1

PRML

1

$$Loss = \sum_{n=1}^{N} \left(\sum_{j=0}^{M} x_n^j \cdot w_j - t_n \right)^2$$

$$\frac{dL}{dw_k} = \sum_{n=1}^{N} 2 \left(x^k \cdot \sum_{j=0}^{M} x_n^j \cdot w_j - x_n^k t_n \right)$$

$$\sum_{n=1}^{N} \sum_{j=0}^{M} x_n^{j+k} \cdot w_j = \sum_{n=1}^{N} x_n^k t_n$$

$$\sum_{j=0}^{M} w_j \cdot \sum_{n=1}^{N} x_n^{j+k} = T_i$$

$$\sum_{j=0}^{M} w_j \cdot A_{ij} = T_i$$

2

$$\frac{dLoss}{dw_k} = \sum_{n=1}^{N} \left(x^k \cdot \sum_{j=0}^{N} x_n^j \cdot w_j - x_n^k T_n \right) + \lambda w_k \cdot N$$

$$\sum_{j=0}^{M} \frac{N}{n} \lambda w_k + w_j \sum_{n=0}^{N} x_n^{j+k} = \sum_{n=1}^{N} x_n^k t_n = T_i$$

or $A_{ij} = \sum_{n=1}^{N} x_n^{j+k} - \lambda \cdot (i == j \,!= 0)$ or $[A] - I \cdot \vec{\lambda}$ $i = 0$ if 1st, λ else

$$\sum_{j=0}^{M} A_{ij} \cdot w_j = T_i$$

3

$P(a) = P(a|r) \cdot P(r) + P(a|b) \cdot P(b) + P(a|g) \cdot P(g) = 0.34$

$P(g|o) = P(o|g) \cdot P(g) / P(o)$ $P(o|g) = 3/10$ $P(g) = 0.6$

$P(o) = \sum_{k \in \{r,b,g\}} P(o|k) \cdot P(k) = 0.38$

$P(g|o) = 102/190$

$\boxed{4}$ $\arg\max_x \; P_x(x) \quad \hat{x} = g(\bar{y})$
$\arg\max_y \; P_x(g(y))|g'| = \hat{y}$

apply; $\frac{d}{dy} = 0$ if arg max to $P_x(g(\bar{y}))$ @ $\bar{y} = \hat{y}$

$$0 = \underbrace{P_x'(g(\bar{y})) \cdot g'(\bar{y}) \, |g'(\bar{y})|}_{=0} + P_x(g(\bar{y})) \cdot \frac{g'(\bar{y})}{|g'(\bar{y})|} \cdot g''(\bar{y})$$

since $\hat{x} = g(\bar{y})$

$P_x(g(\bar{y})) > 0$ if $\bar{y} = \hat{y}$, $g'(\bar{y}) \neq 0$ for $P_y(y)$ to be defined
so $\bar{y} = \hat{y}$ i.f.f. $g''(\bar{y}) = 0$

$\boxed{5}$ $E[f - E(f)] = E[f^2 - 2f E(f)) + E^2(f)] = E(f^2) - 2E(f) \cdot E(f) + E^2(f)$
$= E(f^2) - E^2(f)$

$\boxed{6}$ $\int_x \int_y yx \, p(x,y) \, dy \, dx = \int_x x \, p(x) dx \int_y p(y) \cdot y \, dy = E[X Y] = E[X] E[Y]$

$\boxed{7}$ $I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-1/2\sigma^2 (x^2+y^2)} \quad x = r\cos\theta \quad y = r\sin\theta$

$$\begin{vmatrix} dx/dr & dx/d\theta \\ dy/dr & dy/d\theta \end{vmatrix} = \begin{vmatrix} \cos dr & -r\sin\theta \, d\theta \\ \sin\theta \, dr & r\cos\theta \, d\theta \end{vmatrix} = r(\cos^2 + \sin^2) dr d\theta = r \, dr \, d\theta$$

$$= \int_0^{2\pi} d\theta \int_0^{\infty} e^{-r^2/2\sigma^2} r \, dr = 2\pi \cdot \Big|_0^{\infty} \sigma^2 e^{-r^2/2\sigma^2}$$

$$= 2\pi\sigma^2 (0 - 1) = 2\pi\sigma^2 \rightarrow I = \sqrt{2\pi\sigma^2} \rightarrow I/\sqrt{2\pi\sigma^2} = 1$$

$\boxed{8}$ $\int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi\sigma^2}} e^{-(x-u)^2/2\sigma^2} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (v+u) e^{-v^2/2\sigma^2} dv$
where $v = x - u$

$\int u e^{-v^2/2\sigma^2} dv = u \cdot \frac{\sqrt{2\pi\sigma^2}}{\sqrt{2\pi\sigma^2}} = u$ by $\boxed{7}$

$+ \int v e^{-v^2/2\sigma^2} dv = \Big|_{-\infty}^{\infty} -\sigma^2 e^{-v^2/2\sigma^2} = 0$ for $v = \pm\infty$ $\quad u + 0 = \boxed{u}$

3

6 cont

$\cancel{\frac{d}{d\sigma^2} \int \frac{-v^2/\sigma^4}{\sqrt{2\pi}} \cdot e^{-v^2/2\sigma^2} \, dv}$ $\cancel{v = x - v}$

$\frac{d}{d\sigma^2} \int \frac{1}{\sqrt{2\sigma^2\pi}} e^{-(x-v)^2/2\sigma^2} \, dx = 1$

$\frac{d}{d\sigma^2} \int e^{-(x-v)^2/2\sigma^2} \, dx = \sqrt{2\sigma^2\pi}$ let $V = \sigma^2$

$vp = -(x-v)^2/2\sigma^2$

$dn = \sqrt{2\sigma^2\pi}$

$\frac{d}{d\sigma^2} = \frac{d}{dV}\frac{dV}{d\sigma}$ where $\frac{dV}{d\sigma} = 2\sqrt{V}$

$\int_{-\infty}^{\infty} \frac{d}{dV} 2\sqrt{V} e^{vp} = \frac{d}{dV} 2\sqrt{V} \cdot dn = \frac{d}{dV} 2V\sqrt{2\pi} = 2\sqrt{2\pi}$

$= \int \frac{1}{\sqrt{V}} e^{vp} + \frac{(x-v)^2}{V^{3/2}} e^{vp} = 2\sqrt{2\pi}$

$\sqrt{2\pi} + \frac{1}{V}\int \frac{(x-v)^2}{\sqrt{V}} e^{vp}$

$\frac{V}{\sqrt{2\pi}} \times \to \frac{1}{\sqrt{2\pi V}} \int e^{vp} (x-v)^2 \, dx = V(2-1) = V = \sigma^2$

$\int \frac{x^2}{dn} e^{vp} + \int \frac{-2xv}{dn} e^{vp} + \int \frac{v^2}{dn} e^{vp} = \sigma^2$

$E[x^2] + -2v^2 + v^2 = \sigma^2$

$E[x^2] = \sigma^2 + v^2$

$\boxed{9}$ since log monotonic $\arg\max_x \log\left(\frac{1}{dn} e^{vp}\right)$

$= -\lg(dn) + vp \quad \propto m \; -(x-\vec{v})^2/2\sigma^2$ // $-\lg(dn)$ const

$\to$ minimize $(x-\vec{v})^2 \to x = \vec{v}$

$\boxed{10}$ $E[X+Z] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+z)\cdot p(x,z)\,dx\,dz$ $\quad p(x,z) = p(x)\,p(z)$

$= \int p(z)\left(\int x\,p(x)\,dx + \int z\,p(x)\,dx\right)dz$

$= \int p(z)\,(E[X] + z)\,dz$

$= \int p(z)\,E[X]\,dz + \int p(z)\,z\,dz$

$= E[X] + E[Z]$

$Var(X+Z) = E[(X+Z)^2] - E[X+Z]^2$

$= E[X^2] + E[2XZ] + E[Z^2] - (E[X] + E[Z])^2$

$= (E[X^2] - E[X]^2) + (E[Z^2] - E[Z]^2) + 2(E[XZ] - E[X]\cdot E[Z])$

$Var(X) + Var(Z) + 0$ as

$E[XZ] \neq E[X]E[Z] \to p(x,z) \neq p(x)\cdot p(z)$

$\boxed{11}$ $\frac{d}{dU}\ln(p(x|U,\sigma^2)) = \frac{1}{\sigma^2}\sum_{n=1}^{n}(x_n - U) = 0 \to \sum_{n=1}^{n} x_n = \sum^{N} U = NU$

$\frac{d}{d\sigma^2}\ln(p(x|U,\sigma^2)) =$ $\quad V = \sigma^2$

$\frac{d}{dV}\; -\frac{1}{2}V^{-1}\sum(x_n - U)^2 - \frac{N}{2}\ln(V) - \frac{N}{2}\ln(2\pi)$

$\to 0 = V^{-2}\,2\sigma\sum(x_n - U)^2 - \frac{N}{2V}\cdot 2\sigma$

$\frac{1}{\sigma}\cdot\sigma^3 = \frac{1}{N}\cdot\sum_{N}^{n=1}(x_n - U)^2 = \sigma^2$

$\boxed{12}$ $E[x^2] = U^2 + \sigma^2$ & $E[x_m \cdot x_n] = \int\int x_m \frac{1}{\sigma n} e^{(x_m - U)/2\sigma^2}\cdot x_n \cdot \frac{1}{\sigma n}\cdot e^{(x_n - U)/2\sigma^2}\, dx_m\, dx_n$

$= E[X_m]\cdot E[X_n] = U^2$

$E[U_{ML}] = \frac{1}{N}\sum^{n} E[X_n] = \frac{1}{N}\sum^{n} U = U$

5

12 cnt $E[\sigma_{ML}^2] = \frac{1}{N} \Sigma^N E[(x_n - U_{ML})^2] = \frac{1}{N} \Sigma E[X_n^2] - 2E[X_n \cdot U_{ML}] + E[U_{ML}^2]$

$= \frac{1}{N} \Sigma\, U^2 + \sigma^2 - 2 \cdot (U \cdot U_{ML} + \frac{\sigma^2}{N}) + E[\frac{1}{N} \cdot \frac{1}{N} (\Sigma_{i \neq j}^{N,N} X_i \cdot X_j + \Sigma_{n=1}^{N} X_n^2)]$

$\sigma^2 - U^2 - \frac{2\sigma^2}{N} + U^2 \cdot \frac{n^2 - n}{n^2} + (U^2 + \sigma^2) \cdot N/N^2$

$< \sigma^2 - \frac{\sigma^2}{N} = \sigma^2 \cdot (1 - 1/N)$

13 as $E[X_n \cdot U_{ML}] = E[X_n \cdot \frac{1}{N} \Sigma_{n=1}^{N} X_m] = E[X_n^2 + \frac{1}{N} \Sigma_{n \neq m} X_m]$

$= U^2 + \sigma^2 + \frac{n-1}{N} \cdot U = \sigma^2 \cdot 1/N + U^2$ if U_{ML} is replaced

by a const. $E[X_n \cdot k] = k \cdot U$ w/o a σ^2 term

14 $W_{ij} = W_{ij}^A + W_{ij}^S$ & $W_{ji} = W_{ji}^A + W_{ji}^S = -W_{ij}^A + W_{ij}^S$

$W_{ij} + W_{ji} = 2 W_{ij}^S \to W_{ij}^S = \frac{W_{ij} + W_{ji}}{2}$ $W_{ij}^A = W_{ij}^S - W_{ji}$

Lower triangular

b) $\sum^D \sum^D W_{ij} x_i x_j = \Sigma_{i=1}^{D/2} \Sigma_{j=1}^{D/2} W_{ij} x_i x_j + W_{ji} x_j x_i$ if odd $W_{D/2\, D/2} = W^S_{D/2\, D/2}$

$= \Sigma_{i=1}^{D/2\ LT} \Sigma_{j=1}^{D/2} x_i x_j (W_{ji} + W_{ij})$

$= \Sigma_{i=1}^{D/2\ LT} \Sigma_{i=j}^{D/2} x_i x_j \cdot W_{ij}^S \cdot 2 = \Sigma^{\text{Full matrix}} \Sigma\, x_i x_j \cdot W_{ij}^S$

as $(x_i x_j) \cdot W_{ij}^S = (x_j \cdot x_i) \cdot W_{ji}^S$

15

1.15 For the base case of an induction proof $m=1$ is trivial.

assuming $\sum_{i_1}^{D} \sum_{i_2} \cdots \sum^{i_{m-1}} W_{i \ldots i_{m-1}} x_{i_1} \cdots x_{i_{m-1}} = \sum_{i_1}^{D} \cdots \sum_{i_{m-1}}^{D}$

$\Rightarrow \sum_{i_m}^{D} \sum_{i_1=1}^{i_m} \ldots x_{i_m} (W \ldots x_{i_{m-1}}) = \sum_{i_m}^{D} \sum_{i_{m-1}}^{D} \ldots x_{i_m} (W \ldots$

let $z_{i_{m-1}} = \sum_{i_{m-2}}^{i_{m-1}} \cdots \sum_{i_1=1}^{i_2} W_{i_{m-1} \ldots} \cdots x_{i_{m-2}}$

$\sum_{i_m=1}^{D} \sum_{i_{m-1}=1}^{D} x_{i_m} \cdot x_{i_{m-1}} \cdot z_{i_{m-1}} \cdot w_{i_m z} =$

$= \sum_{i_m=1}^{D} x_{i_m} \left(\sum_{i_{m-1}}^{i_m-1} x_{i_{m-1}} \cdot z_{i_{m-1}} + x_{i_m} z_{i_m} + \sum_{i_{m-1}=i_m+1}^{D} x_{i_{m-1}} z \right)$

but $\sum_{i_m}^{D} \sum^{i_m-1} x_{i_{m-1}} \cdot z_{i_{m-1}} = \sum_{i_m=1}^{D} \sum_{i_{m-1}=i_m+1}^{D} x_{i_m} x_{i_{m-1}} \cdot z_{i_{m-1}}$

as $\sum_{i_m=1}^{D} x_{i_m} \sum_{i=i_{m-1}}^{i_m-1} x_{i_{m-1}} \cdot z_{i_{m-1}} \Rightarrow \sum_{i_{m-1}=1}^{D} \sum_{i_{m-1}=i_{m-1}}^{D} x_{i_m}$

as $x_a x_b = x_b x_a$

gives $(a,b)=(b,a)$ symmetry along cartesian grid $(1, i_{m-1}) \times (1, i_m - 1)$

Thus $\sum_{i_m}^{D} \sum_{i_{m-1}}^{D} x_{i_m} x_{i_{m-1}} \cdot z_{i_{m-1}} \cdot w_{i_m z} = 2 \sum_{i_m}^{D} \sum_{i_{m-1}}^{D} x_{i_m} x_{i_{m-1}} \cdot z_{i_{m-1}} w_{i z} + x_{i_m}^2 z_{i_m} w_{i_m}$

as long as $W_{\ldots i_{m-1} i_m} + W_{\ldots i_m i_{m-1}} = \tilde{W}_{\ldots i_{m-1} i_m}$

b) ~~isomorphic to allocating M indentical balls amongst D~~ bins.

as order irrelevant so $x_{i_1} \ldots x_{i_m} = \prod_{j=D}^{p} x_j^{p_j}$

where $\sum P_j = D$ [so all values in descending order so no possibility of dublicates, like in 1.135]

$N(1,1) = 1$ & $N(1,0) = 0$, since

$n(D,M) = \sum_{i=1}^{D} f(i_1) \cdot x_{i_1}$ where $f(x) = \sum_{i_2=1}^{x} \cdots \sum_{i_{m-1}}^{i_{m-2}} \tilde{W} x_{i_1} \cdots$

a reduction of 1 attribute summation $= N(x, m-1)$

c) $D=1 \rightarrow \sum \frac{(1+M-2)!}{0! \cdot (m-1)!} = 1 = \frac{m!}{0! \cdot m!}$

$M=1 \rightarrow \sum_{i=1}^{D} \frac{(i-1)!}{(i-1)! \, (0!)} = D = \frac{D!}{(D-1)! \cdot 1!}$

7

$\boxed{1.15\ C}$ cont

for D+1

$$\sum^{D} \frac{(i+m-2)!}{(i-1)!\,(M-1)!} = \frac{(D+M-1)!}{(D-1)!\,M!} \rightarrow \frac{((D+1)+m-2)!}{(D+1-1)!\,(m-1)!} + \sum^{D}$$

$$= \frac{(D+M-1)!}{D!\,(M-1)!} + \sum^{D} = \frac{(D+m-D)(D+m-1)!}{D!\,m!} + \sum^{D}$$

$$= \frac{(D+m)! - D(D+M-1)!}{D!\,m!} + \frac{(D+M-1)!}{(D-1)!\,M!}$$

$$= \frac{(D+m)!}{D!\,m!} - \cancel{\frac{(D+M-1)!}{(D-1)!\,M!}} + \cancel{\frac{(D+M-1)!}{(D-1)!\,M!}}$$

d) $N(D,2) = (D+2-1)!/(D-1)!\,2 = (D+1)\cdot D/2$

number of independents from square matrix

$$n(D,M) = \sum n(i,m-1) = \sum_{i=1}^{D} (i+m-2)!/(i-1)!\,(m-1)!$$

$$= (D+m-1)!/(D-1)!\,M!$$

$\boxed{1.16}$ $N(D,M) \geq \sum_{m=0}^{M} N_n(D,m)$ as orders independent but $\leq$ as couldn't have total index greater than constituents. can always add higher order terms to increase that # parameters.

$$N(D,M) = \sum_{m=0}^{M} \frac{(D+m-1)!}{(D-1)!\,m!} \xrightarrow{m+1} \frac{(D+m)!}{(D-1)!\,(m+1)!} + \sum^{m}$$

$$= \frac{(D+m)!}{(D-1)!\,(m+1)!} + \frac{(D+m)!}{D!\,M!} = \frac{(D+m+1)(D+m)!}{D!\,(m+1)!} = \frac{(D+m+1)!}{D!\,(m+1)!}$$

$= N(D,m+1)$

$$D \gg m + \frac{(D+m)^{D+m} e^{-D} e^{-m}}{D^{D} e^{-D}\, m^{m} e^{-m}} = \frac{(D+m)^{D+m}}{D^{D}\, m^{m}} = \frac{D^{D+m}(1+m/D)^{D+m}}{D^{D}\, m^{m}}$$

$$= \frac{D^{D+m}}{D^{D}\cdot m^{m}} \cdot \left(1+\frac{m}{D}(D+m)\right) \approx \frac{D^{m}}{m^{m}}(1+m+m^{2}/D) \approx D^{m}$$

$\boxed{1.17}$ $\Gamma(x+1) = \int_0^\infty u \cdot u^{x-1} e^{-u}\, du$

$a = u \cdot u^{x-1}$ $db = e^{-u}$

$da = x \cdot u^{x-1}$ $b = -e^{-u}$

$= -u^x e^{-u} - \int x u^{x-1} \cdot -e^{-u}\, du =$

$= 0$ $\quad x\int u^{x-1} e^{-u}\, du$

$= $ as $0^x e^{-0} = 0$ & $-u^x e^{-\infty} = 0$ as $d^n u \cdot -u^x/e^u = x!/(x-n)!\, u^{x-n}/e^{u}(-1)^n$

$n=x$ as $0 \leftarrow$

b) $\int_0^\infty u^0 e^{-u}\, du = |_0^\infty - e^{-\infty} - -e^{-0} = 1$

$\boxed{1.18}$ $S_D \int_0^\infty e^{-r^2} (r^2)^{D-1/2}\, dr = \frac{1}{2} S_D \int_0^\infty e^{-r^2} (r^2)^{\frac{D-1}{2}} 2r\, dr$

$2r\,dr = d(r^2) = du$

$= \frac{1}{2} S_D \Gamma(D/2) = (\pi)^{D/2} = \prod^D \int e^{-x_i^2}$ as $r^{2 \cdot -1/2}$ cancels $2r$

$\rightarrow S_D = 2(\pi^{D/2}) / \Gamma(D/2)$

b) ~~$\int_0^\infty u^{D/2-1} e\, du$~~ ~~$a = u^{D/2-1}$~~ ~~$(D/2-1) u^{D/2-2}$~~ ~~$db = e^{-u}$~~ ~~$b = -e^{-u}$~~

~~$= |_0^\infty e^{-u} u^{D/2-1} + \int e^{-u} u^{D/2-2} (D/2-1)\, du$~~

b) $V_D = \int_0^1 S_D \cdot r^{D-1}\, dr = S_D |_0^1 r^D/D = S_D/D$

c) $S_2 = 2\pi$ $\quad V_2 = \pi$; $S_3 =$

$\boxed{1.19}$ $\frac{Vol_{sp}}{Vol_{cub}} = \frac{2\pi^{D/2}}{D\,\Gamma(D/2)} \Big/ (2a)^D = \frac{2\pi^{D/2}}{D\,\Gamma(D/2)\, 2^{D-1}}$ if $a=1$

$(e^2/D)^D$

b) $2\pi^{D/2} / D\, 2^{D-1} (2\pi)^{1/2} e^{-D+1} (D/2-1)^{D/2-1/2}$

$\sqrt{2}\left(\frac{\pi}{2}\right)^{D-1} \cdot e^{D-1}/D\, (D/2)^{D/2}$ $\quad \left(\frac{\sqrt{\pi}}{2}\right)^D \to 0$ & $\frac{e^{2D-1}}{D^D} \to 0$

c) to corner $= \sqrt{\sum a^2}$ to side $= a$ if $a=1$ $\downarrow = \sqrt{D}$

9

1.20

See Muleshkov & Nguyen for proof of jacobian

$$\int_{x_i \in \text{shell}} 1/(2\pi\sigma^2)^D \cdot \exp(-\|X\|/2\sigma^2)\, dx_i\text{'s} = \int_r^{r+\epsilon} p(r) \cdot r^{D-1} \cdot S_D\, dr$$

treating $p(r)$ as constant over shell (& surface area) gives $P(r)\,[\,r^{D-1} S_D \cdot (r+\epsilon - r)$

$$\Rightarrow p(r) = \frac{S_D\, r^{D-1}}{(2\pi\sigma^2)^{D/2}} e^{-r^2/2\sigma^2}$$

b) $p'(r) = (D-1)\, r^{D-2} \cdots + \frac{S_D}{\cdots} \cdots -2r/2\sigma^2\, e^{-r^2/2\sigma^2}$

$$= \left(\frac{D-1}{r} - r/\sigma^2\right) p(r) \rightarrow \sigma^2(D-1) = r^2 \rightarrow \hat{r} \approx \sigma\sqrt{D}$$

c) $p(r+\epsilon) = (S_D/\cdots)\, r^{(D-1)}\, e(-(r+\epsilon)^2/2\sigma^2) = S_D/\cdots\; r^{D-1}(1+\epsilon/r)^{D-1}\, e^{\wedge}\cdots$

$$= P(r)\cdot e^{\wedge}(-2r\epsilon/2\sigma^2 - \epsilon^2/2\sigma^2 + (D-1)\ln(1+\epsilon/r)$$

2nd order taylor approx:

$$= e^{\wedge}(\qquad (D-1)\cdot(\epsilon/r - \epsilon^2/r^2 \cdot 2)$$

$$= P(r)\cdot e^{\wedge}(-r\epsilon/\sigma^2 - \epsilon^2/2\sigma^2 + \cancel{\hat{r}^2/\sigma^2}(\epsilon/r - \epsilon^2/2r^2)$$

$$= P(r)\cdot e^{\wedge}(-2\epsilon^2/2\sigma^2)$$

d) $P(\hat{r})/P(0) = e^{-\hat{r}^2/2\sigma^2} / e^{0} \cdot 0^{D-1} = e^{-(\sqrt{D}\sigma)^2/2\sigma^2} = e^{-D/2}$

1.21

$a = \sqrt{a}\cdot\sqrt{a} \le \sqrt{a}\cdot\sqrt{b}$ since $\sqrt{a} \le \sqrt{b}$

P(mistake) $= P(C_2, C_1) > P(C_2, C_2) \mid P(C_1, C_1) < P(C_1, C_2)$

$$= \int_{x \in C_1} P(x, C_2) > P(X, C_1) + \int_{x\in C_2} P(X, C_1) > P(X, C_2)$$

$$\le \int \sqrt{\cdots} + \int \sqrt{\cdots}$$

$$\le \int_x \sqrt{P(X, C_2)\cdot P(X, C_1)}$$

$\boxed{1.22}$ $L = \min \sum_{k=\text{true}} \sum_{j=\text{assigned}} \int_{R_j} L_{kj}\, P(x, C_k)\, dx$; since $L_{kj} = 1$ if $k \neq j$

$\min = -(N(k) - 1) \cdot \sum_k \int P(x, C_k)\, dx$; $\sum_k \min -\int P(C_k|x)$

For class k max $\int P(C_k|x)$

$\boxed{1.23}$ X A constant loss for miss classifying

$= \sum_k \sum_j L_{kj} \int P(x, C_k)\, dx \quad \sum_k \sum_j L_{kj} \int P(C_k|x)$

$\boxed{1.24}$ minimizing $P[P(C_j|x) > \theta$ & $\exists k \neq j$ s.t. $P(C_k|\,) > P(C_j|\,)] \cdot L_{kj}$
$+ \; P[P(C_k|\,) < \theta] \cdot \lambda$

choose j if $\sum_{k \neq j} E[P(C_k | x \in j)] \cdot L_{kj} < \lambda$

$1 - P(C_j)$ given L

loss $= \sum_{\text{true}=j} 1 - P(C_j|x)$ if $P(C_j|\,) \geq \theta$
λ else

$1 - P(C_j|x) = \lambda$ when $P(C_j|x) = \theta$
$1 - \theta = \lambda$

$\boxed{1.25}$ $E[L(t, y(x))] = \iint \| y(x) - t \|^2 P(x,t)\, dx\, dt$ $\quad E[t|x] + \sigma(x) = y(x)$ Treating each x as independent

$\Rightarrow \frac{\partial E[L]}{\partial y(x)} = 2 \int (y(x) - t)\, P(x,t)\, dt = 0$

$\rightarrow \int y(x)\, P(x,t)\, dt = y(x) \int P(x,t)\, dt = \int t\, P(x,t)\, dt$

$y(x) = E[t|x]$

$\boxed{1.26}$ $\frac{\partial}{\partial y(x)} E[L_q] = 0 + \iint q (y(x) - t)^{q-1} \cdot \text{sign}(y(x) - t)\, P\, dt\, dx = 0$

$q \int -\int_{-\infty}^{y(x)} (y(x) - t)^{q-1} P\, dt + \int_{y(x)}^{\infty} (y(x) - t)^{q-1} P\, dt\, dx$ $\quad$ xs indep. $P(x,t) = P(t|x)$ conditional

$\int_{-\infty}^{y(x)} 1^0\, P(t|x)\, dt = \int_{y(x)}^{\infty} w^0 \cdot P(t|x)\, dt$

$P[t \leq y(x) | x] = P[t \geq y(x) | x]$

11

$\boxed{1.27}$ crit

as $\lim_{q \to 0}$ becomes uniformly 1 w/ discontinuity ~~step~~ fn around $y(x) = t$

minimize loss by maximizing region prob mass $y(x) = t$; mode

$\boxed{1.28}$* what is "information"?

2 independent samples from $p = h(p,p) = h(p) + h(p) = 2h(p)$

if got $1/m$th $p = h(p^{1/m}) = 1/m\, h(p)$ got $1/m$th the information of p. $1/m$th basis for all rational numbers $1/m = 1/q$.

$\forall k > 0 \quad k = p^{\ln p(k)} \qquad h(p^x) = x h(p) \rightarrow h(k) = \ln p(k)\, h(p) \propto \ln(k)$

$\boxed{1.29}$

$H[x] = -\sum_{x_i=1}^{M} \ln(p(x)) \cdot p(x) \qquad E[p(x)] = 1/M$ since $\ln$ is concave $d/dx^2 \ln(x) = -1/x$

$E[H[x]] \leq H[E[x]] = -\sum \ln(1/M) \cdot 1/M = \ln(M)$

$H[x] \leq \ln(M)$ since entropy a functional operator. $E[H[x]] = H[x]$

$\boxed{1.30}$

$\int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-u)^2/2\sigma^2} \cdot \ln(s/\sigma) \cdot \frac{(x-u)^2}{2\sigma^2} \cancel{+} \frac{(x-m)^2}{2s^2}\, dx$

~~$\frac{\ln(s/\sigma)}{\sqrt{2\pi\sigma^2}}$ ... dx~~ $\quad x^2 - 2xu + u^2$

$\ln(s/\sigma) \int N(x) \left(\frac{1}{2s^2} - \frac{1}{2\sigma^2}\right) x^2 + \int N(x)\, 2x \left(-\frac{m}{2s^2} + \frac{u}{2\sigma^2}\right)$

$+ \int N(x) \left(\frac{m^2}{2s^2} - u^2/\sigma^2\right) dx$

$\ln(s/\sigma) \cdot \left[\left(\frac{1}{2s^2} - \frac{1}{2\sigma^2}\right) \cdot (u^2 + \sigma^2) + \left(\frac{u}{\sigma^2} - \frac{m}{s^2}\right) u + \left(\frac{m^2}{s^2} - u^2/\sigma^2\right) \cdot 1\right.$

$\boxed{1.31}$

$-\iint p(x,y) \lg(p(x,y))\, dx\, dy = \iint p(y|x)\, p(x) \lg(p(y|x)\, p(x)) =$

$= -\iint p(y|x) \cdot p(x) \cdot \lg(p(y|x)) + p(y|x)\, p(x) \lg(p(x))$ if indep

$= -\int p(y) \cdot \lg(p(y))\, dy \int p(x)\, dx + \int p(y)\, dy \int p(x) \lg(p(x))\, dx = H[y] + H[x]$

else subtract $H[y] + H[x] = H[y|x] \pm H[y]$ (?)

~~$= -\iint p(y|x)\, p(x) \lg(p(y|x))$~~ ~~$= -KL(p(x,y) \| p(x)) < 0$~~ by assump

~~$+ \iint p(x,y) \cdot \lg p(x)\, dx\, dy$~~

~~first term $= \int p(y) \lg(p(y))\, dy = H[y] = 0$, $H[x]$~~ =

but $H[y|x] < H[y]$ if x, y correlated.

PRML

1.32 $H[y] = -\int_x \left|\frac{dx}{dy}\right| p(x) \cdot \lg\left(p(x)\left|\frac{dx}{dy}\right|\right) \cdot \left|\frac{dy}{dx}\right| dx = -\int p(x)\ln(x)dx - \int p(x)\lg(|A|^{-1})$
$= H[x] + \lg(|A|)$

1.33 $0 = H[y|x] = -\sum_x \sum_y p(x,y) \cdot \lg(p(y|x))$ if $p(y|x) \neq 1 \to p(y|x) \leq 1/2$
$p(x,y) \geq 0$ $\to \lg(p(y|x)) \leq -1$
$p(y|x) \neq 0 \leftrightarrow p(x,y)/p(x) \neq 0 \leftrightarrow p(x,y) \neq 0$
so if $\exists y$ s.t. $p(y|x) \neq 1$ $H[y|x] > 0$

1.34 if $\frac{\partial G}{\partial y} - d/dx\left(\partial G/\partial y'\right) = 0$ where $\partial G/\partial p(x) = \lg(p(x)) + p(x)/p(x)$
$\partial G/\partial y' = 0$
$0 = (-1 + \lambda_1 + \lambda_2 x + \lambda_3 (x-u)^2) + 1$

1.35 $H(x) = -\int p(x) \lg(1/k\, e^{-f(x)}) = -\int p(x)(-\lg(\sqrt{k}) - f(x))$
$= \frac{1}{2}\lg(2\pi\sigma^2)\int p(x) + \frac{1}{2\sigma^2}\int p(x)(x-u)^2 = \frac{1}{2\sigma^2}\cdot\sigma^2 = 1/2$
$= 1/2(\lg(2\pi\sigma^2) + 1)$

1.36 for convex region $[a,b]$, $\epsilon = b-a/2$: of fi $f(a) + f(b) > 2f(a+\epsilon)$
by taylor approx: $f(a) + f(a) + f'(a)2\epsilon + f''(a)(2\epsilon)^2 > 2f(a) + 2f'(a)\epsilon + f''(a)\epsilon^2$
$\leftrightarrow f''(a)2\epsilon^2 + O(\epsilon^3) > f''(a)\epsilon^2 \; 2O(\epsilon^3) \leftrightarrow f''(a)\epsilon^2 > 2O(\epsilon^3) - O((2\epsilon)^3)$
$2 \cdot \max f'''([a,a+\epsilon])/3! \cdot \epsilon^3 - \max f'''([a,b])/3! \cdot 8\epsilon^3 \sim 0$
but since f strictly convex if $-f'''(c) > f''(c)$ then chord on $[c-\epsilon, c+\epsilon]$ would be concave, as 2nd taylor approx would be over est.

1.37 $H[x|y] = -\iint p(x,y)\ln(p(x|y)) = -\iint p(x,y)\ln(p(x,y)) + \iint p(x,y)\ln(p(y))\,dx\,dy$
$\iint p(x,y)\,dx\,dy = \iint p(x|y)\,dx\,p(y)\,dy = \int p(y)\,dy \to$
$= H[x,y] - H[y]$

1.38 1.114 is base case. let $\lambda'_{m+1} = k \in (0,1) \to \lambda'_i = \lambda_i(1-\lambda'_{m+1})$
$f(\sum \lambda_i x_i)\cdot(1-\lambda'_{m+1}) + f(x_{m+1})\cdot\lambda'_{m+1} \leq (\sum^m \lambda_i f(x_i))\cdot(1-\lambda'_{m+1}) + f(x_{m+1})\cdot\lambda'_{m+1}$
but $\sum \lambda_i x_i \in [\min(x_i), \max(x_i)]$ and by base case $\leq \sum \lambda_i f(x_i)$
$f(\sum \lambda_i x_i \cdot (1-\lambda'_{m+1}) + x_{m+1}\cdot\lambda'_{m+1}) \leq$
$= f(\sum^{m+1} \lambda'_i x_i)$

13

[1.39] a) $H[x] = P_x(0) \cdot \ln(2/3) + P_x(1) \cdot \ln(1/3) = 0.63$

b) $H[y] = H[x]$

c) $H[y|x] = P_{xy}(0,0) \cdot \ln(1/2) + P(0,1)\ln(1/2) + P(1,1)\ln(1) = 0.46$

d) $H[x|y] = 0$

e) $H[x,y] = 3 \cdot (1/3 \cdot \ln(1/3)) = 1.09$

f) $I[x,y] = -1/3 \ln(1/2) - 1/3 \ln(1/2) - 1/3 \ln(1) = 0.46$

[1.40] $\ln(x)$ is concave as $d/dx^2 = -1/x^2 < 0 \rightarrow -\ln(x)$ is convex

$-\ln(\sum x_i \cdot 1/n) \leq -\sum 1/n \ln(x_i) = -1/n \ln(\prod x_i) = -\ln(\sqrt[n]{\prod x_i})$

negate; apply monotonic exp $\rightarrow \sum x_i \cdot 1/n \geq \sqrt[n]{\prod x_i}$

[1.41] $-\iint p(x,y)\left[\ln(p(x)) + \ln(p(y)) - \ln(p(x,y))\right] dx\,dy$

$= H[x] + H[y] - H[x,y]$

$= H[x] - H[x|y] = H[y] - H[y|x]$