k}{2\pi}\right)^{D/2} \exp^{-\frac{1}{2}(\mu_k-m_k)^T\beta\Lambda(\mu_k-m_k)} \cdot \left(\frac{\Lambda}{2\pi}\right)^{D/2} \exp^{-\frac{1}{2}(\Lambda-W_k)^T V_k(\Lambda-W_k)}$$

$$\beta_k = \beta_0 + N_k ; \quad m_k = \frac{1}{\beta_k}(\beta_0 m_0 + N_k\bar{x}_k)$$

$$W_k^{-1} = W_0^{-1} + N_k S_k + \frac{\beta_0 N_k}{\beta_0 + N_k}(x_k - m_0)(x_k - m_0)^T$$

$$V_k = V_0 + N_k$$

$$\boxed{\begin{aligned} &= \left(\frac{\beta_0+N_k}{2\pi}\right)^{D/2} \exp^{-\frac{1}{2}(\mu_k - \frac{1}{\beta_k}(\beta_0 m_0 + N_k\bar{x}))^T(\beta_0+N_k)\Lambda(\mu_k - \frac{1}{\beta_k}(\beta_0 m_0 + N_k\bar{x}))} \\ &\cdot \left(\frac{\Lambda}{2\pi}\right)^{D/2} \exp^{-\frac{1}{2}\left(\Lambda - \frac{1}{\frac{1}{W_0} + N_k S_k + \frac{\beta_0 N}{\beta_0+N_k}(x_k-m_0)(x_k-m_0)}\right)^T (V_0+N_k)\cdot\left(\Lambda - \frac{1}{\frac{1}{W_0}+N_k S_k + \frac{\beta_0 N}{\beta_0+N_k}()}\right)} \end{aligned}}$$

10.11 $L_m = \sum_n \sum_z q(z|m) q(m) \ln \left\{ \frac{p(z, X, m)}{q(z|m) q(m)} \right\}$

10.10 Derive $\ln p(X) = L_m - \sum \sum q(z|m) q(m) \ln \left\{ \frac{p(z, m | X)}{q(z|m) q(m)} \right\}$

$\ln p(X) = L(q) + KL(q \| p)$

$= \int q(z) \ln \left\{ \frac{p(X, z)}{q(z)} \right\} dz - \int q(z) \ln \left\{ \frac{p(z|X)}{q(z)} \right\}$

if $KL(q \| p) = - \sum \sum q(z|m) q(m) \ln \left\{ \frac{p(z, m | X)}{q(z|m) q(m)} \right\}$

then $\boxed{L(q|m) = \sum \sum q(z|m) q(m) \ln \left\{ \frac{p(z, m, X)}{q(z|m) q(m)} \right\}}$

10.12 $p(X, z, \pi, \mu, \Lambda) = p(X | z, \mu, \Lambda) p(z | \pi) p(\pi) p(\mu | \Lambda) p(\Lambda)$

if $\ln q_j^*(z_j) = \mathbb{E}_{i \neq j} [\ln p(X, z)] + const$

$\ln q^*(z) = \mathbb{E}_{\pi, \mu, \Lambda} [\ln p(X, z, \pi, \mu, \Lambda)] + const$

$= \mathbb{E}_{\pi, \mu, \Lambda} [\ln p(X | z, \mu, \Lambda) p(z | \pi) p(\pi) p(\mu | \Lambda) p(\Lambda)] + const$

$= \mathbb{E}_\pi [\ln p(z | \pi)] + \mathbb{E}_\pi [\ln p(\pi)] + \mathbb{E}_{\mu, \Lambda} [\ln p(\mu | \Lambda)] + \mathbb{E}_\Lambda [\ln p(\Lambda)]$

$+ \mathbb{E}_{\mu, \Lambda} [\ln p(X | z, \mu, \Lambda)]$

$= \mathbb{E} [\ln p(z | \pi)] + \mathbb{E}_{\mu, \Lambda} [\ln p(X | z, \mu, \Lambda)] + const$

$= \sum_{n=1}^{N} \sum_{k=1}^{K} \Big[\mathbb{E}_\pi [\ln \pi_k] + \frac{1}{2} \mathbb{E}_\Lambda [\ln |\Lambda|] - \frac{D}{2} \ln 2\pi - \frac{1}{2} \mathbb{E}_{\mu, \Lambda} [(x_n - \mu_k)^T \Lambda (x_n - \mu_k)] \Big\} b$

$+ const$

$\Rightarrow \boxed{q^*(z) = b^z}$

10.13 $\ln q^*(\pi, \mu, \Lambda) = \ln p(\pi) + \sum_{k=1}^{k} \ln p(\mu_k, \Lambda_k) + E_z [\ln p(z | \pi)] + \sum_{k=1}^{K} \sum_{n=1}^{N} \mathbb{E}[z_{nk}] \ln N(x_n | \mu_k, \Lambda_k^{-1}) + const$

Derive $q^*(\mu_k, \Lambda_k) = N(\mu_k | m_k, (\beta_k \Lambda_k)^{-1}) W(\Lambda_k | W_k, \nu_k)$

$\beta_k = \beta_0 + N_k$; $m_k = \frac{1}{\beta_k} (\beta_0 m_0 + N_k \bar{x}_k)$; $W^{-1} = W_0^{-1} + N_k S + \frac{\beta_0 N_k}{\beta_0 + N_k} (x - m_0)(x - m_0)^T$

10.13 $\ln q^*(\mu_K, \Lambda_K) = \ln N(\mu_K | m_0, (\beta_0 \Lambda_K)^{-1}) W(\Lambda_K | W_0, \nu_0) = \ln N(\mu_K | m_0, (\beta_0 \Lambda_K)^{-1}) + \ln W(\Lambda_K | W_0, \nu_0)$

$+ \sum_{n=1}^{N} \mathbb{E}[z_{nK}] \ln N(X | \mu_K, \Lambda_K^{-1}) + \text{const} \qquad + \sum_{n=1}^{N} \mathbb{E}[z_{nK}] \ln N(X | \mu_K, \Lambda_K^{-1}) + \text{const}$

$$= -\frac{\beta_0}{2}(\mu_K - m_0)^T \Lambda (\mu_K - m_0) + \frac{1}{2}\ln|\Lambda| - \frac{1}{2}\mathrm{Tr}(\Lambda_K W_0^{-1}) + \frac{(\nu_0 - D - 1)}{2}\ln|\Lambda|$$

$$- \frac{1}{2}\sum_{n=1}^{N} \mathbb{E}[z_{nK}](x_n - \mu_K)^T \Lambda_K (x_n - \mu_K) - \frac{1}{2}\left(\sum_{n=1}^{N} \mathbb{E}[z_{nK}]\right)\ln|\Lambda_K| + \text{const}$$

$$\ln q^*(\mu_K, \Lambda_K) = \ln q^*(\mu_K | \Lambda_K) + \ln q^*(\Lambda_K)$$

$$= -\frac{1}{2}\mu_K^T \left[\beta_0 + \frac{1}{2}\sum_{n=1}^{N} \mathbb{E}[z_{nK}]\right]\Lambda_K \mu_K - \mu_K^T \left[m_0 \beta_0 + \sum_{n=1}^{N} \mathbb{E}[z_{nK}] x\right]\Lambda_K$$

$$+ \text{const}$$

$$= -\frac{1}{2}\mu_K^T [\beta_0 + N_K] \Lambda_K \mu_K + \mu_K^T \Lambda_K [\beta_0 m_0 + N_K \bar{x}_K] + \text{const}$$

Therefore, $\ln q^*(\Lambda_K) = \ln q^*(\mu_K, \Lambda_K) - \ln q^*(\mu_K | \Lambda_K)$

$$= -\frac{\beta_0}{2}(\mu_K - m_0)^T \Lambda (\mu_K - m_0) + \frac{1}{2}\ln|\Lambda| - \frac{1}{2}\mathrm{Tr}(\Lambda_K W_0^{-1}) + \frac{(\nu_0 - D - 1)}{2}\ln|\Lambda|$$

$$- \frac{1}{2}\sum_{n=1}^{N} \mathbb{E}[z_{nK}](x_n - \mu_K)^T \Lambda_K (x_n - \mu_K) - \frac{1}{2}\left(\sum_{n=1}^{N} \mathbb{E}[z_{nK}]\right)\ln|\Lambda_K|$$

$$+ \frac{1}{2}\mu_K^T \left[\beta_0 + \frac{1}{2}\sum_{n=1}^{N} \mathbb{E}[z_{nK}]\right]\Lambda_K \mu_K - \mu_K^T \left[m_0 \beta_0 + \sum_{n=1}^{N} \mathbb{E}[z_{nK}] x\right]\Lambda_K$$

$$= \frac{(\nu_K - D - 1)}{2}\ln|\Lambda_K| - \frac{1}{2}\mathrm{Tr}(\Lambda_K W_0^{-1}) + \frac{1}{2}\ln|\Lambda|$$

$$+ \frac{\beta_0}{2} m_0^T \Lambda m_0 - \frac{1}{2}\sum_{n=1}^{N} \mathbb{E}[z_{nK}] x_n^T x_n$$

$$= \frac{(\nu_K - D - 1)}{2}\ln|\Lambda_K| - \frac{1}{2}\mathrm{Tr}(\Lambda_K W_K^{-1}) - \frac{1}{2}\sum_{n=1}^{N} \mathbb{E}[z_{nK}](x_n - \bar{x}_K)^T (x_n - \bar{x}_K)$$

$W(\Lambda | W, \nu_K)$

1.) $E_{cell} = E_{cathode} - E_{anode}$

$F_2(g) + 2e^- \rightarrow 2F^-(aq)$ } cathode $= 2.87V$

$Ni(s) \rightarrow Ni^{2+}(aq) + 2e^-$ } anode $= 0.28V$

$E_{cell} = 2.87V - 0.28V$

$= \boxed{2.59V}$

ashley

Ashley

Ashley

Ashley

Ashley

Ashley

$$= \iint (x_n-\mu_k)^T \Lambda (x_n-\mu_k) \exp\left\{-\tfrac{1}{2}\left[(\mu_k-m_k)^T \beta\Lambda(\mu_k-m_k)\right]\right\} \cdot \exp\left\{-\tfrac{1}{2}\frac{(\Lambda-W)^T(\Lambda-W)}{\nu_k}\right\} d\mu\, d\Lambda$$

$$u=(x_n-\mu_k)^T\Lambda(x_n-\mu_k) \qquad dv=\exp\left\{-\tfrac{1}{2}(\mu_k-m_k)^T\beta\Lambda(\mu_k-m_k)\right\}$$

$$d\mu=-2[x_n+\mu_k]\Lambda$$

$$= \iint (x_n^T\Lambda x_n - 2x_n\Lambda\mu_k + \mu_k^T\Lambda\mu_k) \exp^{-\frac{1}{2}\mu_k\beta\Lambda\mu_k} \cdot \exp^{-\cancel{\mu_k\beta\Lambda m_k}} \cdot \exp^{-\frac{1}{2}m_k\beta\Lambda m_k} \cdot \exp^{-\frac{1}{2}\frac{(\Lambda-W)^T(\Lambda-W)}{\nu_k}}\, d\mu\, d\Lambda$$

$$= \iint \cancel{x_n^T\Lambda x_n \exp\left\{\frac{-1}{2\nu_k}(\Lambda-W)^T(\Lambda-W)\right\}} \exp\left\{\tfrac{1}{2}m_k\beta\Lambda m_k\right\} \cdot \cancel{\left[-2x_n\Lambda\right] \iint \mu_k \cdot e^{\frac{1}{2}\left[\mu_k\beta\Lambda\mu_k + 2\mu_k\beta\Lambda m_k\right]} d\mu}$$

$$+ \cancel{\Lambda \iint \mu_k^T\mu_k \exp^{-\frac{1}{2}\mu_k\beta\Lambda\mu_k} \cdot \exp^{-\mu_k\beta\Lambda m_k}\, d\mu}$$

$$\int e^{ax}dx = \frac{1}{a}e^{ax}; \quad \int x e^{ax}dx = \frac{e^{ax}}{a^2}(ax-1); \quad \int_0^\infty x^n e^{-x}dx = n!$$

$$\int_0^\infty x^n e^{-ax}dx = \begin{cases} \dfrac{\Gamma(n+1)}{a^{n+1}} & (n>-1,\ a>0) \\[2mm] \dfrac{n!}{a^{n+1}} & (n=0,1,2,\ldots\ a>0) \end{cases}$$

$$\int_0^\infty e^{-ax^2}dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}; \quad \int_{-\infty}^{\infty} e^{2bx-ax^2}dx = \sqrt{\frac{\pi}{a}}\, e^{b^2/a} \quad (a>0)$$

$$\int_0^\infty x^n e^{-ax^2}dx \begin{cases} \frac{1}{2}\Gamma\left(\frac{n+1}{2}\right)/a^{\frac{n+1}{2}} & (n>-1,\ a>0) \\[2mm] \dfrac{(2k-1)!!}{2^{k+1}a^k}\sqrt{\dfrac{\pi}{a}} & (n=2k,\ k,\ a>0) \\[2mm] \dfrac{k!}{2a^{k+1}} & (n=2k,\ a>0) \end{cases}$$

$$\boxed{\begin{aligned} &\int_{-\infty}^{\infty} x e^{-(ax^2+bx)}dx = \frac{\sqrt{\pi}\, b}{2a^{3/2}} \cdot e^{b^2/4a} \\ &\int_{-\infty}^{\infty} x^2 \cdot e^{-(ax^2+bx)} \cdot dx = \frac{\sqrt{\pi}(2a+b^2)}{4a^{5/2}}\, e^{b^2/2a} \end{aligned}}$$

$$\Gamma\left(\tfrac{1}{2}\right)=\sqrt{\pi}; \quad \Gamma(z+1)=z\Gamma(z).$$

$$\mathbb{E}_{\mu,\Lambda}\left[(x-\mu_k)^T\Lambda(x-\mu_k)\right] = D\beta^{-1} + \nu_k(x-m)^T W(x-m)$$

$$\ln\Lambda_k = \mathbb{E}[\ln\Lambda] = \sum_{i=1}^{D}\psi\left(\frac{\nu_k+1-i}{2}\right) + D\ln 2 + \ln W$$

$$\ln\pi \equiv \mathbb{E}[\ln\pi_k] = \psi(\alpha_k) - \psi(\hat{\alpha})$$

$$\int \mu_K^T \mu_K \cdot e^{-\frac{1}{2}(\mu_K - m_K)^T \beta\Lambda(\mu_K - m_K)} \cdot d\mu = e^{-\frac{1}{2} m_K \beta \Lambda m_K} \cdot \int \mu_K^T \mu_K \cdot e^{-\frac{1}{2}\mu_K^T \mu_K \beta\Lambda + \mu_K \cdot \beta\Lambda \cdot m_K} \cdot d\mu$$

$$= e^{-\frac{1}{2} m_K \beta\Lambda m_K} \cdot \frac{\sqrt{\pi}\left(2 \cdot \frac{1}{2}\beta\Lambda + (\beta\Lambda m_K)^2\right)}{4\left(\frac{1}{2}\beta\Lambda\right)^{5/2}} \cdot e^{(\beta\Lambda m_K)^2 / 4 \cdot \frac{1}{2}\beta\Lambda}$$

$$= \frac{\sqrt{\pi}\left(\beta\Lambda + m_K \beta\Lambda m_K\right)}{\frac{1}{\sqrt{32}}(\beta\Lambda)^{5/2}} \, e^{\frac{1}{2} m_K^T \beta\Lambda m_K}$$

$$\iint \left[x_n^T x_n e^{m_K \beta\Lambda m_K} - x_n \frac{\sqrt{\pi} m_K}{\sqrt{2}/2 \beta\Lambda} + \frac{\sqrt{\pi}(\beta\Lambda + m_K \beta\Lambda m_K)}{\sqrt{2}(\beta\Lambda)^{5/2}} \cdot e^{\frac{1}{2} m_K^T \beta\Lambda m_K} \right] \cdot e^{-\frac{1}{2}\frac{(\Lambda - W)^T(\Lambda - W)}{V_K}} d\Lambda$$

$$\int x_n^T x_n e^{m_K \beta\Lambda m_K} \cdot e^{-\frac{1}{2}\frac{(\Lambda - W)^T(\Lambda - W)}{V_K}} d\Lambda = \int x_n^T x_n \cdot e^{m_K \beta\Lambda m_K} \cdot e^{-\frac{1}{2}(\Lambda^T\Lambda - 2\Lambda W + W^T W)/V_K} \cdot d\Lambda = \int x_n^T x_n \cdot e^{-\frac{1}{2}\frac{\Lambda^T\Lambda}{V_K} + (m_K \beta m_K + \frac{W}{V_K})\Lambda + \frac{W^T W}{V_K}} \cdot d\Lambda$$

$$= x_n^T x_n \, e^{W^T W / V_K} \cdot \int e^{-\frac{1}{2}\frac{\Lambda^T\Lambda}{V_K} + (m\beta m_K + \frac{W}{V_K})\Lambda} \cdot d\Lambda \quad a = \frac{1}{2V_K} \, ; \, b = \left(m_K \beta m_K + \frac{W}{V_K}\right)\frac{1}{2}$$

$$2(m_K \beta m_K + W/V_K)^2 V_K$$

$$= x_n^T x_n \cdot e^{W^T W / V_K} \cdot \sqrt{\pi V_K} \cdot e$$

$$\int -x_n \frac{\sqrt{\pi}\, m_K}{\sqrt{2}/2 \cdot \beta \cdot \Lambda} \cdot e^{-\frac{1}{2}\frac{(\Lambda - W)^T(\Lambda - W)}{V_K}} \cdot d\Lambda = \frac{-x_n \sqrt{\pi}\, m_K}{\sqrt{2}/2 \cdot \beta} \int \frac{1}{\Lambda} \cdot e^{-\frac{1}{2}\frac{(\Lambda - W)^T(\Lambda - W)}{V_K}} \cdot d\Lambda =$$

$\boxed{\text{Not solved - order of integrations}}$

Lower order first

$$10.14 \quad = \iint (x_n-\mu_k)^T\Lambda(x_n-\mu_k)\cdot \exp\left\{-\frac{1}{2}(\mu_k-m_k)^T\beta\Lambda(\mu_k-m_k)\right\}\cdot \exp\left\{-\frac{1}{2}\frac{(\Lambda-W)^T(\Lambda-W)}{\nu_k}\right\} d\mu\, d\Lambda$$

$$= \int\left[\int (x_n-\mu_k)^T(x_n-\mu_k)\cdot \exp\left\{-\frac{1}{2}(\mu_k-m_k)^T\beta\Lambda(\mu_k-m_k)\right\}d\mu\right]\cdot \exp\left\{-\frac{1}{2}\frac{(\Lambda-W)^T(\Lambda-W)}{\nu_k}\right\}\Lambda\, d\Lambda$$

$$= \int\left[\int x_n^T x_n\, e^{-\frac{1}{2}(\mu_k-m_k)^T\beta\Lambda(\mu_k-m_k)}\,d\mu - 2x_n^T\int \mu_k\, e^{-\frac{1}{2}(\mu_k-m_k)^T\beta\Lambda(\mu_k-m_k)}\,d\mu + \int \mu_k^T\mu_k\, e^{-\frac{1}{2}(\mu_k-m_k)^T\beta\Lambda(\mu_k-m_k)}\,d\mu\right]\times\cdots$$

$$x_n^T x_n\, e^{+\frac{1}{2}m_k^T\beta\Lambda m_k}\int e^{-\frac{1}{2}[\mu_k^T\mu_k - 2\mu_k m_k]}\,d\mu = x_n^T x_n\, e^{\frac{1}{2}m_k^T\beta\Lambda m_k}\int e^{-\frac{1}{2}\mu_k^T\mu_k + \mu_k m_k}\,d\mu$$

$$= x_n^T x_n\, e^{\frac{1}{2}m_k^T\beta\Lambda m_k}\int_{-\infty}^{\infty} e^{2\cdot\frac{1}{2}\mu_k m_k - \frac{1}{2}\mu_k^T\mu_k}\,d\mu \qquad a=\frac{1}{2}\beta\Lambda;\ b=\frac{1}{2}m_k\beta\Lambda$$

$$= x_n^T x_n\cdot e^{\frac{1}{2}m_k^T\beta\Lambda m_k}\cdot\sqrt{\frac{2\pi}{\beta\Lambda}}\cdot e^{\frac{1}{2}m_k^T\beta\Lambda m_k}$$

$$= x_n^T x_n\cdot e^{m_k\beta\Lambda m_k}, \ldots$$

$$\cancel{-2x_n\int \mu_k\, e^{-\frac{1}{2}(\mu_k^T\mu_k - 2\mu_k m_k + m_k^T m_k)\beta\Lambda}\,d\mu = -2x_n\sqrt{\frac{\pi}{2}}(-m_k)\,\underbrace{\mathrm{erf}\left(\frac{m_k-\mu_k}{\sqrt{2}}\right)} + 2x_n e^{-\frac{1}{2}(m_k-\mu_k)^2}}$$

$$\cancel{\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\,dt}$$

$$\mathrm{erf}\left(\frac{m_k-\mu_k}{\sqrt{2}}\right) = \frac{2}{\sqrt{\pi}}\int_0^{(m_k-\mu_k)/\sqrt{2}} e^{-t^2}\,dt$$

$$\cancel{\int \mu_k^T\mu_k\, e^{-\frac{1}{2}(\mu_k-m_k)\beta\Lambda(\mu_k-m_k)}\,d\mu}$$

$$-2x_n\int \mu_k\, e^{-\frac{1}{2}(\mu_k^T\beta\Lambda\mu_k - 2\mu_k^T\beta\Lambda m_k + m_k^T\beta\Lambda m_k)}\,d\mu = -2x_n\cdot e^{\frac{m_k^T\beta\Lambda m_k}{2}}\cdot\int \mu_k\cdot e^{-\frac{1}{2}\beta\Lambda\mu_k^T\mu_k + \mu_k^T\beta\Lambda m_k}\,d\mu$$

$$= -2x_n\cdot e^{\frac{m_k^T\beta\Lambda m_k}{2}}\cdot\int \mu_k\cdot e^{-a\mu_k^T\mu_k + b\mu_k}\,d\mu$$

$$= -2x_n\cdot e^{\frac{m_k^T\beta\Lambda m_k}{2}}\cdot\frac{\sqrt{\pi}\cdot\beta\Lambda m_k}{4\left(\frac{1}{2}\beta\Lambda\right)^{3/2}}\cdot e^{(\beta\Lambda m_k)^2/4\left(\frac{1}{2}\beta\Lambda\right)}$$

$$= -2x_n\cdot e^{\frac{-m_k^T\beta\Lambda m_k}{2}}\cdot\frac{\sqrt{\pi}\,m_k}{4\sqrt{8\beta\Lambda}}\cdot e^{\beta\Lambda m_k^T m_k/2}$$

$$= -x_n\cdot\frac{\sqrt{\pi}\,m_k}{2\cdot\frac{\pi}{\sqrt{8}}\beta\Lambda}$$

$$\int x^2\,dx \qquad \frac{x^3}{3} \qquad \frac{x^2}{2}\cdot\frac{x^2}{2} \qquad \frac{x^3}{3} - \frac{x^4}{4}$$

10.14 $\mathbb{E}_{\mu,\Lambda}\left[(X_n-\mu_k)^T\Lambda(X_n-\mu_k)\right]=$

Likelihood x prior: $|\Lambda|^{N/2}\exp\left\{-\frac{1}{2}\left(\sum(x_i-\mu_i)^T\Lambda(x_i-\mu_i)\right)\right\}|\Lambda|^{(\nu_0-D-1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}(W_0^{-1}\Lambda)\right\}$

$$\times|\Lambda|^{1/2}\cdot\exp\left\{-\frac{K_0}{2}(\mu-\mu_0)^T\Lambda(\mu-\mu_0)\right\}$$

$$|\Lambda|^{1/2}\cdot|\Lambda|^{(\nu_0+N-D-1)/2}\exp\left\{-\frac{1}{2}\left((K_0+N)\mu^T\Lambda\mu-\mu^T\Lambda(K_0\mu_0+N\bar{x})-(K_0\mu_0+N\bar{x})\Lambda\mu\right.\right.$$
$$\left.\left.+K_0\mu_0^T\Lambda\mu+\sum x_i^T\Lambda x_i+\mathrm{tr}(W_0^{-1}\cdot\Lambda)\right)\right\}$$

1. $(K_0+N)\mu^T\Lambda\mu-\mu^T\Lambda(K_0\mu_0+Nx)-(K_0\mu_0+N\bar{x})^T\Lambda\mu$
$$+\frac{1}{K_0+N}(K_0\mu_0+N\bar{x})^T\Lambda(K_0\mu_0+N\bar{x})$$
$$-\frac{1}{K_0+N}(K_0\mu_0+Nx)^T\Lambda(K_0\mu_0+N\bar{x})$$
$$+K_0\mu_0^T\Lambda\mu_0+\sum x^T\Lambda x+\mathrm{tr}(W_0^{-1}\Lambda)$$

2. $(K_0+N)\left(\mu-\frac{K_0\mu+N\bar{x}}{K_0+N}\right)^T\Lambda\left(\mu-\frac{K_0\mu+N\bar{x}}{K_0+N}\right)$
$$-\frac{1}{K_0+N}(K_0\mu_0+Nx)^T\Lambda(K_0\mu_0+N\bar{x})$$
$$+K_0\mu_0^T\Lambda\mu_0+\sum x^T\Lambda x+\mathrm{tr}(W_0^{-1}\Lambda)$$

3. $(K_0+N)\left(\mu-\frac{K_0\mu+N\bar{x}}{K_0+N}\right)^T\Lambda\left(\mu-\frac{K_0\mu+N\bar{x}}{K_0+N}\right)$
$$-\frac{1}{K_0+N}(K_0\mu_0+N\bar{x})^T\Lambda(K_0\mu_0+N\bar{x})$$
$$+K_0\mu_0^T\Lambda\mu_0+\sum(x_i\Lambda x_i-\bar{x}_i\Lambda x_i+\bar{x}_i\Lambda x_i+\bar{x}_i\Lambda x_i)$$
$$+\mathrm{tr}(W_0^{-1}\Lambda)$$

4. $(K_0+N)\left(\mu-\frac{K_0\mu+Nx}{K_0+N}\right)^T\Lambda\left(\mu-\frac{K_0\mu+Nx}{K_0+N}\right)$
$$\mathrm{tr}\left(\frac{NK_0}{K_0+N}(\bar{x}-\mu_0)\Lambda(\bar{x}-\mu_0)\right)$$
$$\mathrm{tr}\left(\sum(x_i-\bar{x})^T\Lambda(x_i-\bar{x})\right)+\mathrm{tr}(W_0^{-1}\Lambda)$$

10.14 $q^*(\mu_K, \Lambda_K) = N(\mu_K | m_K, (\beta\Lambda)^{-1}) \cdot W(\Lambda_K | W_K, \nu_K)$

$E_{\mu_K, \Lambda_K}[(x_n - \mu_K)^T \Lambda_K (x_n - \mu_K)] = \int (x_n - \mu_K)^T \Lambda_K (x_n - \mu_K) N(\mu_K | m_K, (\beta\Lambda))^{-1} \cdot W(\Lambda_K | W_K, \nu_K) d$

$= \iint (x_n - \mu_K)^T \Lambda (x_n - \mu_K) \cdot \exp\{-\frac{1}{2}(\mu_K - m_K)^T \beta\Lambda (\mu_K - m_K)\} \exp\{\frac{-1}{2\nu_K}(\Lambda_K - W_K)^T (\Lambda_K - W_K)\} d\mu\, d\Lambda$

$- \iint (x_n - \mu_K)^T \Lambda (x_n - \mu_K) \exp\{\frac{1}{2}[(\mu_K - m_K)^T \beta\Lambda (\mu_K - m_K) + \frac{(\Lambda - W)^T (\Lambda - W)}{\nu_K}]\} d\mu\, d\Lambda$

$u = (x_n - \mu_K)^T \Lambda (x_n - \mu_K)$; $du = [-2x_n + 2\mu_K]\Lambda$; $dv = \exp\{\frac{-1}{2}(\mu_K - m_K)^T \beta\Lambda (\mu_K - m_K)\}$; $v =$

10.15 $\mathbb{E}[\pi_K] = \frac{\alpha_K}{\hat{\alpha}}$ if $\mathbb{E}[\pi_K] = \int q^*(\pi) \pi_K \cdot d\pi_K = \int Dir(\pi | \alpha) \pi_K d\pi_K$

$= \int C(\alpha) \cdot \prod_{K=1}^{K} \pi_K^{\alpha_K - 1} \cdot \pi_K\, d\pi_K$

$= \int C(\alpha) \prod_{K=1}^{K} \pi_K^{\alpha_K} d\pi_K$

$= \frac{C(\alpha)}{K\alpha_{+}} \cdot \pi_K^{\alpha_K + 1}$ if $\alpha_K = 0$; $\frac{\alpha_0}{K\alpha_0} = \boxed{\frac{\alpha_K + N_K}{K\alpha_0 + N}}$

10.16 $\mathcal{L} = \sum_Z \iiint q(Z, \pi, \mu, \Lambda) \ln\left\{\frac{p(X, Z, \pi, \mu, \Lambda)}{q(Z, \pi, \mu, \Lambda)}\right\} d\pi\, d\mu\, d\Lambda$

$= \mathbb{E}[\ln p(X, Z, \pi, \mu, \Lambda)] - \mathbb{E}[\ln q(Z, \pi, \mu, \Lambda)]$

$= \mathbb{E}[\ln p(X | Z, \mu, \Lambda)] + \mathbb{E}[\ln p(Z | \pi)] + \mathbb{E}[\ln p(\pi)] + \mathbb{E}[\ln p(\mu, \Lambda)]$

$- \mathbb{E}[\ln q(Z)] - \mathbb{E}[\ln q(\pi)] - \mathbb{E}[\ln q(\mu, \Lambda)]$

$\mathbb{E}[\ln p(X | Z, \mu, \Lambda)] = \iint \ln p(X | Z, \mu, \Lambda) \cdot q^*(\mu_K, \Lambda_K) d\mu_K$

$= \iint \ln \prod_{n=1}^{N} \prod_{K=1}^{K} N(x_n | \mu_K, \Lambda_K^{-1})^{Z_{nK}} \cdot N(\mu_K | m_K, (\beta_K \Lambda_K))^{-1} \cdot W(\Lambda_K | W_K, \nu_K) d\mu_K d\Lambda_K$

$= \iint$ cont: $\boxed{\text{Not solved}}$ still difficult $d\Lambda_K$ before $d\mu_K$

10.16 $\mathbb{E}[\ln p(X|Z,\mu,\Lambda)] = \iint \prod_{n=1}^{N}\prod_{k=1}^{K} \ln \mathcal{N}(x_n|\mu_k,\Lambda_k^{-1})^{z_{nk}} = \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K}\mathbb{E}[z_{nk}]\{\mathbb{E}[\ln\Lambda_k] - \mathbb{E}[(x_n-\mu_k)\Lambda(x_n-\mu_k)] - D\ln(2\pi)$

$$= \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\{\ln\Lambda_k - D\beta_k^{-1} - \nu_k(x-m_k)^T W_k(x-m_k) - D\ln 2\pi\}$$

$$= \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K} N_k\{\ln\Lambda_k - D\beta_k^{-1} - \nu_k \mathrm{Tr}(S_k W_k)$$

$$- \nu_k(x_k-m_k)^T W_k(x_k-m_k) - D\ln(2\pi)\}$$

$$N_k = \sum r_{nk} \;;\; x_k = \frac{1}{N_k}\sum r_{nk} x_n \;;\; S_k = \frac{1}{N_k} r_{nk}(x_n-\bar{x}_k)^T(x_n-x_k)$$

$$\beta_k = \beta_0 + N_k \;;\; m_k = \frac{1}{\beta}(\beta_0 m_0 + N_k\bar{x}_k) \;;\; W_k^{-1} = W_0^{-1} + N_k S_k + \frac{\beta_0 N_k}{\beta_0 + N_k}(\bar{x}_k - m_0)(x_k - m_0)^T$$

$$\nu_k = \nu_0 + N_k$$

$$\mathbb{E}[\ln p(X|Z,\mu,\Lambda)] = \prod_{n=1}^{N}\prod_{k=1}^{K}\ln \mathcal{N}(x_n|\mu_k,\Lambda_k^{-1})^{z_{nk}} = \frac{1}{2}\sum\sum\mathbb{E}[z_{nk}]\{\mathbb{E}[\ln\Lambda_k] - \mathbb{E}[(x_n-\mu_k)\Lambda(x_n-\mu_k)] - D\ln(2\pi)$$

$$= \frac{1}{2}\sum\sum N_k\{\ln\Lambda - \underbrace{\int (x_n-\mu_k)\Lambda(x_n-\mu_k) \mathcal{N}(\mu_k|m_k,(\beta\Lambda)^{-1}) W(\Lambda_k|W_k,\nu_k)\,d\mu\, d\Lambda}_{D\beta_k^{-1} + \nu_k(x_n-m_k)^T W_k(x_n-m_k)}\}$$

Again, $\int (x_n-\mu_k)\Lambda(x_n-\mu_k)\mathcal{N}(\mu_k|m_k,(\beta\Lambda)^{-1}) W(\Lambda_k|W_k,\nu_k)\,d\mu\, d\Lambda$

$$\iint (x_n-\mu_k)\Lambda(x_n-\mu_k)\left(\frac{\beta\Lambda}{2\pi}\right)^{D/2} \cdot e^{-\frac{1}{2}(\mu_k-m_k)\beta\Lambda(\mu_k-m_k)} \cdot e^{-\frac{1}{2}(\Lambda-W_k)\frac{1}{\nu_k}(\Lambda-W_k)} \cdot d\mu \cdot d\Lambda$$

$$\int (x_n-\mu_k)^T(x_n-\mu_k) \cdot \int \Lambda \cdot \left(\frac{\beta\Lambda}{2\pi}\right)^{D/2} \cdot e^{-\frac{1}{2}(\mu_k-m_k)\beta\Lambda(\mu_k-m_k)} \cdot e^{-\frac{1}{2}(\Lambda-W_k)\frac{1}{\nu_k}(\Lambda-W_k)}\, d\Lambda\, d\mu$$

$$\int \Lambda \left(\frac{\beta\Lambda}{2\pi}\right)^{D/2} \cdot e^{-\frac{1}{2}(\mu_k-m_k)\beta\Lambda(\mu_k-m_k)} \cdot e^{-\frac{1}{2}(\Lambda^T\Lambda + 2\Lambda W_k + W^T W)/\nu_k}\, d\Lambda\, d\mu$$

10.17 $E[\ln p(\pi)] = \ln C(\alpha_0) + (\alpha_0 - 1)\sum_{k=1}^{K} \ln \pi_k$

$$= \int \ln p(\pi)\, q(\pi, \mu, \Lambda)\, d\mu\, d\Lambda = \int \ln p(\pi) \cdot q(\pi) \prod_{k=1}^{K} q(\mu_k, \Lambda_k)\, d\mu\Lambda$$

$$= \iiint \ln \mathrm{Dir}(\pi|\alpha_0) \cdot \mathrm{Dir}(\pi|\alpha) \prod_{k=1}^{K} N(\mu_k|m_k, (\beta\Lambda)^{-1}) W(\Lambda_k|W_k, v_k)\, d\mu\, d\Lambda\, d\pi$$

$$= \iiint \ln C(\alpha_0) \prod_{k=1}^{K} \pi_k^{\alpha-1} \cdot C(\alpha) \prod_{k=1}^{K} \pi_k^{\alpha-1} \cdot \prod_{k=1}^{K} \left(\frac{\beta\Lambda}{2\pi}\right)^{1/2} e^{-\frac{\beta\Lambda}{2}(\mu_k - m_k)^T(\mu_k - m_k)} \cdot \left(\frac{v_k}{2\pi}\right)^{1/2} e^{-\frac{v_k}{2}(\Lambda_k - W_k)^T(\Lambda_k - W_k)}\, d\mu\, d\Lambda\, d\pi$$

$$= \ln C(\alpha_0) \iiint \prod_{k=1}^{K} \pi_k^{\alpha_0-1} \cdot C(\alpha) \prod_{k=1}^{K} \pi_k^{\alpha-1} \cdot \prod_{k=1}^{K} \left(\frac{\beta\Lambda}{2\pi}\right)^{1/2} e^{-\frac{\beta\Lambda}{2}(\mu_k - m_k)^T(\mu_k - m_k)} \cdot \left(\frac{v_k}{2\pi}\right) e^{-\frac{v_k}{2}(\Lambda - W)^T(\Lambda - W)}\, d\mu\, d\Lambda\, d\pi$$

$$= \ln C(\alpha_0) \iiint C(\alpha) \prod_{k=1}^{K} \pi_k^{\alpha_0 + \alpha - 2}$$

---

sparse(m, n, density, format, dtype, random state)

- m = rows
- n = columns
- density = 1 or 0 matrix
- Format = used to specify format of a matrix
- dtype - Data type of returned matrix

sparse(i, j, v, m, n)

- i = point x
- j = point y
- v = value @ (x, y)
- m = # of rows at zero
- n = # of cols at zero

W17 $\mathbb{E}[\ln p(\pi)] = \int \ln C(\alpha_0) \prod_{k=1}^{K} \pi_k^{\alpha_0 - 1} \, d\pi = \int \ln C(\alpha_0) + (\alpha_0 - 1) \ln \prod_k^K \pi$

$$= \ln C(\alpha_0) + (\alpha_0 - 1) \ln \prod_{K=1}^{K} \pi$$

$$= \ln C(\alpha_0) + (\alpha_0 - 1) \sum \ln \pi$$

$\mathbb{E}[\ln p(\mu, \Lambda)] = \ln p(\mu|\Lambda) p(\Lambda)$

$$= \ln \prod^{K} N(\mu_k | m_0, (\beta\Lambda)^{-1}) W(\Lambda_k | W_0, v_0)$$

$$\cancel{= \sum \ln \left[\frac{\beta\Lambda}{2\pi}\right]^{D/2} e^{\frac{\beta\Lambda}{2}(\mu_k - m_0)(\mu_k - m_0)} \left(\frac{v}{2\pi}\right)^{D/2} e^{\frac{-1}{2v_0}(\Lambda - W)(\Lambda - W)}}$$

$$\cancel{= \frac{D}{2}\left[\ln\left(\frac{\beta\Lambda}{2\pi}\right)\right] \frac{\beta\Lambda}{2}(\mu_k - m_0)(\mu_k - m_0) + \frac{D}{2}\ln\left[\frac{v}{2\pi}\right] + \frac{-1}{2v_0}(\Lambda - W)(\Lambda - W)}$$

$$\cancel{= \frac{D}{2}\left[\ln\left(\frac{\beta\Lambda}{2\pi}\right)\right] - \frac{\beta\Lambda}{2}}$$

$$= \sum \ln \left[\frac{\beta\Lambda}{2\pi}\right]^{D/2} e^{-\frac{\beta\Lambda}{2}(\mu_k - m_0)(\mu_k - m_0)} \cdot B(W, v) |\Lambda|^{\frac{v-D-1}{2}} e^{-\frac{v_k}{2}\mathrm{Tr}(W^{-1}\Lambda)}$$

$$= \frac{D}{2}\left(\frac{\beta\Lambda}{2\pi}\right) + \frac{-\beta\Lambda}{2}(\mu_k - m_0)(\mu_k - m_0) + \ln B(W, v) + \frac{v - D - 1}{2} \ln \Lambda - \frac{v_k}{2} \mathrm{Tr}(W^{-1}\Lambda)$$

$$= \sum \frac{D}{2} \ln\left(\frac{\beta\Lambda}{2\pi}\right) + \frac{-\beta\Lambda}{2}(\mu_k^T \mu_k - \mu_k m_0 + m_0^T m_0) + \ln B(W, v) + \frac{v - d - 1}{2} \ln \Lambda - \frac{v_k}{2} \mathrm{Tr}(W^{-1}\Lambda)$$

$$m_k = \frac{1}{\beta_k}(\beta_0 m_0 + N_k \bar{x}_k)$$

How is $-\frac{\beta\Lambda}{2}(\mu_k^T \mu_k - \mu_k m_0 + m_0^T m_0)$ equivalent to ...

$$\beta_k = \beta_0 + N_k \,;\; W_k^{-1} = W_0^{-1} + N_k S_k + \frac{\beta_0 N}{\beta_0 + N}(\bar{x}_k - m_0)(\bar{x}_k - m_0)$$

$$v_k = v_0 + N_k$$

$$\ln \Lambda - \frac{D\beta_0}{\beta_k} - \beta_0 v_k (m_k - m_0)^T W (m_k - m_0)$$

Wishart: $p(R) = 2^{-vd/2} \pi^{-d(d-1)/4} |S|^{v/2} \prod^{d} \Gamma\left(\frac{v+1-i}{2}\right)^{-1} |R|^{(v-d-1)/2} \cdot \exp\left(-\frac{1}{2}\mathrm{Tr}|RS|\right)$

$$= 2^{-vD/2} \cdot \pi^{-D(D-1)/4} \cdot |W|^{-v/2} \cdot \prod^{d} \Gamma\left(\frac{v+1-i}{2}\right)^{-1} |\Lambda|^{(v-d-1)/2} \cdot \exp\left(-\frac{1}{2}\mathrm{Tr}(W^{-1}\Lambda)\right)$$

Gaussian: $p(\mu|R) = \frac{|rR|^{1/2}}{(2\pi)^{D/2}} \cdot \exp\left(-\frac{1}{2}\mathrm{Tr}\left[rR(\mu - m)(\mu - m)^T\right]\right)$

$$p(\mu, R) = \frac{1}{Z(d, r, v, S)} |R|^{(v-d)/2} \exp\left(-\frac{1}{2}\mathrm{Tr}\left[R\left(r(\mu - m)(\mu - m)^T + S\right)\right]\right)$$

$$Z(d, r, v, S) = 2^{\frac{(v+1)d}{2}} \pi^{d(d+1)/4} \cdot r^{-d/2} |S|^{-v/2} \cdot \prod_{i=1}^{d} \Gamma\left(\frac{v+1-i}{2}\right)$$

$$\mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)$$

$$\mathrm{tr}\left(W_0^{-1}\Lambda + \sum (x-\bar{x})^T(x-\bar{x})\Lambda + \frac{NK_0}{K_0+N}(\bar{x}-\mu_0)(\bar{x}-\mu_0)^T\Lambda\right)$$

$$= \mathrm{tr}\left(\left(W_0^{-1} + \sum (x-\bar{x})^T(x-\bar{x}) + \frac{NK_0}{K_0+N}(\bar{x}-\mu_0)(\bar{x}-\mu_0)^T\Lambda\right)\right)$$

if $S = \sum (x_i-\bar{x})^T(x_i-\bar{x})$

$$|\Lambda|^{1/2}\cdot\exp\left\{-\frac{K_0+N}{2}\left(\mu-\frac{K_0\mu+N\bar{x}}{K_0+N}\right)^T\Lambda\left(\mu-\frac{K_0\mu+Nx}{K_0+N}\right)\right\}$$

$$\times|\Lambda|^{(V_0+N-D-1)/2}\cdot\exp\left\{-\frac{1}{2}\mathrm{tr}\left(W_0^{-1}+S+\frac{NK_0}{K_0+N}(\bar{x}-\mu_0)(\bar{x}-\mu_0)\right)\Lambda\right\}$$

Switch of variables: $|\Lambda|^{1/2}\cdot\exp\left\{-\frac{B\Lambda}{2}(m_K-m_0)^T(m_K-m_0)\right\}$

$$\times|\Lambda|^{(V_0+N-D-1)/2}\cdot\exp\left\{-\frac{1}{2}\mathrm{Tr}(W^{-1}\Lambda)\right\}$$

$$\mathbb{E}_{\mu_K;\Lambda_K}\left[(x_n-\mu_K)\Lambda(x_n-\mu_K)\right] = \boxed{\text{Still unsure}}$$

$$\mathbb{E}[\ln q(Z)] = \ln \prod_{n=1}^{N}\prod_{K=1}^{K} r_{nk}^{z_{nk}} = \sum\sum \mathbb{E}[z_{nk}]\ln r_{nk} = \sum\sum r_{nk}\cdot \ln r_{nk}$$

$$\mathbb{E}[\ln q(\pi)] = \mathrm{Dir}(\pi|\alpha) = \ln C(\alpha)\prod_{K=1}^{K}\mu_K^{\alpha_K-1} = (\alpha_K-1)\ln\mu_K + \ln C(\alpha)$$

$$\mathbb{E}[\ln q(\mu,\Lambda)] = \ln N(\mu_K|m_K,(\beta\Lambda)^{-1})\,W(\Lambda|W_K,v_K)$$

10.18 $q(Z,\pi,\mu,\Lambda) = q(Z)\,q(\pi,\mu,\Lambda)$; $q(\pi,\mu,\Lambda) = q(\pi)\prod_{K=1}^{K} q(\mu_K,\Lambda_K)$

with $q^*(Z) = \prod_{n=1}^{N}\prod_{K=1}^{K} r_{nK}^{z_{nK}}$; $q^*(\pi) = \mathrm{Dir}(\pi|\alpha)$

$$q^*(\mu_K,\Lambda_K) = N(\mu_K|m_K,(\beta\Lambda)^{-1})\,W(\Lambda_K|W_K,v_K)$$

Substitute into: $\mathcal{L} = \sum\iiint q(Z,\pi,\mu,\Lambda)\ln\left\{\frac{p(X,Z,\pi,\mu,\Lambda)}{q(Z,\pi,\mu,\Lambda)}\right\}d\pi\, d\mu\, d\Lambda$

$$= \mathbb{E}[\ln p(X,Z,\pi,\mu,\Lambda)] - \mathbb{E}[\ln q(Z,\pi,\mu,\Lambda)]$$

$$= \mathbb{E}[\ln p(X|Z,\mu,\Lambda)] + \mathbb{E}[\ln p(Z|\pi)] + \mathbb{E}[\ln p(\pi)] + \mathbb{E}[\ln p(\mu,\Lambda)]$$
$$- \mathbb{E}[\ln q(Z)] - \mathbb{E}[\ln q(\pi)] - \mathbb{E}[\ln q(\mu,\Lambda)]$$

$$= \sum\iiint q(Z)\,q(\pi,\mu,\Lambda)\ln\left\{\frac{p(X,Z,\pi,\mu,\Lambda)}{q(Z)\,q(\pi,\mu,\Lambda)}\right\}d\pi\, d\mu\, d\Lambda$$

$$= \sum\iiint \prod_{n=1}^{N}\prod_{K=1}^{K} r_{nK}^{z_{nK}}\cdot q(\pi)\prod_{K=1}^{K} q(\mu_K,\Lambda_K)\ln\left\{\frac{p(X,Z,\pi,\mu,\Lambda)}{\prod_{n=1}^{N}\prod_{K=1}^{K} r_{nK}^{z_{nK}}\cdot q(\pi)\cdot\prod_{K=1}^{K} q(\mu_K,\Lambda_K)}\right\}d\pi\, d\mu\, d\Lambda$$

$$= \sum\iiint \prod_{n=1}^{N}\prod_{K=1}^{K} r_{nK}^{z_{nK}}\cdot \mathrm{Dir}(\pi|\alpha)\prod_{K=1}^{K} N(\mu_K|m_K,(\beta\Lambda)^{-1})\,W(\Lambda|W_K,v_K)$$
$$\cdot\ln\left\{\frac{p(X,Z,\pi,\mu,\Lambda)}{\prod_{n=1}^{N}\prod_{K=1}^{K} r_{nK}^{z_{nK}}\cdot \mathrm{Dir}(\pi|\alpha)\cdot\prod_{K=1}^{K} N(\mu_K|m_K,(\beta\Lambda)^{-1})\,W(\Lambda|W,v_K)}\right\}d\pi\, d\mu\, d\Lambda$$

10.19 Derive $p(\hat{x}|X) = \frac{1}{\hat{\alpha}} \sum_{k=1}^{K} \alpha_k St(\hat{x}|m_k, L_k, \nu_k + 1 - D)$

$$p(\hat{x}|X) = \sum_{\hat{z}} \iiint p(\hat{x}|\hat{z}, \mu, \Lambda)\, p(\hat{z}|\pi)\, p(\pi, \mu, \Lambda|X)\, d\pi\, d\mu\, d\Lambda$$

$$= \sum_{k=1}^{K} \iiint \pi_k N(\hat{x}|\mu_k, \Lambda_k^{-1})\, p(\pi, \mu, \Lambda|X)\, d\pi\, d\mu\, d\Lambda$$

$$= \sum_{k=1}^{K} \iiint \pi_k N(\hat{x}|\mu_k, \Lambda_k^{-1})\, q(\pi)\, q(\mu_k, \Lambda_k)\, d\pi\, d\mu\, d\Lambda$$

Suppose, $= \frac{1}{\hat{\alpha}} \sum \alpha_k \frac{\Gamma(\nu/2 + 1/2)}{\Gamma(\nu/2)} \left(\frac{\lambda}{\pi\nu}\right)^{1/2} \left[1 + \frac{\lambda(x-\mu)^2}{\nu}\right]^{-\nu/2 - 1/2}$

$$\cong \frac{1}{\hat{\alpha}} \sum_{k=1}^{K} \alpha_k \frac{\Gamma[(\nu_k+1-D)/2 + 1/2]}{\Gamma[(\nu_k+1-D)/2]} \left(\frac{L_k}{\pi(\nu_k+1-D)}\right)^{1/2} \left[1 + \frac{(x-m_k)^2}{\nu_k + 1 - D}\right]^{-(\nu_k-1+D)/2 - 1/2}$$

$$L_k = \frac{(\nu_k + 1 - D)\beta_k}{(1+\beta_k)} W_k$$

$$= \sum_{k=1}^{K} \iiint \pi_k \left(\frac{\Lambda}{2\pi}\right)^{1/2} \exp\left\{-\frac{\Lambda}{2}(\hat{x}-\mu_k)(\hat{x}-\mu_k)\right\} Dir(\pi|\alpha)$$

$$\times N(\mu_k|m_k, (\beta\Lambda)^{-1})\, W(\Lambda|W_k, \nu_k)\, d\pi\, d\mu\, d\Lambda$$

$$= \sum_{k=1}^{K} \iiint \pi_k \left(\frac{\Lambda}{2\pi}\right)^{1/2} \exp\left\{-\frac{\Lambda}{2}(\hat{x}-\mu_k)(\hat{x}-\mu_k)\right\} \cdot C(\alpha) \prod_{k=1}^{K} \mu_k^{\alpha_k - 1} \cdot \left(\frac{\beta\Lambda}{2\pi}\right)^{D/2} \exp\left\{-\frac{\beta\Lambda}{2}(\mu_k - m_k)(\mu_k - m_k)\right\}$$

$$\times B(W, \nu) |\Lambda|^{(\nu-D-1)/2} \cdot \exp\left(-\frac{1}{2} Tr(W^{-1}\Lambda)\right) d\pi\, d\mu\, d\Lambda$$

$$= \sum \iiint \pi_k \cdot \left(\frac{\Lambda}{2\pi}\right)^{1/2} \exp\left\{-\frac{\Lambda}{2}(\hat{x}-\mu_k)(\hat{x}-\mu_k)\right\} C(\alpha) \prod_{k=1}^{K} \pi_k^{\alpha_k - 1} \cdot \left(\frac{\beta\Lambda}{2\pi}\right)^{D/2} \exp\left\{-\frac{\beta\Lambda}{2}(\mu_k - m_k)(\mu_k - m_k)\right\}$$

$$\times |W|^{-\nu/2} \cdot \left(2^{\nu D/2} \cdot \pi^{D(D-1)/4} \cdot \prod_{i=1}^{D} \Gamma\left(\frac{\nu + 1 - i}{2}\right)\right)^{-1} |\Lambda|^{(\nu-D-1)/2} \cdot \exp\left(-\frac{1}{2} Tr(W^{-1}\Lambda)\right) d\pi\, d\mu\, d\Lambda$$

$$= \sum \frac{C(\alpha) \cdot \pi^{\alpha_k}}{\alpha_k} \cdot \iint \boxed{\left(\frac{\Lambda}{2\pi}\right)^{1/2} \exp\left\{-\frac{\Lambda}{2}(\hat{x}-\mu_k)(\hat{x}-\mu_k)\right\} \cdot \left(\frac{\beta\Lambda}{2\pi}\right)^{\nu/2} \exp\left\{-\frac{\beta\Lambda}{2}(\mu_k - m_k)(\mu_k - m_k)\right\}}$$

$$\times |W|^{-\nu/2} \left(2^{\nu D/2} \cdot \pi^{D(D-1)/4} \cdot \prod_{i=1}^{D} \Gamma\left(\frac{\nu+1-i}{2}\right)\right)^{-1} |\Lambda|^{(\nu-D-1)/2} \cdot \exp\left(-\frac{1}{2} Tr(W^{-1}\Lambda)\right) d\mu\, d\Lambda$$

$$\boxed{(\beta)^{D/2} \cdot \left(\frac{\Lambda}{2\pi}\right)^{(\nu/2 + 1/2)}} \cdot \exp\left\{-\frac{\Lambda}{2}\left[(\hat{x}-\mu_k)^T(\hat{x}-\mu_k) + \beta(\mu_k - m_k)(\mu_k - m_k) + Tr(W^{-1}\Lambda)\right]\right\}$$

$$\times |W|^{-\nu/2} \left(2^{-\nu D/2} \cdot \pi^{-D(D-1)/4} \cdot \frac{1}{\Gamma\left(\frac{\nu+1-i}{2}\right)}\right) |\Lambda|^{(\nu-D-1)/2}$$

$$(\beta)^{1/2} \cdot \left(\frac{\Lambda}{2\pi}\right)^{(\nu/2+1/2)} \cdot \exp\left\{-\frac{\Lambda}{2}\left[(\hat{x}-\mu_k)(\hat{x}-\mu_k) + \beta(\mu_k - m_k)(\mu_k - m_k) + Tr(W^{-1})\right]\right\}$$

$$\times |W|^{-\nu/2} \cdot 2^{-\nu D/2} \cdot \pi^{-D(D-1)/4} \cdot \frac{1}{\Gamma\left(\frac{\nu+1-i}{2}\right)} \cdot |\Lambda|^{(\nu-D-1)/2}$$

10.20 $q^*(\Lambda_K) = W(\Lambda_K | W_K, \nu_K)$ (10.63) $\lim_{N\to\infty} \nu_K = \; = N_K$ -and- $\lim_{N\to\infty} W_K = \; = N_K^{-1} S_K^{-1}$

$E[\Lambda] = \nu_K W_K = S_K^{-1}$; (B.82) $\lim_{N\to\infty} H[\Lambda] = \; = 0$

$\lim_{N\to\infty} -\ln B(W_K, \nu_K) = -\frac{N_K}{2}(D\ln N_K + \ln|S| - D\ln 2) + \sum \ln\Gamma\left(\frac{N_K+1-i}{2}\right)$

$q^*(\mu_K|\Lambda_K) = N(\mu_K | m_K, (\beta\Lambda)^{-1}) W(\Lambda_K | W_K, \nu_K)$ $= -\frac{N_K}{2}(\ln|S| + D)$; $\boxed{E[\ln\Lambda] = -\ln|S_K|}$

$$\frac{dq(\mu,\Lambda_K)}{d\mu_K} = \frac{d}{d\mu_K}\left(\frac{\beta\Lambda}{2\pi}\right)^{D/2} \exp\left\{-\frac{\beta\Lambda}{2}(\mu_K - m_K)^T(\mu_K - m_K)\right\} \cdot B(W,\nu)|\Lambda|^{(\nu-D-1)/2} \exp\left(-\frac{1}{2}Tr(W^{-1}\Lambda)\right)$$

$u = (\mu_K - m_K) \qquad du = 1\, d\mu_K$

$$= \left(\frac{\beta\Lambda}{2\pi}\right)^{D/2} \cdot \left[\frac{d}{du}\frac{-\beta\Lambda}{2}u^2\right] \exp\left\{-\frac{\beta\Lambda}{2}(\mu_K - m_K)^T(\mu_K - m_K)\right\} \cdot B(W,\nu) \cdot |\Lambda|^{(\nu-D-1)/2} \cdot \exp\left(-\frac{1}{2}Tr(W^{-1}\Lambda)\right)$$

$$= \left(\frac{\beta\Lambda}{2\pi}\right)^{D/2} \cdot \left[-\beta\Lambda(\mu_K - m_K) \cdot \exp\left\{-\frac{\beta\Lambda}{2}(\mu_K - m_K)^T(\mu_K - m_K)\right\}\right] B(W,\nu)|\Lambda|^{(\nu-D-1)/2} \exp\left(-\frac{1}{2}Tr(W^{-1}\Lambda)\right)$$

$$0 = \exp\left\{-\frac{\beta\Lambda}{2}(\mu_K - m_K)^T(\mu_K - m_K)\right\}$$

$\therefore$ $\boxed{\text{"Too sharp around mean" - Unsolvable}}$

$q^*(\pi) = Dir(\pi|\alpha)$

$$\frac{dq(\pi)}{d\pi} = \frac{d}{d\pi} C(\alpha) \prod_{K=1}^{K} \pi_K^{\alpha_K - 1} = (\alpha_K - 1)\, C(\alpha) \prod_{K=1}^{K} \pi_K^{\alpha_K - 2} = 0$$

$$\pi_K^{\alpha_K - 2} = 0$$

$$\boxed{\pi_K = 0}$$

$\psi(x) = \ln x + O(1/x)$ (10.67) (10.65) (10.66)

$\ln\tilde{\Lambda}_K = E[\ln|\Lambda_K|] = \sum_{i=1}^{D} \psi\left(\frac{\nu_K + 1 - i}{2}\right) + D\ln 2 + \ln|W_K|$

$\ln\tilde{\pi}_K = E[\ln\pi_K] = \psi(\alpha_K) - \psi(\hat{\alpha})$

$r_{nK} \propto \tilde{\pi}_K \tilde{\Lambda}_K^{1/2} \left\{-\frac{D}{2\beta_K} - \frac{\nu_K}{2}(x_n - m_K)^T W_K (x_n - m_K)\right\}$

Finally, $p(\hat{x}|D) = \sum_{K=1}^{K} \frac{\alpha_K}{\hat{\alpha}} \iint N(\hat{x}|\mu_K, \Lambda_K)\, q(\mu_K, \Lambda_K)\, d\mu_K\, d\Lambda_K$

$$\boxed{= \sum_{K=1}^{K} \frac{N_K}{N} N(\hat{x}|\bar{x}, W_K)}$$

10.21 $\boxed{K(K-1)(K-2)(K-3)(K-4)\cdots 1 = K!}$

10.22 $q^*(\pi) = \mathrm{Dir}(\pi|\alpha)$; $\prod_{k=1}^{K} q^*(\pi) = \prod_{k=1}^{K} \mathrm{Dir}(\pi_i|\alpha)$

$$\boxed{\prod_{k=1}^{K} \ln q^*(\pi) = \prod_{k=1}^{K} \ln \mathrm{Dir}(\pi_i|\alpha) < \prod_{k=1}^{K+1} \ln q^*(\pi) = \prod_{k=1}^{K+1} \ln \mathrm{Dir}(\pi_i|\alpha)}$$

10.23 $\pi_K = \frac{1}{N}\sum_{n=1}^{N} r_{nK}$

$$\mathcal{L} = \sum_{z}\iiint q(z,\pi,\mu,\Lambda)\ln\left\{\frac{p(X,z,\pi,\mu,\Lambda)}{q(z,\pi,\mu,\Lambda)}\right\} d\pi\, d\mu\, d\Lambda$$

$$= \mathbb{E}[\ln p(X,z,\pi,\mu,\Lambda)] - \mathbb{E}[\ln q(z,\pi,\mu,\Lambda)]$$

$$= \mathbb{E}[\ln p(X|z,\mu,\Lambda)] + \mathbb{E}[\ln p(z|\pi)] + \mathbb{E}[\ln p(\pi)] + \mathbb{E}[\ln p(\mu,\Lambda)]$$

$$- \mathbb{E}[\ln q(z)] - \mathbb{E}[\ln q(\pi)] - \mathbb{E}[\ln q(\mu,\Lambda)]$$

$$\mathbb{E}[\ln p(z|\pi)] = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nK}\ln\tilde{\pi}_K + \lambda\left[\sum \pi_K - 1\right]$$

$$\frac{N_K}{\pi_K} + \lambda = 0 \ ; \quad N_K = -\lambda\pi_K$$

$$\sum N_K = -\lambda$$

$$\frac{N_K}{\pi_K} - \sum N_K = 0 \ ; \quad \boxed{\pi_K = \frac{N_K}{\sum N_K}}$$

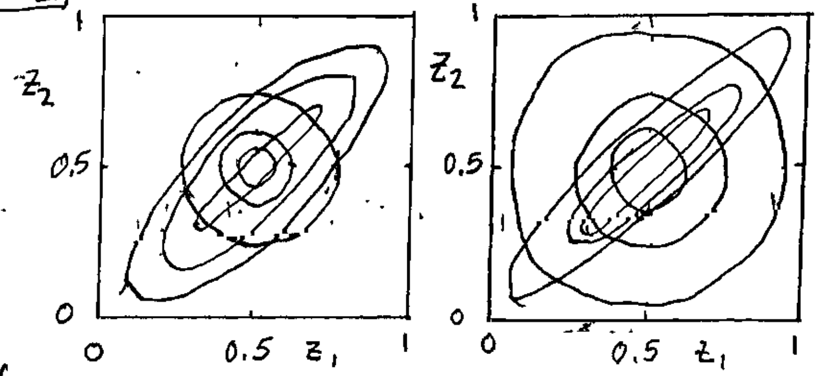
10.24 Maximum Posterior (MAP) Estimation:

Baysion $\underset{\text{posterior}}{p(\theta|X)} = \frac{\overset{\text{liklihood}}{p(X|\theta)}\,\overset{\text{Prior}}{p(\theta)}}{\underset{\text{Evidence}}{p(X)}}$; $\Theta_{MAP} = \arg\max p(\theta|X) = \arg\max p(X|\theta)p(\theta)$

$= \arg\max \sum_{n=1}^{N} \log p(x|\theta) + \log p(\theta)$

Assumming $\arg\max \sum_{n=1}^{N} \log p(X|\theta) + \log p(\theta)$, then the domain is from $\theta \to \infty$ because $p(X|\theta)$ is $0 \to \infty$ and $\log p(X|\theta)$ is from $\theta \to \infty$.

10.25 $q(Z)=\prod_{i=1}^{M} q_i(z_i)$ $\boxed{\text{Figure 10.2}}$



If a Baysian mixture of Gaussians made use of a factorized approximation to the posterior distribution, then the posterior is capable of being underestimated for specific regions of data. As example,

$$\vec{\mu}_{Fit} \begin{pmatrix} \mu_{1F} \\ \mu_{2F} \end{pmatrix} < \vec{\mu}_{Actual} = \begin{pmatrix} \mu_{1A} \\ \mu_{2A} \end{pmatrix} \text{ and } \Lambda_{Fit} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} < \Lambda_{Actual} = \begin{pmatrix} \lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}$$

As the number of components increase, so does $(\mu_{Fit} = \mu_{Actual})$ and $(\Lambda_{Fit} = \Lambda_{Actual})$

10.26. $q(W)\cdot q(\alpha)\cdot q(\beta) = \boxed{\underbrace{p(t|W)}_{q(W)} \cdot \underbrace{p(W|\alpha)\cdot p(\alpha)}_{q(\alpha)} \cdot \underbrace{p(W|\beta)p(\beta)}_{q(\beta)}}$

$$\ln q(W) = \ln p(t|W) + \mathbb{E}_{\alpha}[\ln p(W|\alpha)] + \mathbb{E}_{\beta}[\ln p(W|\beta)] + const$$

$$= -\frac{\beta}{2}\sum_{n=1}^{N}\{W^T\phi_n - t_n\}^2 - \frac{1}{2}\mathbb{E}[\alpha]W^TW - \frac{1}{2}\mathbb{E}[\beta]W^TW$$

$$\ln q(\alpha) = \ln p(\alpha) + \mathbb{E}_{W}[\ln p(W|\alpha)] + const$$

$$= (a_0-1)\ln\alpha - b_0\alpha + \frac{M}{2}\ln\alpha - \frac{\alpha}{2}\mathbb{E}[W^TW] + const$$

$$\ln q(\beta) = \ln p(\beta) + \mathbb{E}[\ln p(W|\beta)] + const$$

$$= (c_0-1)\ln\alpha - d_0\alpha + \frac{M}{2}\ln\alpha - \frac{\alpha}{2}\mathbb{E}[W^TW] + const$$

$$q(\beta) = Gam(\beta|c_0, d_0)$$

$$\boxed{c_0 = \frac{N_0}{2}\ ;\ d_0 = \frac{1}{2}\sum_{n=1}^{N}\{W^T\phi_n - t_n\}^2}$$

10.27 Prove $\mathcal{L}(q) = \mathbb{E}[\ln p(w,\alpha,t)] - \mathbb{E}[\ln q(w,\alpha)]$

$$= \mathbb{E}_w[\ln p(t|w)] + \mathbb{E}_{w,\alpha}[\ln p(w|\alpha)] + \mathbb{E}_\alpha[\ln p(\alpha)] - \mathbb{E}_\alpha[\ln q(w)]_w - \mathbb{E}[\ln q(\alpha)]$$

$$\mathbb{E}_w[\ln p(t|w)]_w = \ln \prod_{n=1}^{N} N(t_n | w^T \phi_n, \beta^{-1})$$

$$\boxed{= \sum_{n=1}^{N} \frac{N}{2} \ln \frac{\beta}{2\pi} - \frac{\beta}{2} t^T t + t_n^T \cdot w_n^T \Phi_n - \frac{\beta}{2} Tr[w^T \phi(x) \cdot w^T \phi(x)]}$$

$$\mathbb{E}_{w,\alpha}[\ln p(w|\alpha)] = \ln N(w | 0, \alpha^{-1} I) \quad ; \mathbb{E}[\ln \tau] = \psi(a) - \ln b$$

$$\mathbb{E}[\tau] = a/b$$

$$= \frac{N}{2} \ln \frac{\mathbb{E}[\alpha]}{2\pi} - \frac{\mathbb{E}[\tau]}{2} \cdot \mathbb{E}[w^T w]$$

$$= -\frac{N}{2} \ln 2\pi + \frac{N}{2} \mathbb{E}[\ln \alpha] - \frac{a_n}{2b_n}[m_n m_n^T + S_n]$$

$$\boxed{= -\frac{N}{2} \ln 2\pi + \frac{N}{2}[\psi(a_n) - \ln b_n] - \frac{a_n}{2b_n}[m_n m_n^T + S_n]}$$

$$\mathbb{E}[\ln p(\alpha)] = \ln Gam(\alpha | a_n, b_n)$$

$$= \ln \frac{1}{\Gamma(a_n)} \cdot b_n^{a_n} \cdot \alpha^{a_n - 1} \cdot e^{-b_n \alpha}$$

$$= -\ln \Gamma(a_n) + a_n \ln b_n + (a_n - 1)\mathbb{E}[\ln \alpha] - b_n \mathbb{E}[\alpha]$$

$$\boxed{= -\ln \Gamma(a_n) + a_n \ln b_n + (a_n - 1)[\psi(a_n) - \ln b_n] - b_n \cdot \frac{a_n}{b_n}}$$

$$-\mathbb{E}[\ln q(w)] = -\ln N(w | m_n, S_N) = -\frac{N}{2} \ln\left(\frac{S_N}{2\pi}\right) + \underbrace{\frac{S_N}{2}(w^T w) + S_N m_N^T w + S_N m_N^T m_N}_{=1}$$

$$\boxed{= \frac{1}{2} \ln |S_N| + \frac{M}{2}[1 + \ln(2\pi)]}$$

$$-\mathbb{E}[\ln q(\alpha)] = -\ln Gam(\alpha | a_n, b_n) = -\ln \frac{1}{\Gamma(a_n)} \cdot b_n^{a_n} \cdot \alpha^{a_n - 1} \cdot e^{-b_n \alpha}$$

$$\boxed{= \ln \Gamma(a_n) - (a_n - 1)\psi(a_n) - \ln b_n + a_n}$$

10.23 $$\ln q^*(z) = \mathbb{E}_\eta[\ln p(X, z|\eta)] + \text{const}$$
$$= \sum_{n=1}^{N} \{\ln h(x_n, z_n) + \mathbb{E}[\eta^T] u(x_n, z_n)\} + \text{const}$$

$$q^*(\eta) = f(\nu_n, \chi_n) g(\eta)^{\nu_n} \exp\{\eta^T \chi_N\}$$

$$\nu_N = \nu_0 + N$$
$$\chi_N = \chi_0 + \sum_{n=1}^{N} \mathbb{E}_{z_n}[u(x_n, z_n)]$$

Use the above to derive: $q^*(z) = \prod_{n=1}^{N} \prod_{k=1}^{K} r_{nk}^{z_{nk}}$

$$q^*(\pi) = \mathrm{Dir}(\pi|\alpha)$$
$$q^*(\mu_k, \Lambda_k) = N(\mu_k | m_k, (\beta_k \Lambda_k)^{-1}) W(\Lambda_k | W_k, \nu_k)$$

$$\ln q^*(z) = \mathbb{E}_\eta[\ln p(X, z|\eta)] + \text{const}$$
$$= \sum_{n=1}^{N} \{\ln h(x_n, z_n) + \mathbb{E}[\eta^T] u(x_n, z_n)\} + \text{const}$$
$$q(z) = h(x_n, z_n) g(\mathbb{E}[\eta]) \exp\{\mathbb{E}[\eta^T] u(x_n, z_n)\}$$
$$\boxed{q^*(z) = \prod_{n=1}^{N} \prod_{k=1}^{K} \frac{h(x_n, z_n) g(\mathbb{E}[\eta]) \exp\{\mathbb{E}[\eta^T] u(x_n, z_n)\}}{\sum_{j=1}^{K} h(x_n, z_n) g(\mathbb{E}[\eta]) \exp\{\mathbb{E}[\eta^T] u(x_n, z_n)\}}}$$

$$\ln q^*(\eta) = \ln p(\eta|\nu_0, \chi_0) g(\mathbb{E}[\eta]) \exp\{\mathbb{E}[\eta^T] u(x_n, z_n)\}$$
$$= \nu_0 \ln g(\eta) + \eta^T \chi_0 + \sum_{n=1}^{N} \{\ln g(\eta) + \eta^T \mathbb{E}[u(x_n, z_n)]\} + \text{const}$$
$$q^*(\eta) = f(\nu_n, \chi_n) g(\eta)^{\nu_n} \exp\{\eta^T \chi_n\} = \boxed{\frac{1}{T(\nu_n, \chi_n)} g(\eta)^{\nu_n} \cdot \exp(\eta^T \chi_n)}$$
$$q(z, \eta) = q(\mu_k|\eta) q(\eta)$$
$$\boxed{= N(x | \mu_0, \eta) W(\eta | W_k, \nu_0)}$$

10.29. $f(x)=\ln(x)$; $f'(x)=\frac{1}{x}$; $\boxed{f''(x)=-\frac{1}{x^2}}$; $g(\lambda)=\min_x\{\lambda x - f(\lambda)\}$

$$= \lambda - \frac{1}{x} = 0$$

$$; \quad x = \frac{1}{\lambda}$$

$$g(\lambda) = 1 + \ln x = 1 + \ln \frac{1}{\lambda}$$

$$f(x) = \min_\lambda \left\{ \lambda x - 1 + \ln \frac{1}{\lambda} \right\}$$

$$= x - \frac{1}{\lambda} = 0$$

$$f(x) = \frac{1}{x} \cdot x - 1 + \ln\left(\frac{1}{1/x}\right)$$

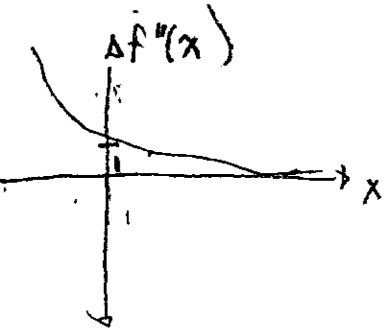
$$= 1 - 1 + \ln x = \boxed{\ln(x)}$$

10.30 $f(x) = -\ln(1+e^{-x})$; $f'(x) = -\frac{-e^{-x}}{1+e^{-x}}$; $f''(x) = \frac{e^{-x}(1+e^{-x})^{-1} - e^{-x}(1+e^{-x})^{-2} \cdot e^{-x}}{(1+e^{-x})^2}$

Derive $\sigma(x) \leq \exp(\lambda x - g(\lambda))$

Taylor Expansion: $f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2$

$$= 1 + \frac{e^{-x}}{1+e^{-x}} + \frac{e^{-x}(1+e^{-x})^{-1} - e^{-x}(1+e^{-x})^{-2} e^{-x}}{(1+e^{-x})^2}$$



$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k$$

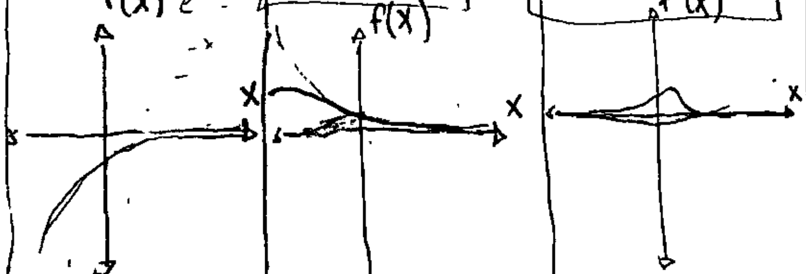
$f(x)$	$f'(x)$	$f''(x)$	$T_{f(x)}$	$g(\lambda)$	λ	$\exp(\lambda x - g(\lambda))$

$\boxed{-\ln(1+e^{-x})}$

$f(x)\, e^{-x}$

$\frac{e^{-x}}{1+e^{-x}}$

$\frac{-e^{-x}(1+e^{-x})^{-1} + e^{-x}(1+e^{-x})^{-2} e^{-x}}{(1+e^{-x})^2}$



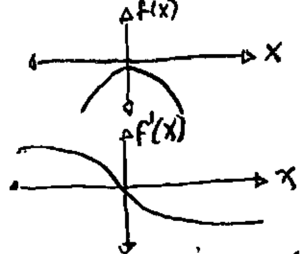
$\sigma(x) = \frac{1}{1+e^{-x}} \leq \exp\{\ln(1+e^{-x})\} = \frac{1}{1+e^{-x}}$

$T_{f(x)} = -\ln(1+e^{-x}) + \frac{e^{-x}}{1+e^{-x}} \cdot x + \frac{e^{-x}(1+e^{-x})^{-1} - e^{-x}(1+e^{-x})^{-2} e^{-x}}{(1+e^{-x})^2}$

$g(\lambda) = \min_\lambda \{\lambda x - \ln(1+e^{-\lambda})\} = x + \frac{e^{-\lambda}}{1+e^{-\lambda}} = 0$

$\boxed{\lambda = -1}$; $\boxed{g(\lambda) = -x - \ln(1+e^{-x})}$

10.31 $f(x) = \boxed{-\ln\left(e^{x/2}+e^{-x/2}\right)}$



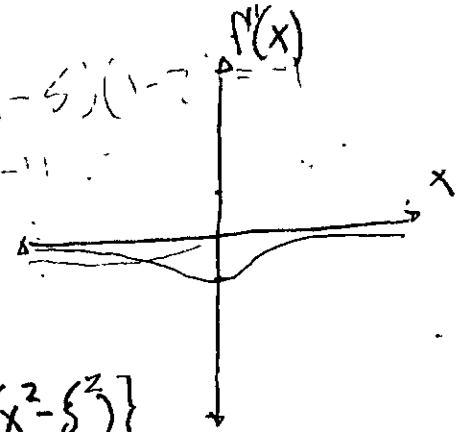
$f'(x) = \boxed{\dfrac{-e^{-x/2} + e^{+x/2}}{2\left(e^{x/2}+e^{-x/2}\right)}}$

$f''(x) = \dfrac{\cancel{\frac{1}{2}\left(e^{x/2}+e^{-x/2}\right)2\left(e^{x/2}+e^{-x/2}\right) - \left(e^{x/2}-e^{-x/2}\right)\frac{2}{2}\left(e^{x/2}-e^{-x/2}\right)}}{\left[2\left(e^{x/2}+e^{-x/2}\right)\right]^{2}}$

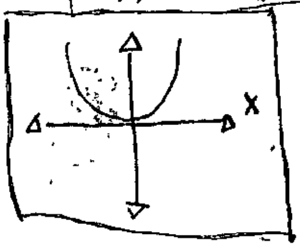
$= \dfrac{-\left(e^{x/2}+e^{-x/2}\right)+\left(e^{x/2}+e^{-x/2}\right)}{4\left(e^{x/2}+e^{-x/2}\right)} = \boxed{\dfrac{-1}{4\left(e^{x/2}+e^{-x/2}\right)}}$

$-(4-\xi)(1-$

$\xi-1$



$\boxed{f(x) = x^2}$; $\boxed{f'(x) = x}$; $\boxed{f''(x) = 1}$



Derive $\sigma(x) \geq \sigma(\xi)\exp\{(x-\xi)/2 - \lambda(\xi)(x^2-\xi^2)\}$

$g(\lambda) = \min_x \{\lambda x - f(x)\}$

$\Rightarrow \lambda - \dfrac{e^{-x/2}-e^{x/2}}{2\left(e^{x/2}+e^{-x/2}\right)} = 0$; $\lambda = \frac{1}{2}$

$\sigma(x) \geq \sigma(\xi)\exp\{\lambda x - \lambda f(x)\}$

$\geq \sigma(\xi)\exp\{x/2 - \lambda\xi + \ln\left(e^{x/2}+e^{-x/2}\right)\}$

$\sigma(x) \geq \sigma(\xi)\exp\{(x-\lambda)/2 + \ln\left(e^{x/2}+e^{-x/2}\right)\}$

10.32

$p(t|w) = e^{at}\sigma(-a) \geq e^{at}\sigma(\xi)\exp\{-(a+\xi)/2 - \lambda(\xi)(a^2-\xi^2)\}$

$\ln\{p(t|w)p(w)\} \geq \ln p(w) + \sum_{n=1}^{N}\{\ln\sigma(\xi_n) + w^T\phi_n t_n - (w^T\phi_n+\xi)/2 - \lambda(\xi)([w^T\phi_n]^2-\xi^2)\}$

$m_N = S_N\left(S_0^{-1}m_0 + \sum_{n=1}^{N}(t_n - 1/2)\phi_n\right)$

$S_N^{-1} = S_0^{-1} + 2\sum_{n=1}^{N}\lambda(\xi)\phi_n\phi_n^T$

$\geq -\frac{1}{2}(w-m_0)^T S_0^{-1}(w-m_0) + \sum_{n=1}^{N}\{w^T\phi_n(t_n - 1/2) - \lambda(\xi)w^T(\phi_n\phi_n^T)w\} + \ln\sigma(\xi)$

$-\xi_n^2 - \xi_n/2$

$$m_N = S_N\left(S_0^{-1}m_0 + \sum_{n=1}^{N}(t_n - 1/2)\phi_n\right)$$

$$= S_N\left(S_0^{-1}m_0 + \sum_{n=1}^{N-1}(t_n - 1/2)\phi_n + (t_n - 1/2)\phi_n\right)$$

$$= S_N\left(S_{N-1}^{-1} S_{N-1}\left(S_0^{-1}m_0 + \sum_{n=1}^{N-1}(t_n - 1/2)\phi_n\right) + (t_n - 1/2)\phi_n\right)$$

$$= S_N\left(S_{N-1}^{-1} m_{N-1} + (t_N - 1/2)\phi_N\right)$$

$$S_N^{-1} = S_0^{-1} + 2\sum_{n=1}^{N}\lambda(\xi_n)\phi_n\phi_n^T$$

$$= S_0^{-1} + 2\sum_{n=1}^{N-1}\lambda(\xi_n)\phi_n\phi_n^T + 2\lambda(\xi)\phi_n\phi_N^T$$

$$= S_{N-1}^{-1} + 2\lambda(\xi_N)\phi_n\phi_n^T$$

(10.163) $(\xi_n^{new})^2 = \phi_n^T\mathbb{E}[ww^T]\phi_n = \phi_n^T(S_N + m_N m_N^T)\phi_n$

$$\boxed{\begin{aligned}\ln\{p(t|w)p(w)\} \geq -\tfrac{1}{2}(w-m_0)^T S_0^{-1}(w-m_0) + \sum_{n=1}^{N}\{w^T\phi_n(t_n - 1/2) - \lambda(\xi)w^T\phi_n\phi_n^T w\} \\ + \ln\sigma(\xi) - \phi_n^T(S_N + m_N m_N^T)\phi_n - \phi_n^T(S_N + m_n m_n^T)^{1/2}_n/2\end{aligned}}$$

10.33 $$Q(\xi, \xi^{old}) = \sum_{n=1}^{N}\{\ln\sigma(\xi_n) - \xi_n/2 - \lambda(\xi_n)(\phi_n^T\mathbb{E}[ww^T]\phi_n - \xi_n^2)\} + const$$

$$\frac{dQ(\xi,\xi^{old})}{d\xi_n} = \frac{d}{d\xi_n}\left[\sum_{n=1}^{N}\{\ln\sigma(\xi_n) - \xi/2 - \lambda(\xi_n)(\phi_n^T\mathbb{E}[ww^T]\phi_n - \xi_n^2)\}\right] + const$$

$$= \sum_{n=1}^{N}\frac{1}{\sigma(\xi_n)}\frac{d\sigma(\xi_n)}{d\xi_n} - \frac{1}{2} - \frac{d\lambda(\xi_n)}{d\xi_n}(\phi_n^T\mathbb{E}[ww^T]\phi_n - \xi_n^2) + const$$

$$- \lambda(\xi_n)(\phi_n^T\mathbb{E}[ww^T]\phi_n - 2\xi_n) = 0$$

$$\lambda(\xi) = -\frac{1}{2\xi}\left[\sigma(\xi) - \frac{1}{2}\right]$$

$$\sigma(\xi) = \frac{1}{1+e^{-\xi}}$$

$$= \sum_{n=1}^{N}\frac{1}{\sigma(\xi_n)}\left[\frac{+1}{(1+e^{-\xi})^2}\right] - \frac{1}{2} + \frac{d}{d\xi_n}\frac{1}{2\xi}\left[\sigma(\xi) - \frac{1}{2}\right](\phi_n^T\mathbb{E}[ww^T]\phi_n - \xi_n^2)$$

$$+ \frac{1}{2\xi}\left[\sigma(\xi) - \frac{1}{2}\right](\phi_n^T\mathbb{E}[ww^T]\phi_n - 2\xi) = 0$$

$$= \sum_{n=1}^{N}\frac{1}{\sigma(\xi)^3} - \frac{1}{2} + \frac{1}{2\xi^2}\left[\sigma(\xi) - \frac{1}{2}\right](\phi_n^T\mathbb{E}[ww^T]\phi_n - \xi_n^2) +$$

$$\ln\int p(t|w)p(w)\,dw = \ln\int e^{at}\sigma(-a)\cdot N(w|m_0,S_0)\,dw \;;\; a = w^T\phi$$

$$= \ln\int e^{w^T\phi t}\cdot\sigma(-w^T\phi)\cdot N(w|m_0,S_0)\,dw$$

$$\geq \ln\int h(w,\xi)p(w)\,dw$$

$$= \sum_{n=1}^{N}\ln\sigma(-w^T\phi) + w^T\phi_n t_n + (w^T\phi_n + \xi)/2$$
$$-\lambda(\xi_n)([w^T\phi_n]^2 - \xi_n^2\} + (w-m_0)^T S_0 (w-m_0)$$

$$= \sum_{n=1}^{N}\ln\sigma(-w^T\phi) + w^T\phi_n t_n - \frac{1}{2}w^T\phi_n - \xi/2 - \lambda(\xi_n)[w^T\phi_n]^2$$
$$+\lambda(\xi_n)\xi_n^2 + \frac{(w-m_0)^T S_0 (w-m_0)}{2}$$

$$= \sum_{n=1}^{N}\{\ln\sigma(-w^T\phi) - \xi/2 + \lambda(\xi)\xi_n^2\}$$
$$+ w^T\phi_n t_n - \frac{1}{2}w^T\phi_n - \lambda(\xi_n)[w^T\phi_n]^2 + \frac{(w-m_0)^T S_0 (w-m_0)}{2}$$

$$= \sum_{n=1}^{N}\{\ln\sigma(-w^T\phi) - \xi/2 + \lambda(\xi)\xi_n^2\}$$
$$+ w^T\phi_n t_n - \frac{1}{2}w^T\phi_n - \lambda(\xi_n)w^T\phi\phi^T w + \frac{w^T S_0 w}{2} - w^T S_0 m_0 + \frac{m_0^T S_0 m}{2}$$

$$= \sum_{n=1}^{N}\{\ln\sigma(-w^T\phi) - \xi/2 + \lambda(\xi)\xi_n^2\}$$
$$+ w^T\phi_n t_n - \frac{1}{2}w^T\phi_n - \lambda(\xi_n)w^T\phi\phi^T w + \frac{w^T S_0^{-1} w}{2} - w^T S_0^{-1} m_0 + \frac{m_0^T S_0^{-1} m_0}{2}$$

$$= \sum_{n=1}^{N}\{\ln\sigma(-w^T\phi) - \xi/2 + \lambda(\xi)\xi^2\} + m_0^T S_0^{-1} m_0/2$$
$$+ m_n - \lambda(\xi)w^T\phi\phi^T w + w^T S_0^{-1} w/2$$

$$= \sum_{n=1}^{N}\{\ln\sigma(-w^T\phi) - \xi/2 + \lambda(\xi)\xi^2\} + m_n - \frac{1}{2}w^T(S_0^{-1} - 2\lambda(\xi)\phi\phi) + m_0^T S_0 m_0/2$$

$$= \sum_{n=1}^{N}\{\ln\sigma(-w^T\phi) - \xi/2 + \lambda(\xi)\xi^2\} + m_n - \frac{1}{2}S_N\cdot S_N^{-1} + m_0^T S_0 m_0/2$$

$$= \sum_{n=1}^{N} \frac{1}{\sigma(\xi)} \sigma^2(\xi) \cdot e^{-\xi} - \frac{1}{2} - \frac{1}{2\xi^2}\left[\sigma(\xi) - \frac{1}{2}\right]\left(\phi_n^T \mathbb{E}[ww^T]\phi_n - \xi_n^2\right) \Big\} = 0$$

$$+\frac{1}{2\xi}\left[[\sigma^2(\xi)] \cdot \left(\phi_n^T \mathbb{E}[ww^T]\phi_n - \xi_n^2\right) + \underbrace{\left[\sigma(\xi) - \frac{1}{2}\right]}_{=0}(-2\xi_n)\right] = 0$$

$$\cancel{\frac{1}{2\xi}}\left[\left(\cancel{\tfrac{1}{2}}\right) : \left(\phi_N^T \mathbb{E}[ww^T]\phi_n - \xi_n^2\right)\right] = 0$$

$$\boxed{\xi_n^2 = \phi_N^T \mathbb{E}[ww^T]\phi_n}$$

10.34 $$\mathcal{L}(\xi) = \frac{1}{2}\ln\frac{|S_N|}{|S_0|} - \frac{1}{2} m_N^T S_N^{-1} m_N + \frac{1}{2} m_0^T S_0^{-1} m_0 + \sum_{n=1}^{N}\left\{\ln\sigma(\xi_n) - \frac{1}{2}\xi_n - \lambda(\xi_N)\xi_N^2\right\}$$

$$\frac{d}{d\alpha}\ln|A| = \mathrm{Tr}\left(A^{-1}\frac{d}{d\alpha}A\right);\; m_N = S_N\left(S_0^{-1} m_0 + \sum_{n=1}^{N}(t_n - 1/2)\phi_n\right)$$

$$S_N^{-1} = S_0^{-1} + 2\sum_{n=1}^{N}\lambda(\xi_n)\phi_n\phi_n^T$$

$$\mathcal{L}(\xi) = \ln\int h(w,\xi)p(w)\,dw = \frac{1}{2}\ln\frac{|S_N|}{|S_0|} - \frac{1}{2}m_N^T S_N^{-1} m_N + \frac{1}{2}m_0^T S_0^{-1} m_0 + \sum_{n=1}^{N}\left\{\ln\sigma(\xi) - \frac{1}{2}\xi_n - \lambda(\xi)\xi^2\right\}$$

$$\frac{d\mathcal{L}(\xi)}{d\xi} = \frac{-\sum\lambda'(\xi_n)\phi_n\phi_n^T}{|S_N|^2} + m_N^T\frac{\sum\lambda'(\xi_N)\phi_n\phi_n^T}{|S_N|}m_n + \underbrace{\sum_{n=1}\left\{\frac{1}{\sigma(\xi)}\sigma(\xi)^2 e^{-\xi} - \frac{1}{2}\right.}_{=0} - \lambda'(\xi)\xi^2 - 2\lambda(\xi)\xi\Big\}$$

$$\lambda(\xi) = \frac{1}{2\xi}\left[\sigma(\xi) - \frac{1}{2}\right]$$

$$\frac{d\mathcal{L}(\xi)}{d\xi} = \frac{+\frac{1}{\cancel{2\xi^2}}[\cancel{\sigma^2(\xi)}]\phi_n\phi_n^T}{|S_N|^{2}} + \frac{m_N^T \frac{1}{\cancel{2\xi^2}}[\cancel{\sigma^2(\xi)}]m_N}{|S_N|} = \frac{1}{\cancel{2\xi^2}}[\cancel{\sigma^2(\xi)}]\xi^2 - \frac{1}{\xi}\left[\sigma(\cancel{\xi}) - \frac{1}{2}\right]\xi = 0$$

$$\boxed{\frac{\phi_N\phi_N^T}{|S_N|^2} + \frac{m_N^T m_N}{|S_N|} = \xi^2}$$

10.35 $$\mathcal{L}(\xi) = \frac{1}{2}\ln\frac{|S_N|}{|S_0|} - \frac{1}{2}m_N^T S_N^{-1} m_N + \frac{1}{2}m_0^T S_0^{-1} m_0 + \sum_{n=1}^{N}\left\{\ln\sigma(\xi) - \frac{1}{2}\xi_n - \lambda(\xi_n)\xi_n^2\right\}$$

$$q(w) = N(w|m_0, S_0);$$

$$\ln p(t) = \ln\int p(t|w)p(w)\,dw \geq \ln\int h(w,\xi)p(w)\,dw = \mathcal{L}(\xi)$$

10.36 $p_j(D) \simeq p_{j-1}(D) Z_j$; $Z_j = \int f_j(\theta) q^{\setminus j}(\theta) d\theta$; $p(D) \simeq \prod_j Z_j$

$$\simeq \int \prod_{i-1} f_{i-1}(\theta) d\theta \cdot \hat{Z}_j = \int \prod_{j-1} f_{i-1}(\theta) d\theta \cdot \int f_j(\theta) q^{\setminus j}(\theta) d\theta$$

$$\boxed{= \prod_j Z_j}$$

10.37 $f_0(\theta)$; $p(D,\theta) = \prod_i f_i(\theta)$ <u>EP Algorithm</u>: Given: $p(D,\theta) = \prod_i f_i(\theta)$

Approximated by: $q(\theta) = \frac{1}{Z} \prod_i \tilde{f}_i(\theta)$

1. Initialize all the approximating factors $\tilde{f}_i(\theta)$
2. Initialize the posterior approximation by setting $q(\theta) \propto \prod_i \tilde{f}_i(\theta)$
3. Until convergence:
   a) Choose a factor $\tilde{f}_j(\theta)$ to refine
   b) Remove $\tilde{f}_j(\theta)$ from the posterior by division $q^{\setminus j}(\theta) = \frac{q(\theta)}{\tilde{f}_j(\theta)}$
   c) Evaluate the new posterior by sufficient statistics (moments) of $q^{new}(\theta)$ equal to $q^{\setminus j}(\theta) f_j(\theta)$ $Z_j = \int q^{\setminus j}(\theta) f_j(\theta) d\theta$
   d) Evaluate and store a new factor $\tilde{f}_j(\theta) = Z_j \frac{q^{new}(\theta)}{q^{\setminus j}(\theta)}$
4. Evaluate the approximation to the model evidence $p(D) \simeq \int \prod_i \tilde{f}_i(\theta) d\theta$

(10.219) $\rho_n = 1 - \frac{w}{Z_n} N(x_n|0, aI)$

(10.220) $v_n^{-1} = (v^{new})^{-1} - (v^{\setminus n})^{-1}$

(10.221) $m_n = m^{\setminus n} + (v_n + v^{\setminus n})(v^{\setminus n})^{-1}(m^{new} - m^{\setminus n})$

(10.222) $$s_n = \frac{Z_n}{(2\pi v_n)^{D/2} N(m_n | m^{\setminus n}, (v_n + v^{\setminus n}) I)}$$

(10.223) $$p(D) \simeq (2\pi v^{new})^{D/2} \exp(B/2) \prod_{n=1}^{N} \left\{ s_n (2\pi v_n)^{-D/2} \right\}$$

(10.224) $$B = \frac{(m^{new})^T m^{new}}{v} - \sum_{n=1}^{N} \frac{m_n^T m_n}{v_n}$$

(10.214) $$q^{\setminus j}(\theta) = \frac{q(\theta)}{\tilde{f}_j(\theta)} = \frac{N(\theta | m, vI)}{s_n N(\theta | m_n, v_n I)} = \frac{1}{s_N}\left(\frac{1}{2\pi}\right)(vI)^{-1} e^{-\frac{1}{2vI}(\theta-m)^2} \cdot e^{-\frac{(\theta-m_n)^2}{2v_n I}}$$

$$= -\frac{(\theta-m)(\theta-m)}{2vI} - \frac{(\theta-m_n)(\theta-m_n)}{2v_n I}$$

$$= -\frac{1}{2}\left[\frac{\theta^2 - 2m\theta + m^2}{vI} + \frac{\theta^2 - 2m_n\theta + m_n^2}{v_n I}\right]$$

$$= -\frac{1}{2}\left[\left(\frac{1}{v+v_n}\right)\theta^2 - 2\left(\frac{m}{vI} + \frac{m_n}{v_n I}\right)\theta + \frac{m^2}{vI} + \frac{m_n^2}{v_n I}\right]$$

$$= -\frac{1}{2}\left[\left(\frac{1}{v+v_n}\right)\theta^2 - 2(m + v^{\setminus n} \cdot v_n^{-1} m_n)\theta + m^2 + v^{\setminus n} \cdot v_n^{-1} \cdot m_n^2\right]$$

$$= -\frac{1}{2}\left(\frac{1}{v+v_n}\right)\left[\theta^2 - 2(v+v_n)(m + v^{\setminus n} \cdot v_n^{-1} \cdot m_n)\theta + (v+v_n)(m^2 + v^{\setminus n} \cdot v_n^{-1} m_n^2)\right]$$

$$= -\frac{1}{2}\left(\frac{1}{v+v_n}\right)\left[\left[\theta - (v+v_n)(m + v^{\setminus n} \cdot v_n^{-1} m_n)\right]^2 - (v+v_n)^2(m+v^{\setminus n} v_n^{-1} m_n)^2 + (v+v_n)(m^2 + v^{\setminus n} \cdot v_n^{-1} \cdot m_n^2)\right]$$

$$= -\frac{1}{2}\left(\frac{1}{v+v_n}\right)\left(\left[\theta - (v+v_n)(m + v^{\setminus n} \cdot v_n^{-1} \cdot m_n)\right]\right)^2 + \frac{1}{2}\left[(v+v_n)(m + v^{\setminus n} \cdot v_n^{-1} \cdot m_n)^2 - (m^2 + v^{\setminus n} \cdot v_n^{-1} \cdot m_n^2)\right]$$

$$\boxed{= N(\theta | m + v^{\setminus n} \cdot v_n^{-1} \cdot m_n, v^{\setminus n} + 1) \cdot \boxed{N(m | 0, I)}}$$

Form of exponential Family of Functions:

Binary: Bernoulli: $[\mathrm{Bern}(x|\mu)] = \mu^{x}(1-\mu)^{1-x}$

Binomial $[\mathrm{Bin}(m|N,\mu)] = \binom{N}{m}\mu^{m}(1-\mu)^{N-m}$; $\binom{N}{m} = \frac{N!}{(N-m)!\,m!}$

Beta Distribution $[\mathrm{Beta}(\mu|a,b)] = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}\cdot(1-\mu)^{b-1}$

Multinomials $[\mathrm{Mult}(m_1, m_2, \ldots, m_K|\mu, N)] = \binom{N}{m_1 m_2 \cdots m_K}\prod_{k=1}^{K}\mu_k^{m_k}$

$$\binom{N}{m_1 m_2 \cdots m_K} = \frac{N!}{m_1!\,m_2!\cdots m_K!}$$

Dirichlet $[\mathrm{Dir}(\mu|\alpha)] = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_K)}\prod_{k=1}^{K}\mu_k^{\alpha_k-1}$

Gaussian $[N(x|\mu,\sigma^2)] = \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$

$$[N(x|\mu,\Sigma)] = \frac{1}{(2\pi)^{D/2}}\cdot\frac{1}{|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

Any factor of exponential family is a multiplicative

10.38 Prove. (10.214) $m^{\backslash n} = m + v^{\backslash n} \cdot v_n^{-1}(m - m_n)$

(10.215) $(v^{\backslash n})^{-1} = v^{-1} - v_n^{-1}$

(10.216) $Z_n = (1-w)\,N(x_n|m^{\backslash n}, (v^{\backslash n}+1)I) + w\,N(x_n|0, aI)$

(10.217) $m = m^{\backslash n} + \rho_n \frac{v^{\backslash n}}{v^{\backslash n}+1}(x_n - m^{\backslash n})$

(10.219) $v = v^{\backslash n} - \rho_n \frac{(v^{\backslash n})^2}{v^{\backslash n}+1} + \rho_n(1-\rho_n)\frac{(v^{\backslash n})^2\|x_n - m^{\backslash n}\|^2}{D(v^{\backslash n}+1)^2}$

$$q^{new}(\theta) = q^{\backslash n}(\theta) f_n(\theta) = N(X \mid m^{\backslash n}, v^{\backslash n}+1)\, N(X \mid 0, aI)$$

$$= \frac{1}{2\pi}(v^{\backslash n}+1)^{1/2}(a)^{1/2} N(X \mid m^{\backslash n}, v^{\backslash n}+1)\, N(X \mid 0, aI)$$

$$= \frac{\sqrt{(v^{\backslash n}+1)(a)}}{2\pi} \exp\left\{\frac{-1}{2(v^{\backslash n}+1)}(x-m^{\backslash n})^2\right\} \exp\left\{-\frac{1}{2a} x^T x\right\}$$

$$= \frac{\sqrt{(v^{\backslash n}+1)(a)}}{2\pi} \exp\left\{-\frac{1}{2}\left[\frac{x^T x_n - 2xm^{\backslash n} + m^{\backslash n 2}}{v^{\backslash n}+1} + \frac{x^T x}{a}\right]\right\}$$

$$= \frac{\sqrt{(v^{\backslash n}+1)(a)}}{2\pi} \exp\left\{-\frac{1}{2}\left[x^T\left(\frac{1}{v^{\backslash n}+1} + \frac{1}{a}\right)x - 2xm^{\backslash n} + m^{\backslash n 2}\right]\right\}$$

$$= \frac{\sqrt{(v^{\backslash n}+1)(a)}}{2\pi} \exp\left\{-\frac{1}{2}\left(\frac{1}{v^{\backslash n}+1} + \frac{1}{a}\right)\left[x^T x - 2\left(\frac{a+v^{\backslash n}+1}{a(v^{\backslash n}+1)}\right) x m^{\backslash n} + \left(\frac{a+v^{\backslash n}+1}{a(v^{\backslash n}+1)}\right) m^{\backslash n 2}\right]\right.$$

$$\cancel{= \frac{\sqrt{(v^{\backslash n}+1)(a)}}{2\pi} \exp\left\{\frac{-1}{2}\right.}$$

10.39 $m = m^{\backslash n} + \rho_n \frac{v^{\backslash n}}{v^{\backslash n}+1}(x_n - m^{\backslash n})$

$$v = v^{\backslash n} - \rho_n \frac{(v^{\backslash n})^2}{v^{\backslash n}+1} + \rho_n(1-\rho_n)\frac{(v^{\backslash n})^2 \|x_n - m^{\backslash n}\|^2}{D(v^{\backslash n}+1)^2}$$

$$E[\theta] = \int q^{new}(\theta)\cdot\theta\, d\theta = \int N(\theta \mid m, vI)\cdot\theta\, d\theta = \frac{1}{(2\pi vI)^{1/2}} \int \exp^{-\frac{1}{2v}(\theta-m)^2}\cdot\theta\cdot d\theta$$

$$= \frac{e^{-\frac{1}{2v}m^2}}{(2\pi vI)^{1/2}} \int\left[e^{-\frac{1}{2v}[\theta^2 + 2m\theta]}\cdot\theta\cdot d\theta\right] = \frac{e^{-\frac{1}{2v}m^2}}{(2\pi v)^{1/2}} \frac{\sqrt{\pi}\cdot\frac{1}{v} e^{(m^2/v^2)/\frac{2}{v}}}{2\left(\frac{1}{2v}\right)^{3/2}} \quad 4\cdot\frac{1}{2v}$$

$$E[\theta^T\theta] = \int q^{new}(\theta)\cdot\theta^T\theta\, d\theta = \int N(\theta \mid m, vI)\,\theta^T\theta\, d\theta = \frac{1}{(2\pi vI)^{1/2}} \int \exp^{-\frac{1}{2v}(\theta-m)^2}\cdot\theta^T\theta\, d\theta$$

$$= \frac{e^{-\frac{1}{2v}m^2}}{(2\pi vI)^{1/2}} \int\left[e^{-\frac{1}{2v}[\theta^2+2m\theta]}\cdot\theta^T\theta\, d\theta\right] = \frac{e^{-\frac{1}{2v}m^2}}{(2\pi v)^{1/2}} \frac{\sqrt{\pi}\left[\left(\frac{m}{v}\right)^2 + \frac{1}{v}\right] e^{(m^2/v^2)/\frac{2}{v}}}{4\left(\frac{1}{2v}\right)^{5/2}}$$

$$= \frac{e^{-\frac{m^2}{2v}}}{(2\pi v)^{1/2}} \cdot \frac{\sqrt{\pi}\left(\frac{1}{v}\right) e^{\frac{m^2}{2v}}}{2\left(\frac{1}{2v}\right)^{3/2}} = \frac{\sqrt{\pi}\left(\frac{1}{v}\right)}{\sqrt{2\pi v}\cdot 2 \frac{1}{2v}\cdot\frac{1}{\sqrt{2v}}} = \frac{\sqrt{\pi}(1/v)}{\sqrt{\pi}\left(\frac{1}{v}\right)} = \boxed{1}$$

$$= \frac{e^{-\frac{m^2}{2v}}}{(2\pi v)^{1/2}} \frac{\sqrt{\pi}\left[\left(\frac{m}{v}\right)^2 + 1/v\right] e^{\frac{m^2}{2v}}}{4\left(1/2v\right)^{5/2}} = \frac{\sqrt{\pi}\left[\left(\frac{m}{v}\right)^2 + 1/v\right]}{(2\pi v)^{1/2}\cdot 4(1/2v)^{5/2}} = \frac{\sqrt{\pi}\left[\left(\frac{m}{v}\right)^2 + 1/v\right]}{(2\pi v)^{1/2}\cdot 4\cdot\frac{1}{(2v)^2}\cdot\frac{1}{\sqrt{2v}}} = \frac{\left(\frac{m}{v}\right)^2}{1/v^2} + \frac{1/v}{1/v^2}$$

$$= \boxed{m^2 + v}$$

Chapter 11

1. $\hat{f} = \frac{1}{L}\sum_{l=1}^{L} f(z^{(l)})$; $E[f] = \int f(z)p(z)\,dz = \boxed{\frac{1}{Z}\sum_{z=1}^{Z} f(z)}$

$$\mathrm{Var}[\hat{f}] = \frac{1}{L}\mathbb{E}\left[(f - \mathbb{E}[f])^2\right] = \boxed{\frac{1}{L}\mathbb{E}[f^2] - \mathbb{E}[f]^2}$$

2. $z = h(y) \equiv \int_{-\infty}^{y} p(\hat{y})\,d\hat{y}$

$$\boxed{y = h^{-1}(z) = \int_{-\infty}^{y} p(\hat{z})\,d\hat{z} = p(\hat{y})}$$

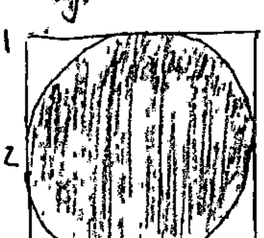
$$\frac{\cos\theta\cdot\cos\theta + \sin\theta\cdot\theta}{\cos^2\theta}$$

3. $y = f(z) = \frac{1}{\pi}\frac{1}{1+y^2}$; $z = h(y) \equiv \frac{1}{\pi}\int_{-\infty}^{y}\frac{1}{1+y^2}\,dy$; $y = \tan\theta$; $\frac{dy}{d\theta} = \sec^2\theta$

$$\equiv \frac{1}{\pi}\int_{-\infty}^{y}\frac{\sec^2\theta}{1+\tan^2\theta}\,d\theta = \frac{1}{\pi}\int_{-\infty}^{y} d\theta = \frac{1}{\pi}\tan^{-1}y\Big|_{-\infty}^{y}$$

$$= \frac{1}{\pi}\left[\tan^{-1}y - \tan^{-1}(-\infty)\right] = \boxed{\frac{1}{\pi}\left[\tan^{-1}y - \frac{\pi}{2}\right]}$$

4. Figure 11.3



1, z_2, -1, -1, z_1, 1

$$y_1 = z_1\left(\frac{-2\ln z_1}{r^2}\right)^{1/2} \qquad p(y_1, y_2) = p(z_1, z_2)\left|\frac{\partial(z_1, z_2)}{\partial(y_1, y_2)}\right| \text{ "Jacobian"}$$

$$y_2 = z_2\left(\frac{-2\ln z_2}{r^2}\right)^{1/2} \qquad = \frac{1}{\pi}\begin{vmatrix}\frac{\partial z_1}{\partial y_1} & \frac{\partial z_1}{\partial y_2}\\ \frac{\partial z_2}{\partial y_1} & \frac{\partial z_2}{\partial y_2}\end{vmatrix}$$

$$y_1 = \cos(2\pi z_2)(-2\ln z_1)^{1/2}$$

$$y_2 = \sin(2\pi z_2)(-2\ln z_1)^{1/2}$$

$$z_2 = \arctan\left(\frac{y_2}{y_1}\right);\ z_1 = \exp^{-\frac{1}{2}(y_1^2 + y_2^2)}$$

$$= \frac{1}{\pi}\begin{vmatrix} \frac{\partial}{\partial y_1} \exp^{-\frac{1}{2}(y_1^2+y_2^2)} & \frac{\partial}{\partial y_2} \exp^{-\frac{1}{2}(y_1^2+y_2^2)} \\ \frac{\partial}{\partial y_2} \arctan\left(\frac{y_2}{y_1}\right) & \frac{\partial}{\partial y_2} \arctan\left(\frac{y_2}{y_1}\right) \end{vmatrix}$$

$r^2, -\frac{y_1^2}{2}, z_1^2 \ln z_1$

$$= \frac{1}{\pi}\begin{vmatrix} -e^{-\frac{1}{2}(y_1^2+y_2^2)} y_1 & -y_2 e^{-\frac{1}{2}(y_1^2+y_2^2)} \\ \frac{-y_2/y_1^2}{1+(y_2/y_1)^2} & \frac{1/y_1}{1+(y_2/y_1)^2} \end{vmatrix}$$

$$= \frac{1}{\pi}\left[-y_1 e^{-\frac{1}{2}(y_1^2+y_2^2)} \frac{1/y_1}{1+(y_2/y_1)^2} - y_2 e^{-\frac{1}{2}(y_1^2+y_2^2)} \frac{y_2/y_1^2}{1+(y_2/y_1)^2}\right.$$

$$= \frac{1}{\pi}\left[e^{-\frac{1}{2}(y_1^2+y_2^2)} \cdot \frac{y_1^2}{y_1^2+y_2^2} - e^{-\frac{1}{2}(y_1^2+y_2^2)} \cdot \frac{y_2^2}{y_1^2+y_2^2}\right] \boxed{= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_1^2+y_2^2)}}$$

11.5. $z = N(z|0,\Sigma) = N(z|0, LL^T)$; $L = \begin{vmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{vmatrix}$; Prove $y = \mu + Lz$:

$\mathbb{E}[y] = \mathbb{E}[\mu + Lz] = \boxed{\mu}$

$\mathrm{cov}[y] = \mathbb{E}[y^T y] - \mathbb{E}[y]\mathbb{E}[y]$

$= \mathbb{E}[(\mu+Lz)(\mu+Lz)] - \mu\mu^T$

$= L^T L = \boxed{\Sigma}$

$\frac{y_1 - y_2}{y_1^2+y_2^2}$

$(a-b)(a+b)$

$a^2 - b$

11.6 $p(\text{accept}) = \int_0^{\tilde{p}(z)} \frac{1}{Kq(z)} du = \frac{\tilde{p}(z)}{Kq(z)}$

$q(z)p(\text{acceptance}) = q(z)\frac{\tilde{p}(z)}{Kq(z)} = \frac{\tilde{p}(z)}{K}$

$K\int q(z)p(\text{acceptance})\,dz = \int \tilde{p}(z)\,dz = \text{Norm}$

$$\boxed{\frac{q(z)p(\text{acceptance}|z)}{p(\text{acceptance})} = \frac{1}{Z_p}\tilde{p}(z) = p(z)}$$

11.7 $y = b\tan z + c$; $q(z) = \frac{K}{1+(z-c)^2/b^2}$; $c = a-1$, $b^2 = 2a-1$

Cauchy Distribution $p(y) = \frac{1}{\pi}\frac{1}{1+y^2}$; $Kq(z) \geq \tilde{p}(z)$; $\frac{K}{\pi}\frac{1}{1+y^2} \geq \tilde{p}(z)$

$$\frac{K}{\pi} \cdot \frac{1}{1+(b\tan z + c)^2} = \frac{K}{\pi} \frac{1}{1+b^2\tan^2 z + 2b\tan z + c^2} =$$

11.9 $q(z) = K_i \lambda_i \exp\{-\lambda_i (z - z_{i-1})\} \quad z_{i-1} < z \leq z_i$

$$p(y) = p(z) \left|\frac{dz}{dy}\right| ; \quad z = h(y) = \int_{-\infty}^{y} p(\hat{y})\, d\hat{y}$$

$$y = h(z - z_i) = \int K_i \lambda_i \exp\{-\lambda_i (z - z_i)\}\, d(z - z_i)$$

$$= -K_i \exp\{-\lambda_i (z - z_i)\}\Big|_0^{z - z_i} = -K_i - K_i \exp\{-\lambda_i (z - z_i)\}$$

$$\boxed{h(z - z_i) = K_i (1 - \exp\{-\lambda_i (z - z_i)\})}$$

$$\boxed{z - z_i = -\frac{1}{\lambda_i}\left(\ln\left\{1 - \frac{y_i}{K_i}\right\}\right)}$$

$$\boxed{Z = \sum_{i=0}^{n} \left[\frac{-1}{\lambda_i}\left(\ln\left\{1 - \frac{y_i}{K_i}\right\}\right) + z_i\right]}$$

<u>Sample Distribution</u>

$$p(z)\left|\frac{dz}{dy}\right| \underset{\text{Integral}}{\overset{\text{Jacobian}}{\rightleftarrows}} p(y)$$

<u>Algorithm:</u>

1. Test sample distribution
2. Fit distribution to $q(z) = K_i \lambda_i \exp\{-\lambda_i (z - z_i)\}$
3. Integrate quantitative Az
4. Discover changing processes.

11.10. $p(z^{(\tau+1)} = z^{(\tau)}) = 0.5$

$p(z^{(\tau+1)} = z^{(\tau)}+1) = 0.25$

$p(z^{(\tau+1)} = z^{(\tau)}-1) = 0.25$

Prove $\mathbb{E}[(z^{(\tau)})^2] = \mathbb{E}[(z^{(\tau-1)})^2] + 1/2$

and by induction $\mathbb{E}[(z^{(\tau)})^2] = \tau/2$

$$\mathbb{E}[p(z^{(\tau+1)} = z^{(\tau)})^2] = \boxed{\mathbb{E}[p(z^{(\tau-1)})] + 0.5}$$

$$\boxed{= \frac{\mathbb{E}[p(z^{(\tau-1)})] + 1}{2}}$$

$$\boxed{= \frac{\tau}{2}}$$

11.11 <u>Gibbs Sampling</u>: 1. Initialize $\{z_i : i = 1, \ldots, M\}$

2. For $\tau = 1, \ldots, T$:

- Sample $z_1^{(\tau+1)} \sim p(z_1 | z_2^{(\tau)}, z_3^{(\tau)}, \ldots, z_m^{(\tau)})$
- Sample $z_2^{(\tau+1)} \sim p(z_2 | z_1^{(\tau+1)}, z_3^{(\tau)}, \ldots, z_m^{(\tau)})$

$\vdots$

- Sample $z_j^{(\tau+1)} \sim p(z_j | z_1^{(\tau+1)}, \ldots, z_{j-1}^{(\tau+1)}, z_{j+1}^{(\tau)}, \ldots z_m^{(\tau)})$

$\vdots$

- Sample $z_M^{(\tau+1)} \sim p(z_m | z_1^{(\tau+1)}, z_2^{(\tau+1)}, \ldots z_{m-1}^{(\tau+1)}$

(11.40) $p^*(z)T(z, z^*) = p^*(z)T(z', z)$

(11.41) $\sum_{z'} p^*(z')T(z', z) = \sum_{z'} p^*(z)T(z, z') = p^*(z)\sum_{z'} p(z'|z) = p^*(z)$

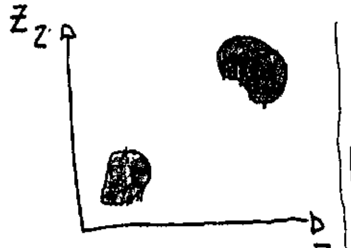
1. Initialize $\{z' : i = 1, \ldots, M\}$

2. For $\tau = 1, \ldots, T$:

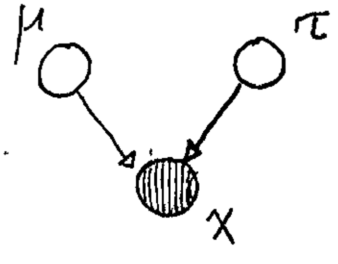
- Sample $z_1'^{(\tau+1)} \sim p(z_1 | z_2'^{(\tau)}, z_3^{(\tau)}, \ldots, z_m^{(\tau)})$
- Sample $z_2'^{(\tau+1)} \sim p(z_2 | z_1^{(\tau+1)}, z_3^{(\tau)}, \ldots, z_m^{(\tau)})$

$\vdots$

- Sample $z_j'^{(\tau+1)} \sim p(z_j | z_1^{(\tau+1)}, \ldots, z_{j-1}^{(\tau+1)}, z_{j+1}^{(\tau)}, \ldots z_m^{(\tau)}$

11.12 

The Gibbs distribution would sample correctly for Figure 11.15, and be an ergodic fit. The key condition for this sampling algorithm is seperation of z_1 condition by z_2, or converse. As Gibbs subsequent steps occur, then each infinitesimal piece of z_1 and z_2 are individualy obtained.

11.13 

$\boxed{N(x|\mu,\tau^{-1})}$ Marginal Distribution

$\boxed{N(\mu|\mu_0,s_0)\ Gam(\tau|a,b)\ Gam(\cdot|;\cdot)}$

Write down expressions for the conditional distributions $p(\mu|x,\tau)$ and $p(\tau|x,\mu)$ for $p(\mu,\tau|x)$

$$\text{If } p(\mu,\tau|x) = \frac{p(\mu,\tau,x)}{p(x)} = \frac{p(\mu)p(\tau)p(x|\mu,\tau)}{p(x)}$$

$$= \frac{N(\mu|\mu_0,s_0)Gam(\tau|a,b)\cdot N(x|\mu,\tau^{-1})}{p(x)}$$

$$\boxed{= \frac{N(\mu|\mu_0,s_0)Gam(\tau|a,b)N(x|\mu,\tau^{-1})}{p(x_1)+p(x_2)+p(x_3)+\cdots}}$$

For the Gibbs algorithm to be applied, each x_i value must be sampled. This would include $x_1, x_2, \ldots, x_n$, and the relative probability $p(x_1|x_2,x_3,\ldots x_n)$, $p(x_2|x_1,x_3,\ldots,x_n)$, etc.

11.14. $z_i' = \mu_i + \alpha(z_i - \mu_i) + \sigma_i(1-\alpha_i^2)^{1/2}\nu$; $N(z_i|\mu_i,\sigma_i)$

$= \mu_i + \alpha z_i - \alpha\mu_0 + \sigma_i(1-\alpha_i^2)^{1/2}\sigma_i$

$\boxed{= \alpha z_i - (1+\alpha)\mu_i + \sigma_i^2(1-\alpha_i^2)^{1/2}}$?

11.15 $K(r) = \frac{1}{2}\|r\|^2 = \frac{1}{2}\sum_i r_i^2$; $H(z,r) = E(z) + K(r)$; $\boxed{\frac{dz_i}{d\tau} = \frac{\partial H}{\partial r_i} = \frac{\partial K(r)}{\partial r_i} = r}$

$\boxed{\frac{\partial r_i}{\partial \tau} = -\frac{\partial H}{\partial z_i} = -\frac{\partial E(z)}{\partial z_i}}$?

11.16 $K(r) = \frac{1}{2}\|r\|^2 = \frac{1}{2}\sum_i r_i^2$; $H(z,r) = E(z) + K(r)$; $p(z,r) = \frac{1}{Z_H}\exp(-H(z,r))$

$$= \frac{1}{Z_H}\exp(-E(z) - K(r))$$

$$p(r|z) = \frac{1}{Z_H}\exp\left(-\frac{1}{2}kz^2 - \frac{1}{2}r^2\right)$$

$$\boxed{= \frac{1}{Z_H}\exp^{-\frac{1}{2}r^2} \cdot \exp^{-\frac{1}{2}kz^2}}$$

11.17

$$\cancel{\frac{1}{Z_H}}\exp(-H(R))\cancel{\delta V}\cdot \cancel{\frac{1}{2}}\min\{1, \exp(-H(R)) + H(R')\}$$
$$= \cancel{\frac{1}{Z_H}}\exp(-H(R'))\cancel{\delta V}\,\frac{1}{2}\min\{1, \exp(-H(R') + H(R))\}$$

$$\boxed{\exp(-H(R)) = \exp(-H(R'))}$$

Chapter 12:

12.1 $S = \frac{1}{N}\sum_{n=1}^{N}(x_n - \bar{x})(x_n - \bar{x}_n)^T$; $u_1^T S u_1 + \lambda_1(1 - u_1^T u_1)$; $u_{M+1}^T S u_{M+1} + \lambda_{M+1}(1 - u_{M+1}^T \cdot u_{M+1}) + \sum_{i=1}^{M}\eta_i u_{M+1}^T \cdot u_i$

$$\frac{d}{du_{M+1}} u_{M+1} S u_{M+1} + \lambda_{M+1}(1 - u_{M+1}^T u_{M+1}) + \sum_{i=1}^{M}\eta_i u_{M+1}^T u_i$$
$$= 2u_{M+1}^T S - 2\lambda_{M+1} \cdot u_{M+1} + \sum_{i=1}^{M}\eta_i u_{M+1}^T = 0$$
$$u_{M+1}^T S = \lambda_{M+1} \cdot u_{M+1}$$
$$\boxed{u_{M+1}^T \cdot S \cdot u_{M+1} = \lambda_{M+1}}$$

12.2 $J = \frac{1}{N}\sum_{n=1}^{N}\sum_{i=M+1}^{D}(x_n^T u_i - \bar{x}^T u_i)^2 = \sum_{i=M+1}^{D} u_i^T S u_i$; $u_i^T u_j = \delta_{ij}$

$\tilde{J} = \mathrm{Tr}\{\hat{U}^T S \hat{U}\} + \mathrm{Tr}\{H(I - \hat{U}^T \hat{U})\}$; $\hat{U}_{D\times(D-M)} \| u_i$

Lagrange Multipliers

$$\frac{dJ}{d\hat{U}} = \frac{d}{d\hat{U}}\left[\mathrm{Tr}\{\hat{U}^T \cdot S \cdot \hat{U}\} + \mathrm{Tr}\{H(I - \hat{U}^T \hat{U})\}\right] = \mathrm{Tr}\{2\hat{U}^T \cdot S\} - \mathrm{Tr}\{2H \cdot \hat{U}\} = 0$$

$$\boxed{\hat{U}^T \cdot S = H \cdot \hat{U}}$$

$$\hat{U}^T \cdot S \cdot \hat{U} = H$$

$$\boxed{D\times(D-M) \cdot (D-M)\times(D-M) \cdot (D-M)\times D = D\times D}$$

$$J = \mathrm{Tr}\{H\} - \mathrm{Tr}\{H\} + \mathrm{Tr}\{H \cdot I\}$$

$$\boxed{= \mathrm{Tr}\{\hat{U}^T S \cdot U\} + \mathrm{Tr}\{H(I - \hat{U}^T \hat{U})\}}$$

12.3 $u_i = \frac{1}{(N\lambda_i)^{1/2}} X^T \cdot v_i$; $u_i^T \cdot u_i = \frac{1}{(N\lambda)^{1/2}} X^T \cdot v_i \cdot \frac{1}{(N\lambda)^{1/2}} X^T \cdot v_i = \frac{1}{(N\lambda)}$

12.4 $p(z) = N(z|0,I)$; $p(z) = N(z|m,\Sigma)$; $p(x) = \int p(x|z)p(z)\,dz = \int N(x|Wz+\mu, \sigma^2 I)\cdot N(z|m,\Sigma)\,dz$

$$= \frac{1}{(2\pi)(\sigma^2 \cdot I \cdot \Sigma)^{1/2}} \int \exp^{-\frac{1}{2}(x-Wz+\mu)\frac{1}{\sigma^2}(x-Wz+\mu) - \frac{1}{2}(z-m)\Sigma^{-1}(z-m)} \cdot dz$$

$$= \frac{1}{(2\pi)(\sigma^2 \cdot I \cdot \Sigma)^{1/2}} \int \exp^{-\frac{1}{2\sigma^2}(x^2 - 2Wzx + 2\mu x + W^2 z^2 + \mu^2 - 2Wz\mu) - \frac{1}{2}(z^2 - 2mz + m^2)\Sigma} \cdot dz$$

$$= \frac{1}{(2\pi)(\sigma^2 \cdot I \cdot \Sigma)^{1/2}} \int \exp^{-\frac{1}{2\sigma^2}[(W^2z^2 - 2Wzx - 2Wz\mu) + (x^2 + 2\mu x)] - \frac{1}{2\Sigma}[(z^2 - 2mz) + m^2]} \, dz$$

$$= \frac{1}{(2\pi)(\sigma^2 \cdot I \cdot \Sigma)^{1/2}} \int \exp^{-\frac{1}{2\sigma^2}[(W^2z^2 - 2Wzx - 2Wz\mu)] - \frac{1}{2\Sigma}[(z^2 - 2mz)]} \times \exp^{-\frac{1}{2\sigma^2}[x^2 + 2\mu x] - \frac{1}{2\Sigma}m^2} \, dz$$

$$= \frac{1}{(2\pi)(\sigma^2 \cdot I \cdot \Sigma)^{1/2}} \int \exp^{-\frac{1}{2\sigma^2}[W^2 z^2] - \frac{1}{2\Sigma}[z^2] - \frac{1}{2\sigma^2}[-2Wzx - 2Wz\mu] - \frac{1}{2\Sigma}(-2mz)} \times \exp^{-\frac{1}{2\sigma^2}[x^2 + 2\mu x] - \frac{1}{2\Sigma}m^2} \, dz$$

$$= \frac{1}{(2\pi)(\sigma^2 \cdot I \cdot \Sigma)^{1/2}} \int \exp^{-\frac{1}{2}\left[\frac{W^2}{\sigma^2} - \frac{1}{\Sigma}\right]z^2 - \frac{1}{2\sigma^2}[-2Wx - 2W\mu] - \frac{1}{2\Sigma}(-2m)\,z} \times \exp^{-\frac{1}{2\sigma^2}[x^2 + 2\mu x] - \frac{1}{2\Sigma}m^2} \cdot dz$$

$$= \frac{1}{(2\pi)(\sigma^2 \cdot I \cdot \Sigma)^{1/2}} \int \exp^{-\frac{1}{2}\left[\frac{W^2}{\sigma^2} - \frac{1}{\Sigma}\right]\left(z^2 - \frac{m/\Sigma - (x+\mu)/\sigma^2}{\left(\frac{W^2}{\sigma^2} - \frac{1}{\Sigma}\right)} z\right)} \times \exp^{-\frac{1}{2\sigma^2}[x^2 + 2\mu x] - \frac{1}{2\Sigma}m^2} \cdot dz$$

$$= \frac{1}{(2\pi)(\sigma^2 \cdot I \cdot \Sigma)^{1/2}} \int \exp^{-\frac{1}{2}\left[W^T W - \frac{\sigma^2}{\Sigma}\right]} \cdots$$

$\boxed{\text{covariance}[C]}$

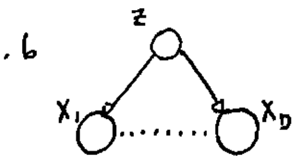
12.5 $N(x|\mu,\Sigma)$, $y = Ax + b$, $A_{M\times D}$

$\mathbb{E}[y] = \mathbb{E}[Ax+b] = \mathbb{E}[Ax] + \mathbb{E}[b] = \boxed{b}$.

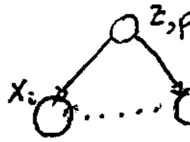
$\mathrm{cov}[y] = \mathbb{E}[y^T y] - \mathbb{E}[y]\mathbb{E}[y] = \mathbb{E}[(Ax+b)(Ax+b)] - \mathbb{E}[Ax+b]\mathbb{E}[Ax+b] = b^T b + A^T A - b^T b = \boxed{A^T A}$

M < D	M = D	M > D
b : $M\times M$ $A^T A$: $M\times M$ $N(y\|b, A^T A)$: Antisymmetric	b : $M\times M$ $A^T A$: $M\times M$ $N(y\|b, A^T A)$: Symmetric	b : $M\times M$ $A^T A$: $M\times M$ $N(y\|b, A^T A)$: Antisymmetric

12.6



"Naive Bayes Model"



$z, p(z) = N(z|0, I)$

$x_D, p(x|z) = N(x|Wz+\mu, \sigma^2 I)$

$$p(x) = \sum_{i=1}^{D} p(x|z)p(z) = \int p(x|z)p(z)\,dz = \boxed{N(x|\mu, C)}$$

12.7. $\mathbb{E}[x] = \mathbb{E}_y[\mathbb{E}_x[x|y]]$

$\text{var}[x] = \mathbb{E}_y[\text{var}_x[x|y]] + \text{var}_y[\mathbb{E}_x[x|y]]$; Derive $p(x) = N(x|\mu, C)$

$\boxed{\mathbb{E}[x] = \mathbb{E}_z[\mathbb{E}_y[x|y]] = \mathbb{E}_z[Wz+\mu] = \mu}$

$\boxed{\text{var}[x] = \mathbb{E}_z[\text{var}_x[x|z]] + \text{var}_z[\mathbb{E}_x[x|z]] = \mathbb{E}_z[C] + \text{var}_z[\mu] = C}$

12.8. $p(x|y) = N(x|\Sigma\{A^T \cdot L(y-b) + \Lambda\mu\}, \Sigma)$; $\Sigma = (\Lambda + A^T \cdot L \cdot A)^{-1}$

Prove $p(z|x) = N(z|M^{-1} \cdot W^T(x-\mu), \sigma^{-2} \cdot M)$

If $p(y|x) = N(y|Ax+b, L^{-1}) \cong p(x|z) = N(x|Wz+\mu, \sigma^2 I)$;

$$\begin{matrix} x & W & z & \mu & \sigma^2 I \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ y & A & x & b & L^{-1} \end{matrix}$$

$$p(z|x) = N(z|(I + W\sigma^{-2}W)^{-1}\{W^T \sigma^{-2}(x-\mu) + 0\}, (I + W^T\sigma^{-2}W))$$

$$\boxed{= N(z|M^{-1} \cdot W^T(x-\mu), \sigma^{-2} \cdot M)}$$

12.9 $\ln p(X|\mu, W, \sigma^2) = \sum_{n=1}^{N} \ln p(x_n|W, \mu, \sigma^2) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|C| - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T C^{-1}(x_n-\mu)$

$$\frac{d\ln p(X|\mu, W, \sigma^2)}{d\mu} = \frac{d}{d\mu}\left[\sum_{n=1}^{N}\ln p(x_n|W, \mu, \sigma^2)\right] = \frac{d}{d\mu}\left[-\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|C| - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T C^{-1}(x_n-\mu)\right]$$

$$= -[-x_n + \mu_n]C^{-1} = 0 \;;\; \boxed{x_n = \mu_n}$$

12.10 $$\frac{d^2\ln p(X|\mu, W, \sigma^2)}{d^2\mu} = -\frac{1}{C} \doteq 0$$

12.11 $\lim_{\sigma^2 \to 0} M = \lim_{\sigma^2 \to 0}(W^T W + \sigma^2 I) = W^T W$; $\mathbb{E}[z|x] = M^{-1} W_{ML}^T (x - \bar{x}_n)$ (12.41)

$= (W_{ML}^T \cdot W_{ML})^{-1} W_{ML}^T (x - \bar{x}_n)$ (12.4?)

$(W_{ML}^T)(W_{ML})^{-1} W_{ML}^T (x - \bar{x})$

$$\doteq \frac{(x - \bar{x}_n)}{U_M(L_M - \sigma^2 I)^{1/2} \cdot R}$$

(12.45) $R = I$

$U^T U = I$

$\sigma^2 = 0$

$$\boxed{= (x - \bar{x}_n) U_M^{-T} \cdot L_M^{-1/2}}$$

$L^{-1/2} U_M^T (x - \bar{x})$

12.12 For $\sigma^2 > 0$, show the posterior mean of prob-PCA is shifted towards the origin relative to orthogonal projection.

$$p(z|X) = N(z \mid M^{-1} \cdot W^T (X-\mu), \sigma^2 M), \quad W_{ML} = U(L_M - \sigma^2 I)^{1/2} \cdot R$$

Posterior mean: $$(W_{ML}^T \cdot W_{ML})^{-1} \cdot W_{ML}^T (X - \bar{X}) = \frac{\cancel{U}(L_M - \sigma^2 I)^{1/2} \cdot \cancel{R}(X - \bar{X})}{U^T \cancel{U}(L_M - \sigma^2 I) \cdot R^T \cdot \cancel{R}}$$

$$\boxed{= \frac{(X - \bar{X})}{U^T (L_M - \sigma^2 I)^{1/2} \cdot R}} \quad : \sigma^2 > 0$$

$$\boxed{= \frac{(X - \bar{X})}{U^T (L_M)^{1/2}}} \quad \sigma^2 = 0$$

12.13 Line: $W_{ML} \tilde{X} = M \mathbb{E}[z|X]$

$$(W_{ML}^T \cdot W_{ML}) \tilde{X} = W_{ML} \cdot M \mathbb{E}[z|X]$$

$$\boxed{\tilde{X} = (W_{ML}^T \cdot W_{ML})^{-1} \cdot W_{ML} \cdot M \cdot \mathbb{E}[z|X]}$$

12.14 $DM + 1 - M(M-1)/2$; $M = D-1$, $D(D-1) + 1 - (D-1)(D-2)/2 = D^2 - D + \cancel{1} - \frac{D^2}{2} + \frac{3D}{2} - \cancel{\frac{2}{2}} = \boxed{D(D+1)/2}$

$M = 0$, $D(0) + 1 - (0)(0-1)/2 = \boxed{1}$

12.15 Derive $$W_{new} = \left[\sum_{n=1}^{N} (X_n - \bar{X}) \mathbb{E}[z_n]^T\right] \left[\sum_{n=1}^{N} \mathbb{E}[z_n z_n^T]\right]^{-1} ; \quad \sigma^2_{new} = \frac{1}{ND} \sum_{n=1}^{N} \{ \|X_n - \bar{X}\|^2 - 2\mathbb{E}[z_n]^T W_{new}^T (X - \bar{X})$$

$$+ Tr(\mathbb{E}[z_n z_n^T] \cdot W_{new}^T \cdot W_{new})$$

from $$\mathbb{E}[\ln p(X, Z|\mu, W, \sigma^2)] = -\sum_{n=1}^{N} \left\{ \frac{D}{2} \ln(2\pi\sigma^2) + \frac{1}{2} Tr(\mathbb{E}[z_n z_n^T]) + \frac{1}{2\sigma^2} \|X_n - \mu\|^2 \right.$$

$$\left. - \frac{1}{\sigma^2} \mathbb{E}[z_n]^T W^T (X_n - \mu) + \frac{1}{2\sigma^2} Tr(\mathbb{E}[z_n z_n^T] \cdot W^T W) \right\}$$

$$\frac{d\mathbb{E}[\ln p(X, z|\mu, W, \sigma^2)]}{dW} = -\frac{1}{\cancel{\sigma^2}} \mathbb{E}[z_n]^T (X_n - \mu) + \frac{1}{2\cancel{\sigma^2}} Tr(\mathbb{E}[z_n z_n^T] W) = 0$$

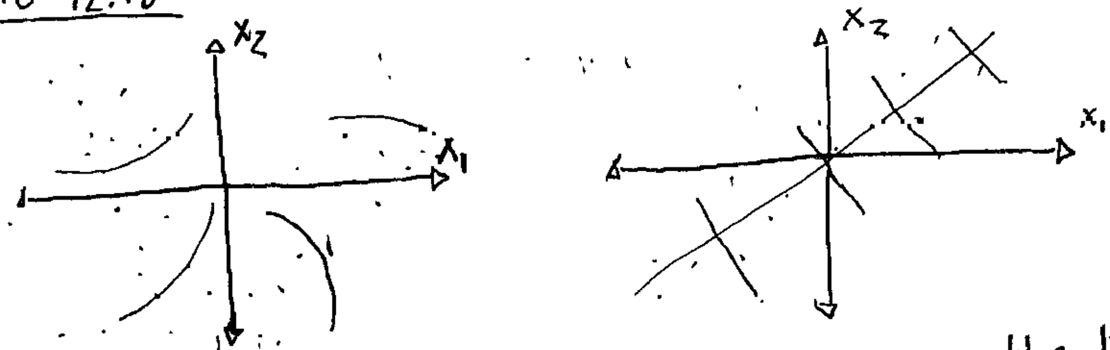
$$\boxed{W_{new} = \sum \mathbb{E}[z_n]^T (X_n - \mu) \cdot 2\, Tr(\mathbb{E}[z_n z_n^T])^{-1}}$$

$$\frac{\partial \mathbb{E}[\ln p(X, Z|\mu, W, \sigma^2)]}{d\sigma^2_{new}} = -\sum_{n=1}^{N} \left\{ \frac{D}{2} - \frac{1}{\sigma^3} \|X_n - \mu\|^2 + \frac{2}{\sigma^3} \mathbb{E}[z_n]^T W^T (X_n - \mu) - \frac{1}{\sigma^3} Tr(\mathbb{E}[z_n z_n^T] W^T W) \right\}$$

$$\boxed{\sigma^2_{new} = \frac{1}{ND} \sum_{n=1}^{N} \{ \|X_n - X\|^2 - 2\mathbb{E}[z_n]^T W_{new}^T (X_n - \bar{X}) + Tr(\mathbb{E}[z_n z_n^T] W^T W) \}}$$

12.2 ~~For $\sigma^2 > 0$; Prove the posterior mean of prob-PCA is shifted towards the origin relative to the projection~~

12.16 Figure 12.16



Derive an EM algorithm for maximizing the liklihood function for prob-PCA model.

Traditional Expectation Maximazation Algorithm:

1. Initialize μ_K, Σ_K, and mixing coefficients π_K, and evaluate log liklihood.
2. E Step: Evaluate the responsibilities using current parameter values

$$\gamma(z_{nk}) = \frac{\pi_K \cdot N(x \mid \mu_K, \Sigma_K)}{\sum_{j=1}^{K} \pi_j \cdot N(x_n \mid \mu_j, \Sigma_j)}$$

3. M Step: Re-estimate the parameters.

$$\mu_K^{new} = \frac{1}{N_K} \sum \gamma(z_{nk}) x_n$$

$$\Sigma_K^{new} = \frac{1}{N_K} \sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_K^{new})(x_n - \mu_K^{new})^T$$

$$\pi_K^{new} = \frac{N_K}{N}$$

4. Evaluate log liklihood

$$\ln p(X \mid \mu, \Sigma, \pi) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \cdot N(x_n \mid \mu_K, \Sigma_K) \right\}$$

Prob-PCA Expectation-Maximization Algorithm:

1. Initialize μ, W, and σ^2
2. E Step: Evaluate $E[z_n] = M^{-1} W^T (x_n - \bar{x})$

$$E[z_n z_n^T] = \sigma^2 \cdot M^{-1} + E[z_n] E[z_n]^T$$

3.M Step: Re-estimate the parameters

$$W_{new} = \left[\sum_{n=1}^{N}(x_n - \bar{x})\mathbb{E}[z_n]^T\right]\left[\sum_{n=1}^{N}\mathbb{E}[z_n z_n^T]\right]$$

$$\sigma^2_{new} = \frac{1}{ND}\sum_{n=1}^{N}\{\|x_n - \bar{x}\|^2 - 2\mathbb{E}[z_n]^T W_{new}^T (x_n - \bar{x})$$

$$+ Tr(\mathbb{E}[z_n z_n^T] W_{new}^T \cdot W_{new})$$

4. Evaluate log liklihood:

$$\mathbb{E}[\ln p(X, z|\mu, W, \sigma^2)] = -\sum_{n=1}^{N}\{\frac{D}{2}\ln(2\pi\sigma^2) + \frac{1}{2}Tr(\mathbb{E}[z_n z_n^T])$$

$$+ \frac{1}{2\sigma^2}\|x_n - \mu\|^2 - \frac{1}{\sigma^2}\mathbb{E}[z_n]^T W^T (x_n - \mu)$$

$$+ \frac{1}{2\sigma^2} Tr(\mathbb{E}[z_n z_n^T] W^T W)$$

12.17. $W_{D\times M}$, μ_D; $X = \{x_n\} = \{x_1, x_2, x_3, \ldots, x_n\}$, $x_n = W z_n + \mu$

$$J = \sum_{n=1}^{N}\|x_n - \mu - W z_n\|^2$$

$$\frac{dJ}{d\mu} = -2\sum_{n=1}^{N}\|x_n - \mu - W z_n\| = 0$$

$$\boxed{X = W z_n + \mu}$$

$$\boxed{(X - \bar{X}) = W(z - \bar{z}) + \mu}$$

$$\frac{dJ}{dW} = -2\sum_{n=1}^{N}\|x_n - \mu - W z_n\| \cdot z_n = 0$$

$$\cancel{-2}\sum_{n=1}^{N}\|z_n x_n - \mu \cdot z_n - W z_n^2\| = 0,$$

$$\mathbb{E}[z_n] \cdot x_n - \mu\,\mathbb{E}[z_n] - W\,\mathbb{E}[z_n z_n^T] = 0$$

$$\mathbb{E}[z_n](X - \bar{X}) = W\,\mathbb{E}[z_n z^T]$$

$$W_{old} \cdot \mathbb{E}[z_n](X - \bar{x}) = W_{old}^T W_{old}\,\mathbb{E}[z_n z_n^T]$$

$$\boxed{\frac{W_{old} \cdot \mathbb{E}[z_n](X - \bar{X})}{(W_{old} \cdot W_{old})} = \mathbb{E}[z_n z_n^T]}$$

$$\frac{\partial J}{\partial z_i} = -2\sum_{n=1}^{N}\|x_n - \mu - W z_n\| W = 0$$

$$W(x - \mu) = W^T W z_n$$

$$\boxed{\frac{W(x - \mu)}{W^T W} = z_n}$$

12.18 $\sum_{i=1}^{D}\sum_{j=1}^{D} \Psi_{ij}$

12.19. $C = WW^T + \Psi$; $\boxed{\Psi = C - W \cdot W^T}$

12.20. $\frac{\partial^2 p(x|z)}{\partial^2 \mu} = \frac{\partial^2}{\partial^2 \mu}\left[N(x|Wz+\mu, \Psi)\right] = \frac{\partial^2}{\partial \mu} \frac{1}{(2\pi\Psi)^{D/2}} e^{-\frac{1}{2}(x-Wz-\mu)^2/\Psi}$

$= \frac{\partial}{\partial \mu}\left[\frac{+2/2(x-Wz-\mu)}{(2\pi\Psi)^{D/2}} \cdot e^{-\frac{1}{2}(x-Wz-\mu)^2/\Psi}\right]$

$= \frac{-e^{-\frac{1}{2}(x-Wz-\mu)^2/\Psi}/\Psi - \Psi(x-Wz-\mu)^2 e^{-\frac{1}{2}(x-Wz-\mu)^2/\Psi}/\Psi^2}{(2\pi\Psi)^{D/2}}$

$= -\frac{e^{-\frac{1}{2}(x-Wz-\mu)^2/\Psi}}{(2\pi\Psi)^{D/2}\cdot \Psi} - \frac{(x-Wz-\mu)^2 \cdot e^{(x-Wz-\mu)^2/\Psi}}{(2\pi\Psi)^{D/2}\cdot \Psi^2} = 0$

$1 = \frac{(x-Wz-\mu)^2}{\Psi}$; $\Psi = (x-Wz-\mu)^2$

$\Psi = x^2 - 2Wzx + W^2z^2 + 2Wz\mu - 2x\mu + \mu^2$

$= \frac{\partial^2}{\partial^2 \mu}\left[\log N(x|Wz+\mu, \Psi)\right]$

$= \frac{\partial^2}{\partial^2 \mu}\left[-\frac{D}{2}\log 2\pi\Psi - \frac{1}{2}(x-Wz-\mu)^2/\Psi\right]$

$= \frac{\partial}{\partial \mu}\left[(x-Wz-\mu)/\Psi\right] = -\frac{1}{\Psi} = 0$ "Undefined..."

12.21. Derive $\mathbb{E}[z_n] = GW^T\Psi^{-1}(x_n - \bar{x})$

$\mathbb{E}[z_n z_n^T] = G + \mathbb{E}[z_n]\mathbb{E}[z_n]^T$; $G = (I + W^T \Psi^{-1} W)^{-1}$

Prob-PCA:

$\mathbb{E}[z_n] = M^{-1} \cdot W^T(x_n - \bar{x})$

$\mathbb{E}[z_n z_n^T] = \sigma^2 M^{-1} + \mathbb{E}[z_n]\mathbb{E}[z_n]^T$

Factor Analysis:

$\boxed{\mathbb{E}[z_n] = \frac{W^T(x_n - \bar{x})}{\frac{\Psi(\Psi + W^TW)}{\Psi}}}$

$\boxed{\mathbb{E}[z_n z_n^T] = \frac{1}{\Psi(\Psi + W^TW)} + \mathbb{E}[z_n]\mathbb{E}[z_n]^T}$

E & M for Student T-Distribution: $St(x|\mu,\lambda,\nu) = \frac{\Gamma(\nu/2+1/2)}{\Gamma(\nu/2)}\left(\frac{\lambda}{\pi\nu}\right)^{1/2}\left[1+\frac{\lambda(x-\mu)^2}{\nu}\right]^{-\frac{\nu}{2}-\frac{1}{2}}$

1. Initialize μ, λ, ν
2. E Step: Evaluate the responsibilities: $\gamma(z_{nk}) = \frac{\pi_k \cdot St(x|\mu,\lambda,\nu)}{\sum_{n=1}^{N} \pi_j \cdot St(x|\mu,\lambda,\nu)}$
3. M Step: Re-estimate the parameters, μ, λ, ν

$$\mu_k = \frac{1}{N}\sum \gamma(z_{nk}) \cdot x_n$$
$$\lambda = a/b$$
$$\nu = 2a$$

4. Evaluate log liklihood;

$$\ln p(x|\mu,\lambda,\nu) = \sum_{n=1}^{N} \ln\{\pi_k \cdot St(x|\mu,\lambda,\nu)\}$$

12.25 $p(z) = N(x|0,I)$, $p(x|z) = N(x|Wz+\mu,\phi)$, $x \to Ax$, $A_{D\times D}$

$$\ln p(X,z) = \ln p(x|z)p(z)$$
$$= \ln N(Ax|Wz+\mu,\phi)\cdot N(x|0,I)$$
$$= +\frac{D}{2}\ln 2\pi\phi - \frac{1}{2\phi}(Ax-Wz-\mu)^2 - \frac{1}{2}x^2$$

$\mu_{ML} \cong A\mu_{ML}$
$W_{ML} \cong AW_{ML}$
$\Phi_{ML} \cong A\phi^{-1}A^T$

$$\frac{d\ln p(X,z)}{d\mu} = \frac{(Ax-Wz-\mu)}{\phi} = 0;\ \mu = Ax - Wz$$

$$\frac{d\ln p(X,z)}{dW} = \frac{(Ax-Wz-\mu)\cdot z}{\phi} = 0;\ W = \frac{Ax-\mu}{\phi z}$$

$$\frac{d\ln p(X,z)}{d\phi} = \frac{-D}{2\phi} + \frac{1}{2\phi^2}(Ax-Wz-\mu)^2 = 0;\ \phi = \frac{2(Ax-Wz-\mu)^2}{D}$$

$$L(\mu,W,\phi) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|WW^T+\phi| - \frac{1}{2}\sum_{n=1}^{N}\{(x_n-\mu)^T(WW^T+\phi)^{-1}(x_n-\mu)\} \quad 12.43$$

$$\frac{dL(\mu,W,\phi)}{d\mu} = \sum_{n=1}^{N}(Ax-\mu) = 0;\ \mu_{ML} = \sum_{n=1}^{N} Ax$$

$(Ax-A\mu)(WW^T+\phi)^{-1}(Ax-A\mu)$
$(WW^T+\phi)A\cdot A^T(x-\mu)^T\cancel{(x-\mu)}\, S$

$$L(\mu,W,\phi) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|WW^T+\phi| - \frac{1}{2}\sum_{n=1}^{N}\{(Ax_n-A\mu)(WW^T+\phi)^{-1}(Ax_n-A\mu)\}$$
$$= -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|WW^T+\phi| - \frac{1}{2}\sum_{n=1}^{N}\{A^TSA(WW^T+\phi)^{-1}\}$$
$$= -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|WW^T+\phi| - \frac{1}{2}\sum_{n=1}^{N}\{A^TSA^{-1}(W_{ML}W_{ML}^T+\phi)\}$$

12.26 $K a_i = \lambda_i N a_i$ & $K^2 a_i = \lambda_i N K a_i$; $a_i = \langle a_1, a_2, \cdots, a_i \rangle$

$$\frac{v_i}{\lambda_i} = \sum \phi(x_n) \; ; \; K = \frac{1}{N}\sum_{n=1}^{N} k(x_1, x_n) = \frac{1}{N}\sum_{n=1}^{N} \phi(x_n)\phi(x_n)^T = \boxed{\frac{1}{N}\sum_{n=1}^{N}\left(\frac{v_i}{a_i}\right)\left(\frac{v_i}{a_i}\right)^T}$$

$$K^n = \left[\frac{1}{N}\sum_{n=1}^{N} K(x_1, x_n)\right]^n = \boxed{\left[\frac{1}{N}\sum_{n=1}^{N} \phi(x_n)\phi(x_n)\right]^n}$$

$$y_i(x) = \phi(x)^T v_i = \sum_{n=1}^{N} a_{in}\phi(x)^T\phi(x_n) = \sum_{n=1}^{N} a_{in} K(x, x_n)$$

$$y_i^n(x) = \left[\phi(x)^T v_i\right]^n = \left[\sum_{n=1}^{N} a_{in}\phi(x)^T\phi(x_n)\right]^n = \boxed{\left[\sum_{n=1}^{N} a_{in} K(x, x_n)\right]^n}$$

$$y_i(x) + y_i(x) = 2y_i(x) = \boxed{2\sum_{n=1}^{N} a_{in} K(x, x_n)}$$

12.27 $k(x, x') = x^T x$

$$\frac{1}{N}\sum_{n=1}^{N} k(x, x')\sum_{m=1}^{m} a_{im} k(x, x') = \lambda_i \sum_{n=1}^{N} a_{in} k(x, x')$$

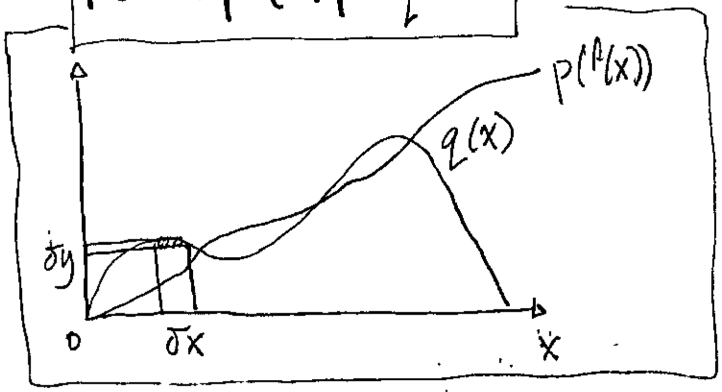
$$\frac{1}{N} x^T x \cdot x^T x \sum_{n=1}^{N} a_{im} = \lambda_i \sum_{n=1}^{N} a_{in} x^T x \;=\; K^2 a_i = \lambda N K a_i$$

$$\boxed{K a_i = \lambda N a_i}$$

12.28 $p_y(y) = p_x(x)\left|\frac{dx}{dy}\right| = p_x(g(y))\,|g'(y)|$

$$p_y(y) = q(x)\left|\frac{dx}{dy}\right| \; ; \; p_y(f(x)) = q(x)\left|\frac{dx}{dy}\right| \; ; \; p_y(f(x))\frac{dy}{dx} = q(x)$$

$$\boxed{p(f(x))\,|f'(x)| = q(x)}$$



12.29. $p(z_1, z_2) = p(z_1)p(z_2)$;

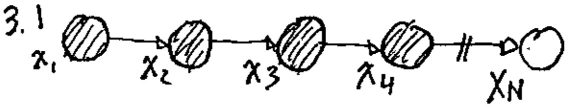
$$\cancel{S = \frac{1}{N}\sum_{n=1}^{N} p(z_1, z_2) = \frac{1}{N}\sum_{n=1}^{N} p(z_1)p(z_2)}$$

$$\mathrm{cov}[z_1, z_2] = \iint (z_1 - \bar{z}_1)(z_2 - \bar{z}_2)\,p(z_1, z_2)\,dz_1\,dz_2$$
$$= \iint (z_1 - \bar{z}_1)(z_2 - \bar{z}_2)\,p(z_1)p(z_2)\,dz_1\,dz_2$$
$$= \int (z_1 - \bar{z}_1)p(z_1)\,dz_1 \cdot \int (z_2 - \bar{z}_2)p(z_2)\,dz_2 = 0$$

$$\cancel{S = \frac{1}{N}\sum_{n=1}^{N} p(y_1, y_2) = \frac{1}{N}\sum_{n=1}^{N} p(y_1)\cdot p(y_2|y_1) = \frac{1}{N}\sum_{n=1}^{N} p(y_1)\,p(y_1^2|y_1)}$$

$$\mathrm{cov}[y_1, y_2] = \iint (y_1 - \bar{y}_1)(y_2 - \bar{y}_2)\,p(y_1, y_2)\,dy_1\,dy_2 = \iint y_1(y_2 - \bar{y}_2)\,p(y_2|y_1)\,p(y_1)\,dy_1\,y_2 = \iint (y_1^3 - y_1\bar{y}_2)\,p(y_2|y_1)\,p(y_1)\,dy_1\,dy_2$$
$$= 0$$

Chapter 13:

13.1 

<u>D-seperation</u>: $p(D|\mu)=\prod_{n=1}^{N} p(x_n|\mu)$

(13.3) $p(x_1,\dots,x_N)=p(x_1)\prod_{n=2}^{N} p(x_n|x_{n-1})$

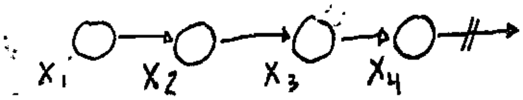
$m<n-2$ $=p(x_1)\prod_{n=2}^{N} p(x_n|x_1,\dots,x_{n-1})$

$$\boxed{=p(x_1)\cdot p(x_2)\prod_{n=3}^{N} p(x_n|x_1,\dots,x_{n-1})}$$

13.2 $p(x_1,\dots,x_n)=p(x_1)\prod_{n=2}^{N} p(x_n|x_{n-1})$

<u>Product Rule</u>
$p(X,Y)=p(Y|X)p(X)$

<u>Sum Rule</u>
$p(X)=\sum_Y p(X,Y)$



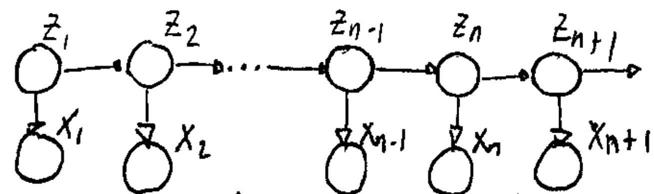
$$\boxed{\frac{p(x_1,x_2,\dots,x_n)}{p(x_1)}=p(x_n|x_1,\dots,x_{n-1})=p(x_n|x_{n-1})}$$

$$\boxed{p(x_1,\dots,x_n)=p(x_1)p(x_2|x_1)\prod_{n=3}^{N} p(x_n|x_{n-1},x_{n-2})}$$

$$\boxed{\frac{p(x_1,\dots,x_n)}{p(x_1)p(x_2|x_1)}=p(x_n|x_{n-1},x_{n-2})}$$

13.3 <u>D-seperation</u>: $p(D|\mu)=\prod_{n=1}^{N} p(x_n|\mu)$

$p(x_1,\dots,x_n)$



$$\boxed{p(X|Z)\neq\prod_{n=1}^{N} p(x_n|z_n)=\sum_{n=1}^{N} p(x_n|z_n)}$$

13.4 $p(X|Z,W)$ <u>Linear Regression Model</u>: $\sum_{n=1}^{N} W_i X_i =$

<u>Nueral Network Model</u>: $F\left(\sum_{j=1}^{M} w_j \phi_j(x)\right)$

<u>Hidden Markov Model</u>:

$p(z_n|z_{n-1},A)=\prod_{k=1}^{K}\prod_{j=1}^{K} A_{jk}^{z_{n-1,j}\cdot z_{nk}}$

A liner regression or neural network model under log maximum liklihood would enable solving for $z_{n-1,j}\cdot z_{nk}$, which would be W_i and W_j.

13.5 $\pi_k = \dfrac{\gamma(z_{1k})}{\sum_{j=1}^{K} \gamma(z_{1j})}$

$A_{jk} = \dfrac{\sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nk})}{\sum_{l=1}^{K}\sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nl})}$

$$Q(\theta,\theta^{old}) = \sum_{k=1}^{K} \gamma(z_{1k}) \ln \pi_k + \sum_{n=2}^{N}\sum_{j=1}^{K}\sum_{k=1}^{K} \xi(z_{n-1,j}, z_{nk}) \ln A_{jk} + \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \ln p(x_n|\phi_k)$$

$$\frac{dQ(\theta,\theta^{old})}{d\pi_k} = \frac{d}{d\pi_k}\left[\sum_{k=1}^{K} \gamma(z_{1k}) \ln \pi_k + \lambda\left(1 - \sum \pi_k\right)\right] = 0$$

$$\sum_{k=1}^{K} \gamma(z_{1k}) \cdot \frac{1}{\pi_k} - \lambda = 0$$

$$\lambda = \sum \gamma(z_{1k}) \cdot \frac{1}{\pi_k}$$

$$\sum_{k=1}^{K} \gamma(z_{1k}) - \sum \gamma(z_{1k}) \cdot \frac{1}{\pi_k} = 0$$

$$\sum_{k=1}^{K} \gamma(z_{1k}) \cdot \frac{1}{\pi_k} = \sum \gamma(z_{1k})$$

$$\boxed{\frac{\sum_{k=1}^{K} \gamma(z_{1k})}{\sum \gamma(z_{1k})} = \pi_k}$$

$$\frac{dQ(\theta,\theta^{old})}{dA_{jk}} = \sum\sum\sum \xi(z_{n-1,j}, z_{nk}) \cdot \frac{1}{A_{jk}} + \frac{d}{dA_{jk}}\left[\lambda\left(1 - \sum_k A_{jk}\right)\right]$$

$$= \sum_{n=2}^{N}\sum_{j=1}^{K}\sum_{k=1}^{K} \xi(z_{n-1,j}, z_{nk}) \cdot \frac{1}{A_{jk}} - \lambda = 0$$

$$\lambda = \sum_{n=2}^{N}\sum_{j=1}^{K}\sum_{k=1}^{K} \xi(z_{n-1,j}, z_{nk}) \cdot \frac{1}{A_{jk}}$$

$$\sum_{n=2}^{N}\sum_{j=1}^{K}\sum_{k=1}^{K} \xi(z_{n-1,j}, z_{nk}) \cdot \frac{1}{\cancel{A_{jk}}} - \sum_{n=2}^{N}\sum_{j=1}^{K}\sum_{k=1}^{K} \xi(z_{n-1,j}, z_{nk}) \cdot \frac{1}{A_{jk}^{2}} = 0$$

$$\boxed{A_{jk} = \frac{\sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nk})}{\sum_{l=1}^{K}\sum_{n=2}^{N} \delta(z_{n-1,j}, z_{nk})}}$$

$$= \sum\sum \gamma(z_{nk})\left[-\frac{1}{\cancel{\Sigma}} + \frac{1}{2\Sigma^2}(x-\mu)^2\right] = 0$$

$$\sum\sum \gamma(z_{nk}) = \frac{\sum\sum \gamma(z_{nk})}{2\Sigma}(x-\mu)^2$$

$$\boxed{\Sigma = \frac{\sum\sum \gamma(z_{nk})(x-\mu)(x-\mu)}{\sum\sum \gamma(z_{nk})}}$$

13.8 $p(x|z) = \prod_{i=1}^{D}\prod_{k=1}^{K} \mu_k^{x_i \cdot z_k}$; $\sum_{i=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \ln p(x|z) = \sum_{i=1}^{N}\sum_{k=1}^{K} \gamma($

$= \sum_{i=1}^{N}\sum_{k=1}^{K} \gamma(z$

$$\frac{d}{d\mu}\left[\sum_{i=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \sum_{i=1}^{D} x_i \ln \mu_{ki} + \sum^{K} \lambda\right.$$

$$\left[\sum_{i=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \sum_{i=1}^{D} x_i / \mu_{ki} + \sum_{k=1}^{K} \lambda\right.$$

$$\lambda_k = -\sum_{n=1}^{N} \gamma(z_{nk})$$

$$\sum_{i=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \ln p(x|z) = \sum_{i=1}^{N}\sum_{k=1}^{K} \gamma(z$$

$$\frac{d}{d\mu_k}\left[\sum_{i=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \sum_{i=1}^{D} \{x_i \ln \mu_{ki} + (1-x_n\right.$$

$$\sum_{i=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \left\{\sum_{i=1}^{D} x_i / \mu_{ki} - \frac{(1-x_{ni})}{1-\mu_k}\right\} =$$

$$\sum_{i=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \sum_{i=1}^{D} \frac{x_i}{\mu_{ki}} = \frac{(1-x_{ni})}{1-\mu_k}$$

$$\sum_{i=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \sum_{i=1}^{D} x_i (1-\mu_k) = \mu_k - \mu$$

$$\sum_{i=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) x_i - \sum_{i=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) x_i \mu_k =$$

$$\sum_{i=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) x_i = \mu_k\left[\sum_{i=1}^{N}\sum_{k=1}^{K} x_i - x_i\right.$$

$$\boxed{\frac{\sum_{i=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) x_i}{\sum \gamma(z_{nk})} = \mu_k}$$

13.6 <u>Expectation & Maximization Algorithm:</u>

1. Initialize π_k, and A_{jk} to zero

$$Q(\theta, \theta^{old}) = \sum_{k=1}^{K} \gamma(z_{1k}) \ln \theta + \sum_{n=2}^{N} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi(z_{n-1,j}, z_{nk}) \ln \theta$$

$$+ \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \ln p(x_n \mid \phi_k)$$

$$= \text{Undefined}$$

$$\boxed{\pi_k = \frac{\gamma(z_{1k})}{\sum_{j=1}^{K} \gamma(z_{1j})} = 0 \text{ ; then } \gamma(z_k) = 0}$$

$$\boxed{A_{jk} = \frac{\sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nk})}{\sum_{l=1}^{K} \sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nl})} = 0 \text{; then } \sum_{n=2}^{N} \xi(z_{n-1,j}\, z_{nk}) = 0}$$

13.7 Prove $Q'(\theta, \theta^{old})$ for μ_k and Σ_k

$$\frac{dQ(\theta, \theta^{old})}{d\mu_k} = \frac{d}{d\mu_k}\left[\sum_{k=1}^{K} \gamma(z_{1k}) \ln \pi_k + \sum_{n=2}^{N} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi(z_{n-1,j}, z_{nk}) \ln A_{jk} + \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \ln N(x \mid \mu_k, \Sigma) \right]$$

$$= \frac{d}{d\mu_k}\left[\sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \left(\tfrac{-1}{2} \ln 2\pi\Sigma - \frac{1}{2\Sigma}(x - \mu_k)^2 \right) \right]$$

$$= \mp \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \cdot \frac{1}{2\Sigma}(x - \mu_k) = 0$$

$$\sum \sum \gamma(z_{nk}) \cdot x = \sum \sum \gamma(z_{nk}) \cdot \mu$$

$$\boxed{\frac{\sum \sum \gamma(z_{nk}) \cdot x}{\sum \sum \gamma(z_{nk})} = \mu}$$

$$\frac{dQ(\theta, \theta^{old})}{d\Sigma_k} = \frac{d}{d\Sigma_k}\left[\sum \sum \gamma(z_{nk}) \left(-\ln 2\pi\Sigma - \frac{1}{2\Sigma}(x - \mu_k)^2 \right) \right].$$

13.9 $p(X|z_n)=p(x_1,\dots,x_n|z_n)$

$p(x_{n+1},\dots,x_n|z_n)$

$p(x_1,\dots,x_{n-1}|x_n,z_n)=p(x_1,\dots,x_{n-1}|z_n)$

$p(x_{n+1},\dots,x_{n-1}|z_1,\dots,z_{n-1})=p(x_{n+1},\dots,x_n|\dots z_{n-1})$

$p(x_{n+1},\dots,x_n|z_n,z_{n+1})=p(x_{n+1},\dots,x_n|z_{n+1})$

$p(x_{n+2},\dots,x_n|z_{n+1},x_{n+1})=p(x_{n+2},\dots,x_n|z_{n+1})$

$p(X|z_{n-1},z_n)=p(x_1,\dots,x_{n-1}|z_{n-1})$

$p(x_n|z_n)p(x_{n+1},\dots,x_n|z_n)$

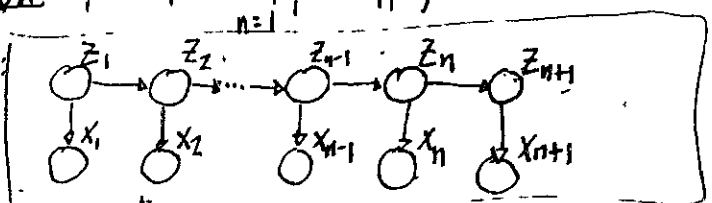
$p(x_{n+1}|X,z_{n+1})=p(x_{n+1}|z_{n+1})$

$p(z_{n+1}|z_n,X)=p(z_{n+1}|z_n)$

Joint Distribution Model: $p(x_1,\dots,x_n,z_1,\dots,z_n)=p(z_1)\prod_{n=2}^{N}p(z_n|z_{n-1})\cdot\prod_{n=1}^{N}p(x_n|z_n)$

D-seperation: $p(D|\mu)=\prod_{n=1}^{N}p(x_n|\mu)$

Figure 13.5:



$\cancel{\mu_{f_s\to x}(x)\equiv\sum_{X_s}F_s(x,X_s)\;;\;\alpha(z_n)=\mu_{f_n\to z_n}(z_n)}$

$\equiv\sum_{X_s}F_n(z_n,\{z_1,\dots,z_{n-1}\})=h(z_1)\prod_{i=2}^{n}f_i(z_i,z_{i-1})$

$\cancel{h(z_1)=p(z_1)p(x_1|z_1)\quad f_n(z_{n-1},z_n)=p(z_n|z_{n-1})p(x_n|z_n)}$

13.10 Sum Rule: $p(x) = \sum p(X, Y)$ Product Rule: $p(X, Y) = p(Y|X)p(X)$

- $p(X|z_n) = p(x_1, \dots, x_n|z_n)\,p(x_{n+1}, \dots, x_N|z_n)$

Sum Rule: $\sum_{i=1}^{N} p(x_i|z_i)\,p(x_{n+1}, x_n|z_n)$, Product Rule: $p(x_1, \dots, x_n|z_n)\,p(x_{n+1}|z_n)\,p(x_n|z_n)$

- $p(x_1, \dots, x_{n-1}|x_n, z_n) = p(x_1, \dots, x_{n-1}|z_n)$

Sum Rule: $\sum_{i=1}^{N-1} p(x_i|z_i)x_i$ Product Rule: $p(x_1, \dots x_{n-2}|z_n)\,p(x_{n-1}|z_n)$

- $p(x_{n+1}, \dots, x_{n-1}|z_{n-1}, z_n) = p(x_1, \dots, x_{n-1}|z_{n-1})$

Sum Rule: $\sum_{i=1}^{N-1} p(x_i|z_i)$ Product Rule: $p(x_1, \dots, x_{n-2}|z_{n-1})\,p(x_{n-1}|z_{n-1})$

- $p(x_{n+1}, \dots, x_n|z_n, z_{n+1}) = p(x_{n+1}, \dots, x_n|z_{n+1})$

Sum Rule: $\sum_{i=n}^{n+1} p(x_i|z_{n+1})$ Product Rule: $p(x_{n+1}|z_{n+1})\,p(x_n|z_{n+1})$

- $p(x_{n+2}, \dots, x_n|z_{n+1}, x_{n+1}) = p(x_{n+2}, \dots, x_n|z_{n+1})$

Sum Rule: $\sum_{i=n+2}^{n} p(x_i|z_{n+1})$ Product Rule: $p(x_{n+2}, x_{n+1}|z_{n+1})\,p(x_n|z_{n+1})$

- $p(X|z_{n-1}, z_n) = p(x_1, \dots, x_{n-1}|z_{n-1})$ Error - Not Rewriting

Sum Rule: $\sum_{i=1}^{N-1} p(x_i|z_{n-1})$ Product Rule: $p(x_1 \dots x_{n-2}|z_{n-1})\,p(x_{n-1}|z_{n-1})$

- $p(x_{n+1}|X, z_{n+1}) = p(x_{n+1}|z_{n+1})$

Sum Rule: $p(x_{n+1}|z_{n+1})$ Product Rule: $p(x_{n+1}|z_{n+1})$

- $p(z_{n+1}|z_n, X) = p(z_{n+1}|z_n)$

Sum Rule: $p(z_{n+1}|z_n)$ Product Rule: $p(z_{n+1}|z_n)$

13.11 $p(x_s) = f_s(x_s) \prod_{i \in ne(f_s)} \mu_{x_i \to f_s}(x_i)$ Derive: $\dfrac{\alpha(z_{n-1})p(x_n|z_n)p(z_n|z_{n-1})\beta(z_n)}{p(X)}$

$$\xi(z_{n-1}, z_n) = p(z_{n-1}, z_n|X)$$
$$= \frac{p(X|z_{n-1}, z_n)\,p(z_{n-1}, z_n)}{p(X)}$$

$$f_s(X) = p(X|z_{n-1}, z_n) = p(x_1, \dots, x_{n-1}|z_{n-1})$$

$$\boxed{= \frac{f_s(X)\cdot p(x_n|z_n)p(x_{n+1},\dots,x_N|z_n)p(z_n|z_{n-1})p(z_{n-1})}{p(X)} = \frac{\alpha(z_{n-1})p(x_n|z_n)\cdot p(z_n|z_{n-1})\beta(z_n)}{p(X)}}$$

13.12 $X^{(r)} = \{r_1, \dots r_K\}$

Hidden Markov Model: $Q(\theta, \theta^{old}) = \sum_{k=1}^{K} \gamma(z_{1k}) \ln \pi_k + \sum\sum\sum \xi(z_{n-1,j}, z_{nk}) \ln A_{jk}$
$$+ \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \ln p(x_n|\phi)$$

$$\alpha(z_n) = p(x_n|z_n) \sum_{z_{n-1}} \alpha(z_{n-1}) p(z_n|z_{n-1})$$

$$\beta(z_n) = p(x_{n+1}, \dots, x_N|z_n)$$

$$\frac{dQ(\theta,\theta^{old})}{d\pi} = \frac{d}{d\pi}\Big[\sum_{k=1}^{K}\gamma(z_{1k})\ln\pi_k + \sum\sum\sum\xi(z_{n-1,j}, z_{nk})\ln A_{jk} + \sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk})\ln p(x_n|\phi) + \lambda\Big[\sum \pi_k - 1\Big]\Big] = 0$$

$$= \lambda = \sum \gamma(z_{1k})$$

$$\boxed{\pi_k = \frac{\sum \gamma(z_{1k})}{\sum\sum \gamma(z_{1j})}}$$

$$\frac{dQ(\theta,\theta^{old})}{dA_{jk}} = \frac{d}{dA_{jk}}\Big[\sum_{k=1}^{K}\gamma(z_{nk})\ln\pi_k + \sum\sum\sum\xi(z_{n-1,j}, z_{nk})\ln A_{jk} + \sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk})\ln p(x_n|\phi) + \lambda\Big[\sum A_{jk} - 1\Big]\Big] = 0$$

$$\lambda = \sum\sum\sum\xi(z_{n-1,j}, z_{nk})\frac{1}{A_{jk}}$$

$$\boxed{A_{jk} = \frac{\sum\sum\sum \xi(z_{n-1,j}, z_{nk})}{\sum\sum\sum \xi(z_{n-1,j}, z_{nl})}}$$

$\alpha(z_n) = p(z_1|x_1,\dots,x_n)\,p(x_1,\dots,x_n) = \left(\prod_{m=1}^{n} c_m\right)\hat{\alpha}(z_n)$

$\beta(z_n) = \left(\prod_{m=n+1}^{N} c_m\right)\hat{\beta}(z_n)$

$\boxed{\gamma(z_n) = \hat{\alpha}(z_n)\hat{\beta}(z_n)}$ $\boxed{\xi(z_{n-1}, z_n) = c_N\,\hat{\alpha}(z_{n-1})\,p(x_n|z_n)\,p(z_n|z_{n-1})\,\hat{\beta}(z_n)}$

13.16 $p(x_1,\dots,x_N, z_1,\dots,z_n) = p(z_1)\left[\prod_{n=2}^{N} p(z_n|z_{n-1})\right]\prod_{n=1}^{N} p(x_n|z_n)$

$\ln p(x_1,\dots,x_N, z_1,\dots,z_n) = \ln p(z_1)\sum_{n=2}^{N} p(z_n|z_{n-1})\sum_{n=1}^{N} p(x_n|z_n)$

$\mu_{f\to x}(x) = \max_{x_1,\dots x_m}\left[\ln f(x, x_1,\dots x_m) + \sum_{m\in ne(f_s)\setminus x} \mu_{x_m\to f}(x_m)\right]$

$f_n(z_{n-1}, z_n) = p(z_n|z_{n-1})\,p(x_n|z_n)$

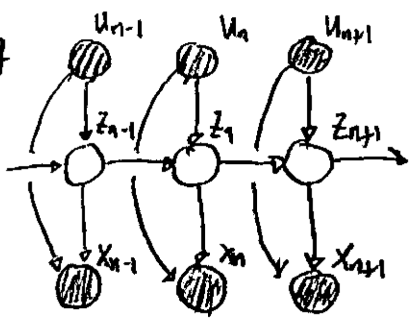
$\mu_{z_{n-1}\to f_n}(z_{n-1}) = \mu_{f_{n-1}\to z_{n-1}}(z_{n-1})$; $\mu_{f_n\to z_n}(z_n) = \sum f_n(z_{n-1}, z_n)\,\mu_{z_{n-1}\to f_n}(z_{n-1})$

Δ $\ln p(x_1,\dots,x_n, z_1,\dots,z_n) = \ln \mu_{f_n\to z_n}(z_n) = \ln\sum f_n(z_{n-1}, z_n)\,\mu_{z_{n-1}\to f_n}(z_{n-1})$

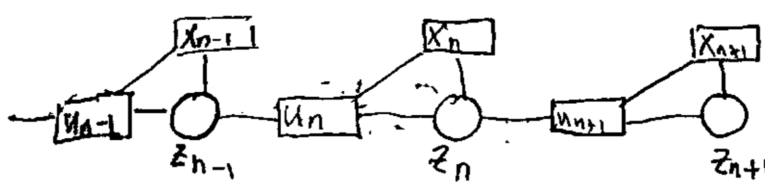
?

$\ln \mu_{f_n\to z_n}(z_n) = \ln p(z_1)$

13.17



$\simeq$



$\boxed{h(z_1) = p(z_{n-1}|x_{n-1})\,p(u_{n-1})}$

13.18 $f_n(z_{n-1}, z_n) = p(z_n|z_{n-1})\cdot p(x_n|z_n)$

$\boxed{\mu_{f_n\to z_n}(z_n) = \sum_{z_{n-1}} p(z_n|z_{n-1})\,p(x_n|z_n)\,\mu_{z_{n-1}\to f_n}(z_{n-1})}$

13.19 $N(x|\mu,\Sigma) = N(x_1,\dots x_n|\mu,\Sigma)$

$p(x_0) = N(x_a|\mu_a,\Sigma_{aa})$

13.20 $p(y) = N(y|A\mu+b, L^{-1}+A\Lambda^{-1}A^T)$; $\int N(z_n|Az_{n-1},T)N(z_{n-1}|\mu_{n-1},V_{n-1})\,dz_{n-1}$

$= N(z_n|A\mu_{n-1}, P_{n-1})$

$$\int \frac{1}{(2\pi T)^{D/2}}\cdot\frac{1}{(2\pi V_{n-1})^{D/2}} e^{-\frac{1}{2T}(z_n-Az_{n-1})^2} \cdot e^{-\frac{1}{2V_{n-1}}(z_{n-1}-\mu_{n-1})^2}\, dz_{n-1}$$

$$= \frac{1}{2\pi(T\cdot V_{n-1})^{D/2}}\int e^{-\frac{1}{2}\left[\frac{1}{T}(z_n-Az_{n-1})^2+\frac{1}{V_{n-1}}(z_{n-1}-\mu_{n-1})^2\right]}\cdot dz_{n-1}$$

$$= \frac{1}{2\pi(TV_{n-1})^{D/2}}\int e^{-\frac{1}{2}\left[\frac{1}{T}(z_n^2-2z_nAz_{n-1}+A^2z_{n-1}^2)+\frac{1}{V_{n-1}}(z_{n-1}^2-2z_{n-1}\mu_{n-1}+\mu_{n-1}^2)\right]} dz_{n-1}$$

$$= \frac{1}{2\pi(TV_{n-1})^{D/2}}\int e^{-\frac{1}{2}\left[\frac{1}{T}(A^2z_{n-1}^2-2z_nAz_{n-1})+\frac{1}{V_{n-1}}(z_{n-1}^2-2z_{n-1}\mu_{n-1})+\frac{1}{T}(z_n^2)+\frac{1}{V_{n-1}}(\mu_{n-1}^2)\right]} dz_{n-1}$$

$$= \frac{1}{2\pi(TV_{n-1})^{D/2}}\cdot\int e^{-\frac{1}{2}\left[(\frac{1}{T}A^2+\frac{1}{V_{n-1}})z_{n-1}^2-2(z_nA-\mu_{n-1})z_{n-1}\right]}\cdot e^{(\frac{1}{T}z_n^2+\frac{1}{V_{n-1}}(\mu_{n-1}^2))\cdot -\frac{1}{2}}\, dz_{n-1}$$

$$= \vdots \int e^{-\frac{1}{2}(\frac{1}{T}A^2+\frac{1}{V_{n-1}})\left[z_{n-1}^2-2\frac{z_nA-\mu_{n-1}}{(\frac{1}{T}A^2+\frac{1}{V_{n-1}})}z_{n-1}\right]}\cdot e^{-\frac{1}{2}(\frac{1}{T}z_n^2+\frac{1}{V_{n-1}}(\mu_{n-1}^2))}\, dz_{n-1}$$

$$= \vdots \int e^{-\frac{1}{2}(\frac{1}{T}A^2+\frac{1}{V_{n-1}})\left[z_{n-1}-\frac{TV_{n-1}(z_nA-\mu_{n-1})}{V_{n-1}\cdot A^2+T}\right]^2-\left[\frac{TV_{n-1}(z_nA-\mu_{n-1})}{V_{n-1}A^2+T}\right]^2}\cdot e^{-\frac{1}{2}(\frac{z_n^2}{T}+\frac{\mu_{n-1}^2}{V_{n-1}})}\, dz_{n-1}$$

$$= \vdots \int e^{-\frac{1}{2}(P_{n-1}/T\cdot V_{n-1})\left[z_{n-1}-\frac{TV_{n-1}(z_nA-\mu_{n-1})}{P_{n-1}}\right]^2-\left[\frac{TV_{n-1}(z_nA-\mu_{n-1})}{P_{n-1}}\right]}\cdot e^{-\frac{1}{2}(\frac{z_n^2}{T}+\frac{\mu_{n-1}^2}{V_{n-1}})}\cdot dz_{n-1}$$

$= \vdots \int$

13.87, 2.113; 2.114, 2.115

$\int N(z_n|Az_{n-1},T)N(z_{n-1}|\mu_{n-1},V_{n-1})dz_{n-1} = N(z_n|A\mu_{n-1},P_{n-1})$

$P(x)=N(x|\mu,\Lambda^{-1})$; $p(y|x)=N(y|Ax+b, L^{-1})$; $p(y)=N(y|A\mu+b, L^{-1}+A\Lambda^{-1}A^T)$

$y=z_n$; $Ax=Az_{n-1}$; $L^{-1}=T$; $x=z_{n-1}$; $\mu=\mu_{n-1}$; $\Lambda^{-1}=V_{n-1}$

$\boxed{P(y)=N(z_n|A\mu_{n-1},T+AV_{n-1}A^T)}$

$z_n|Az_{n-1},T) \Rightarrow N(z_n|Az_{n-1}+a,T)$

$n|Cz_n,\Sigma) \Rightarrow N(x_n|Cz_n+c,\Sigma)$

$=\begin{bmatrix}V_0 & 0\\0 & 0\end{bmatrix}$ $T=\begin{bmatrix}T & 0\\0 & 0\end{bmatrix}$

$=\begin{bmatrix}C & c\end{bmatrix}$ $(P'_{n-1})^{-1}=\begin{bmatrix}P_{n-1}^{-1} & 0\\0 & 0\end{bmatrix}$

$\Sigma)^{-1}$

$|Az_{n-1},T')$

$|Cz_n,\Sigma|$,

$m_0=\mu_0$; $V_0=\sigma_0^2$; $\Sigma=\sigma^2$

$x_n-CA\mu_{n-1})$; $\mu_1=\mu_0+K_1(x_1-C\mu_0)$

; $V_1=(I-K_1C)V_0$

$\boxed{\frac{1}{+\Sigma}=0}$; $K_1=\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}$; $\boxed{IV_0=\frac{1}{\sigma_0^2}}$; $K_1V_0=\frac{N\sigma_0^2\sigma^2}{N\sigma_0^2+\sigma^2}$

; $\Sigma=\sigma^2I$; $(A+BD^{-1}C)^{-1}=A^{-1}-A^{-1}B(D+CA^{-1}B)^{-1}CA^{-1}$: $C=W$

$\mu_{n-1}+K_n(x_n-CA\mu_{n-1})$ & $V_n=(I-K_nC)P_{n-1}$

$p(z|x)=N(z|M^{-1}W^T(x-\mu),\sigma^{-2}M)$

$\mu_{n-1}+K_n(x_n-C\cdot 0)$; $V_n=(I-K_n\cdot W)P_{n-1}$

$=(I-K_n\cdot W)(0\cdot V_{n-1}\cdot 0+I)=I-K_nW$

x_n

$=I-(I\cdot W(WIW^T+\sigma^2I))W$

$=I-(W+\sigma^2I)W=I-(I)+\sigma^2I=\sigma^2$

$=N(z_n|A\mu_{n-1};P_{n-1})$

$=\int N(z_n|Az_{n-1};T)N(z_{n-1}|\mu_{n-1},V_{n-1})dz_{n-1} = \int N(z_n|0\cdot z_{n-1},\sigma^2)N(z_{n-1}|Kx_n, I$

$\boxed{=N(z_n|Kx_n,\sigma^2)}$

$N(z_n|Az_{n-1},T)$

$N(x_n|Cz_n,\Sigma)$; $\Sigma=0$; $N(x_n|Cz_n,0)$

x_n ; $\sigma^2=0$; $\boxed{N(x_n|Cz_n)}$

$N(z_n|Az_{n-1},T)$ $\quad -N(z_n|Az_{n-1},T)N(z_{n-1}|\mu_{n-1},V_{n-1})$

$(x_n|Cz_n,\Sigma)$ $\quad \Rightarrow N(z_n|z_{n-1},0)N(z_{n-1}|\mu_{n-1},\infty)$

; $V_0\to\infty$

$\boxed{=N(z_n|z_{n-1})}$

10.

13.21 $p(y) = N(y|A\mu + b, L^{-1} + A\Lambda^{-1}A^T)$

$p(x|y) = N(x|\Sigma\{A^T L(y-b) + \Lambda\mu\}, \Sigma)$; $\Sigma = (\Lambda + A^T L A)^{-1}$

$(P^{-1} + B^T R^{-1} B)^{-1} B^T R^{-1} = PB^T(BPB^T + R)^{-1}$

$(I + AB)^{-1} A = A(I + BA)^{-1}$

$\boxed{V = P(I - KC)}$ $\boxed{PC(CPC^T + \Sigma)^{-1}}$

13.22. $c_1 \hat{\alpha}(z_1) = p(z_1)p(x_1|z_1) = \boxed{N(x_1|Cz_n, \Sigma + CV_0C^T)}$

$p(x_n|z_n) = N(x_n|Cz_n, \Sigma) = N(y|Ax, L^{-1})$ $\boxed{V_1 = V_0(I - K_1C)}$

$p(z_1) = N(z_1|\mu_0, V_0) = N(x|\mu, \Lambda^{-1})$ $\boxed{\mu_1 = \mu_0 + K_1(x_1 - C\mu_0)}$

13.23 $c_1 \hat{\alpha}(z_1) = p(z_1)p(x_1|z_1)$

$p(x_n|z_n) = N(x_n|Cz_n, \Sigma)$

$p(z_1) = N(z_1|\mu_0, V_0)$

$\boxed{K_1 = V_0C^T(CV_0C^T + \Sigma)}$

13.29 $c_{n+1}\hat{\beta}(z_n) = \int \hat{\beta}(z_{n+1}) p(x_{n+1}|z_{n+1}) p(z_{n+1}|z_n) dz_{n+1}$

$$\hat{\alpha}(z_n) c_{n+1} \hat{\beta}(z_n) = \int \hat{\alpha}(z_n)\hat{\beta}(z_{n+1}) \underbrace{p(x_{n+1}|z_{n+1})} p(z_{n+1}|z_n) dz_{n+1}$$

$$\hat{\alpha}(z_n) c_{n+1} \hat{\beta}(z_n) = \int \hat{\alpha}(z_n)\hat{\beta}(z_{n+1})\, p(z_{n+1}|z_n)\, p(z_{n+1}|z_n) dz_{n+1}$$

$p(z_n|z_{n-1}) = N(z_n|Az_{n-1}, T)$

$p(x_n|z_n) = N(x_n|Cz_n, \Sigma)$

$$= \int \hat{\alpha}(z_n)\hat{\beta}(z_{n+1}) N(z_{n+1}|Az_n, T) N(z_{n+1}|Az_n, T) dz_{n+1}$$

$\mu_n = A\mu_{n-1} + K_n(x_n - CA\mu_{n-1})$

$V_n = (I - K_n C) P_{n-1}$

$c_n = N(x_n | CA\mu_{n-1}, (P_{n-1}C^T + \Sigma))$

$\gamma(z_n) = \hat{\alpha}(z_n)\hat{\beta}(z_n) = N(z_n|\hat{\mu}_n, \hat{V}_n)$ ✓

$\frac{1}{2} = e^{-kt}$

$\frac{A+Bt+Ct^2 \cdots}{A+Bt+C \cdots} = e^{kt}$

$(A+Bt+Ct^2+) = 2(\cdots)$

$(A+Bt+C\cdots) - 4C$

$$= \int N(z_n|\hat{\mu}_n, \hat{V}_n) N(z_{n+1}|Az_n, T) N(z_{n+1}|Az_n, T)$$

$$= \int N(z_n|A\mu_{n-1} + K_n(x_n - CA\mu_{n-1}), (I-K_nC)P_{n-1}) \cdot N(z_{n+1}|Az_n, T) \cdot N(z_{n+1}|Az_n, T) dz_{n+1}$$

$$= \int$$

13.30 $\xi(z_{n-1}, z_n) = c_n \hat{\alpha}(z_{n-1}) p(x_n|z_n) p(z_n|z_{n-1}) \hat{\beta}(z_n)$

$\xi(z_{n-1}, z_n) = (c_n)^{-1} \hat{\alpha}(z_{n-1}) p(x_n|z_n) p(z_n|z_1) \hat{\beta}(z_n)$

$= \underline{N(z_n|\mu_n, V_n) N}$

13.31 ?

13.32 $\mu_0^{new} = \mathbb{E}[z_1]$; $V_0^{new} = \mathbb{E}[z_1 z^T] - \mathbb{E}[z_1][z_1^T]$

$$Q(\theta, \theta^{old}) = -\frac{1}{2}\ln|V_0| - \mathbb{E}_{z|\theta^{old}}\left[\frac{1}{2}(z_1-\mu_0)V_0^{-1}(z_1-\mu_0)\right] + const$$

$$\frac{dQ(\theta,\theta^{old})}{d\mu_0} = \mathbb{E}\left[(z_1-\mu_0)V_0^{-1}\right] = 0 \quad ; \boxed{\mu_0^{new} = \mathbb{E}[z_1]}$$

$$\frac{dQ(\theta,\theta^{old})}{dV_0} = \frac{-1}{2V_0} + \mathbb{E}\left[\frac{1}{2}(z_1-\mu_0)V_0^{-2}(z_1-\mu_0)\right] = 0$$

$$\boxed{V_0^{new} = \mathbb{E}[(z_1-\mu_0)(z_1-\mu_0)]}$$

~~13.33~~ Verify $A^{new} = \left(\sum_{n=2}^{N} \mathbb{E}[z_n z_{n-1}^T]\right)\left(\sum_{n-2}^{N} \mathbb{E}[z_{n-1} z_{n-1}^T]\right)^{-1}$

$$\Gamma^{new} = \frac{1}{N-1} \sum_{n=2}^{N} \left\{ \mathbb{E}[z_n z_n^T] - A^{new}\mathbb{E}[z_{n-1} z_n^T] - \mathbb{E}[z_n z_{n-1}^T]A^{new} + A^{new}\mathbb{E}[z_{n-1} z_{n-1}^T](A^{new})^T \right\}$$

$$Q(\theta,\theta^{old}) = -\frac{N-1}{2}\ln|\Gamma| - \mathbb{E}_{z|\theta^{old}}\left[\frac{1}{2}\sum_{n=2}^{N}(z_n - Az_{n-1})^T\Gamma^{-1}(z_n - Az_{n-1})\right] + const$$

$$\frac{dQ(\theta,\theta^{old})}{d\Gamma} = -\frac{N-1}{2\Gamma} + \mathbb{E}_{z|\theta^{old}}\left[\frac{1}{2}\sum_{n=2}^{N}(z_n - Az_{n-1})^T\Gamma^{-2}(z_n - Az_{n-1})\right] = 0$$

$$\boxed{\Gamma = \frac{1}{(N-1)} \underset{\mathbb{E}_{z|\theta^{old}}}{\,} \sum_{n=2}^{N}(z_n - Az_{n-1})^T(z_n - Az_{n-1})}$$

$$\frac{dQ(\theta,\theta^{old})}{dA} = \mathbb{E}_{z|\theta^{old}}\left[\sum_{n=2}^{N}(z_n - Az_{n-1})\right]z_{n-1} = 0$$

$$\boxed{A = \mathbb{E}_{z|\theta^{old}}[z_n z_{n-1}^T]\,\mathbb{E}_{z|\theta^{old}}[z_{n-1} z_{n-1}]^{-1}}$$

13.34 $Q(\theta,\theta^{old}) = -\frac{N}{2}\ln|\Sigma| - \mathbb{E}_{z|\theta^{old}}\left[\frac{1}{2}\sum_{n=1}^{N}(x - Cz_n)^T\Sigma^{-1}(x - Cz_n)\right] + const$

$$\frac{dQ(\theta,\theta^{old})}{dC} = \mathbb{E}_{z|\theta^{old}}\left[\sum(x_n - Cz_n)\right]z_{n-1} = 0$$

$$\boxed{C^{new} = \left(\sum_{n=1}^{N} x_n \mathbb{E}[z_n^T]\right)\left(\sum \mathbb{E}[z_n z_n^T]\right)^{-1}}$$

$$\frac{dQ(\theta,\theta^{old})}{d\Sigma} = -\frac{N}{\Sigma} + \mathbb{E}_{z|\theta^{old}}\left[\sum_{n=1}^{N}(x - Cz_n)^T\Sigma^{-2}(x - Cz_n)\right] = 0$$

$$\boxed{\Sigma^{new} = \frac{1}{N}\sum_{n=1}^{N}\left[(x - Cz_n)^T(x - Cz_n)\right]}$$

Chapter 14:

1. $p(t|x, z_n, \theta_h, h)$; x = input vector
   t = target vector
   h = indexes of different models
   z_n = latent variable for model h
   θ_h = set of parameters for model h

Write down the formulae needed to evaluate $p(t|x,X,T)p(h)p(z_n)$

$$p(t|x,X,T) = \sum p(h) \sum p(z_n) \int p(t|x,\theta_h,z_n,h)\,p(\theta_h|X,T,h)\,d\theta$$

where $p(\theta_h|X,T,h) = \dfrac{p(T|X,\theta,h)\,p(\theta|h)}{p(T|X,h)}$

$$= p(\theta|h)\prod_{n=1}^{N} p(t_n|x_n,\theta,h)$$

$$= p(\theta|h)\prod_{n=1}^{N}\left(\sum p(t_n, z_{nh}|x_n,\theta,h)\right)$$

14.2 $E_{AV} = \frac{1}{M}\sum_{m=1}^{M} \mathbb{E}_x[\epsilon_m(x)^2]$; $E_{com} = \mathbb{E}_x\left[\left\{\frac{1}{M}\sum_{m=1}^{M} y_m(x) - h(x)\right\}^2\right] = \mathbb{E}_x\left[\left\{\frac{1}{M}\sum_{m=1}^{M}\epsilon_m(x)\right\}^2\right]$

Assuming $\mathbb{E}_x[\epsilon_m(x)] = 0$; $\mathbb{E}_x[\epsilon_m(x)\epsilon_l(x)] = 0$

$$E_{com} = \mathbb{E}_x\left[\left\{\frac{1}{M}\sum_{m=1}^{M}\epsilon_m(x)\right\}^2\right] = \frac{1}{M}\mathbb{E}_x\left[\left\{\sum_{m=1}^{M}\epsilon_m(x)\right\}^2\right] = \boxed{\frac{1}{M}E_{AV}}$$

14.3 Jensen's Inequality: $f\left(\sum_{i=1}^{M}\lambda_i x_i\right) \leq \sum_{i=1}^{M}\lambda_i f(x)$; $\lambda_i \geq 0$; $\sum_i \lambda_i = 1$

if $f(x) = x^2$; Prove $\frac{1}{M}\sum_{m=1}^{M}\mathbb{E}_x[\epsilon_m(x)^2]$

$$E_{com} = \left[\left\{\frac{1}{M}\sum_{m=1}^{M}\epsilon_m(\lambda x)\right\}^2\right] \leq \frac{\lambda}{M}\mathbb{E}_x\left[\left\{\sum_{m=1}^{M}\epsilon_m(x)\right\}^2\right] \quad ; \quad \boxed{\lambda^2 < \lambda}$$

$$\lambda^2\left\{\frac{1}{M}\sum \epsilon_m(x)\right\}^2 \leq \frac{\lambda}{M}\mathbb{E}_x\left[\left\{\sum_{m=1}^{M}\epsilon_m(x)\right\}^2\right]$$

$$\boxed{E_{com} \leq E_{AV}}$$

14.4 Jensen's Inequality $f\left(\sum_{i=1}^{M}\lambda_i x_i\right) \leq \sum_{i=1}^{M}\lambda_i f(x)$; $\lambda_i \geq 0$; $\sum \lambda_i = 1$

Prove $E_{com} < E_{AV}$ for convex function.

Definition convex $F(x)'' > 0$; $\boxed{\lambda^2\frac{1}{M} \leq \frac{\lambda}{M}}$

14.5 $y_{com}(x) = \sum_{m=1}^{M}\alpha_m y_m(x)$; $y_{min}(x) \leq y_{com}(x) \leq y_{max}(x)$

Show $\alpha_m \geq 0$; $\sum_{m=1}^{M}\alpha_m = 1$; $y_{min}(x) = \min\left(\sum_{m=1}^{M}\alpha_m y_m(x)\right)$; $y'_{min}(x) = \min(\alpha_m) = 0$

$y_{com}(x) = \sum_{m=1}^{M}\alpha_m y_m(x)$; $y'_{com}(x) = \sum \alpha_m = 0$

$y_{max}(x) = \max\left(\sum_{m=1}^{M}\alpha_m y_m(x)\right)$; $y'_{max}(x) = \max(\alpha_m) = 0$

14.6 $E = e^{-\alpha_m/2}\sum_{n\in T_m} w_n^{(m)} + e^{\alpha_m/2}\sum_{n\in M} w_n^{(m)} = \left(e^{\alpha_m/2} - e^{-\alpha_m/2}\right)\sum_{n=1}^{N} w_n^{(m)} I(y_m(x_n) \neq t_n) + e^{-\alpha_m/2}\sum_{n=1}^{N} w_n^{(m)}$

$$\frac{dE}{d\alpha_m} = \frac{e^{\alpha_m/2} + e^{-\alpha_m/2}}{2}\sum_{n=1}^{N} w_n^{(m)} I(y_m(x_n) \neq t_n) - \frac{e^{-\alpha_m/2}}{2}\sum_{n=1}^{N} w_n^{(m)} = 0$$

$$\alpha = \frac{\sum_{n=1}^{N} w_n^{(m)} I(y_m(x_n) \neq t_n)}{\sum_{n=1}^{N} w_n^{(m)}}$$

14.7 $\mathbb{E}_{x,t}[\exp\{-ty(x)\}] = \sum\int \exp\{-ty(x)\}p(t|x)p(x)dx$; $y(x) = \frac{1}{2}\ln\left\{\frac{p(t=1|x)}{p(t=-1|x)}\right\}$

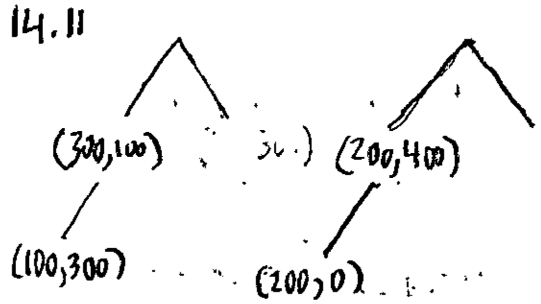
$\frac{d\mathbb{E}_{x,t}}{dy(x)} = -te^{-ty(x)}p(t|x)p(x) = 0$; $y(x) = \frac{1}{2}\ln\left\{\frac{p(t=1|x)}{p(t=-1|x)}\right\}$

14.8 $E=\sum_{n=1}^{N}\exp\{-t_n f_m(x_n)\}$; $\frac{dE}{df(x)}=-t_n\sum_{n=1}^{N}\exp\{-t_n f(x)\}=0$; $-t_n f(x)\neq 1$; $\boxed{f(x)=\frac{-1}{t_n}\neq 1}$

14.9 $f_m(x)=\frac{1}{2}\sum_{l=1}^{m}\alpha_l y_l(x)$ $\frac{df_m(x)}{d\alpha_L}=\frac{1}{2}\sum_{i=1}^{m}y_L(x)=f'(x)$ $\boxed{E=\sum_{n=1}^{N}\exp\{-t_n\frac{1}{2}\sum_{i=1}^{m}y(x)\}}$

14.10 $f(x)=\sum_{i=1}^{n}(y(x)-t_n)^2$; $\frac{dF(x)}{dt}=-\sum_{i=1}^{n}(y(x)-t_n)=0$; $\boxed{y(x)=\frac{t_n}{N}}$

14.11



$Q_T(T)=\sum_{k=1}^{K}p_{Tk}\ln p_{Tk}$

$Q_T(T)=\sum_{k=1}^{K}p_{Tk}(1-p_{Tk})$

14.12 $p(t|\theta)=\sum\pi_k N(t|w_k^T\phi,\beta^{-1})$; $y(x,w)=w^T\phi(x)$; $p(t|x,W,\beta)=N(t|W^T\phi(x),\beta^{-1}I)$

$\cancel{\ln p(T|X,W,\beta)=\sum\ln N(t_n}|W^T\phi(x),\beta^{-1}I)=\frac{NK}{2}\ln\left(\frac{\beta}{2\pi}\right)-\frac{\beta}{2}\sum\|t_n-W^T\phi(x)\|^2$

$W_{ML}=(\phi^T\phi)^{-1}\phi^T T$; $w_k=(\phi^T\phi)^{-1}\phi^T t_k=\phi^\dagger t$

$p(t|\theta)=\sum\pi_k N(t|W_k^T\phi(x),\beta^{-1})=\sum\pi_k N(t|y,\beta^{-1})$

$\boxed{\ln(p(t|\theta))=\sum\pi_k\ln N(t|W^T\phi(x),\beta^{-1})=\frac{NK}{2}\ln\frac{\beta}{2\pi}-\frac{\beta}{2}\sum\|t-W^T\phi(x)\|^2}$

14.13

14.14 $\frac{dQ(\theta,\theta^{old})}{d\pi_k}=\frac{d}{d\pi_k}\mathbb{E}[\ln p(t,z|\theta)]=\frac{d}{d\pi_k}\sum_{n=1}^{N}\sum_{k=1}^{K}\gamma_{nk}\{\ln\pi_k+\ln N(t_n|w_k^T\phi_n,\beta^{-1})\}+\lambda\left(\sum_{n=1}^{K}\pi_k-1\right)$

$=\frac{\gamma_{nk}}{\pi_k}-\lambda=0$; $\lambda=\frac{\gamma_{nk}}{\pi_k}$; $\sum\frac{\gamma_{nk}}{\pi_k}-\gamma_{nk}=0$; $\boxed{\pi_k=\sum\frac{\gamma_k}{N}}$

14.15 $\boxed{p(t|\theta)=\sum\pi_k N(t|w_k^T\phi,\beta^{-1})=\pi_1 N(t|w_k^T\phi,\beta^{-1})+\pi_2 N(t|w_k^T\phi,\beta^{-1})+\cdots}$

14.16 $p(t|\phi,\theta)=\sum_{k=1}\pi_k y_k^t[1-y_k]^{1-t}$ Soft Max Classifiers : $K>2$

$\ln p(t|\phi,\theta)=\ln\sum\pi_k y_n^t[1-y_k]^{1-t}$

$p(c|\lambda)=\frac{p(t|\phi)p(\phi|\theta)}{\sum p(t|\theta)p(\phi|\theta)}$

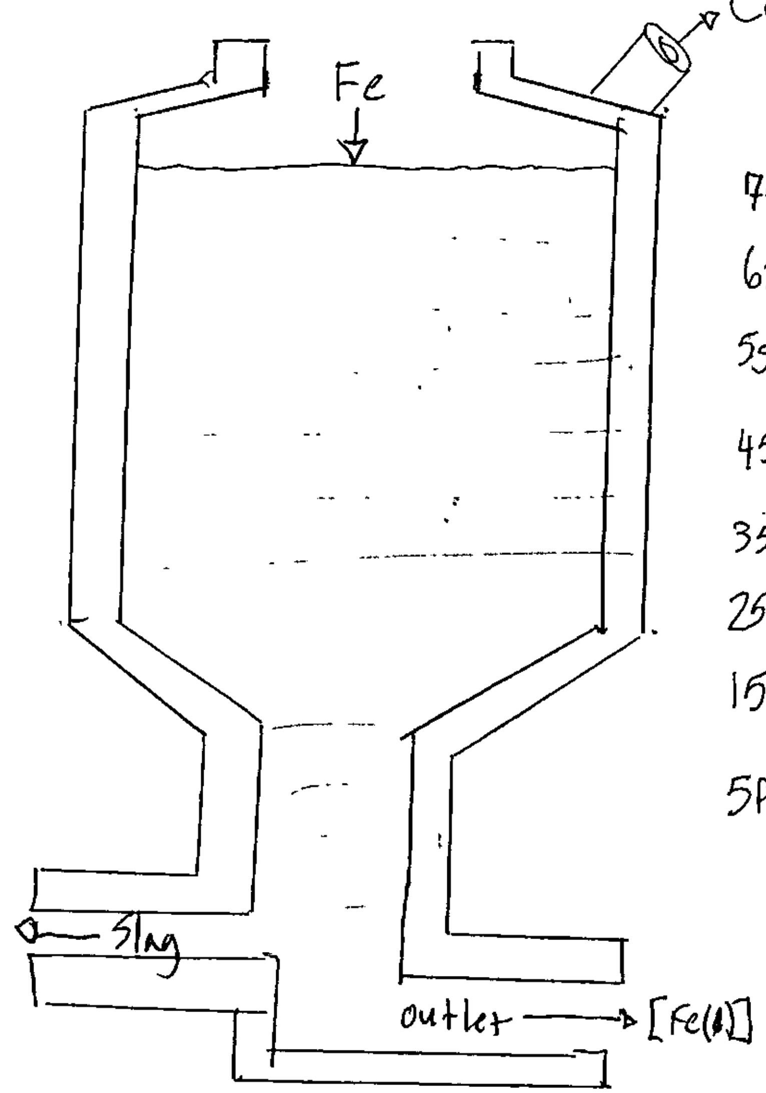
$p(C_k|x)=\frac{p(x|C_k)p(C_k)}{\sum p(x|C_j)p(C_j)}$

$=\frac{\exp(a_k)}{\sum_j\exp(a_j)}$

14.17 $p(t|x)=\sum_{n=1}^{K}\pi_k \psi_k(t|x)$

$a_k=\ln p(x|C_k)p(C_k)$

Isolation of Iron:



75ft 230°C $Fe_2O_3 + CO \rightarrow 2Fe_3O_4 + CO_2$

65ft 410°C $Fe_3O_4 + CO \rightarrow 3FeO + CO_2$

55ft 525°C $FeO + CO \rightarrow Fe + CO_2$

45ft 865°C $C + CO_2 \rightarrow 2CO$

35ft 945°C $CaCO_3 \rightarrow CaO + CO_2$; $C + CO_2 \rightarrow 2CO$

25ft 1125°C $CaO + SiO_2 \rightarrow CaSiO_3$; $C + CO_2 \rightarrow 2CO$

15ft 1300°C $C + CO_2 \rightarrow 2CO$

5ft 1500°C $Fe(l)$

Isolation of Copper:

Reduction:

$$2Cu_2S(l) + 3O_2(g) \rightarrow 2Cu_2O(l) + 2SO_2(g)$$

$$2Cu_2O(l) + Cu_2S(l) \rightarrow 6Cu(l) + SO_2(l)$$