$X \sim N(\mu, S)$ | $y = w^T x + b$ | $P(y) = c\, e^{-(y-\mu_y)^2}$

Constraint $\mu_y^2 = \sigma_y^2$

$E(y) = w^T E(x) + b = w^T \mu + b$ [0.5]

$\sigma_y^2 = w^T S w$ [0.5]

Since N iids are available: likelihood function

$P(Y|w) = \prod_{i=1}^{N} \cancel{\frac{1}{\sigma}}\, c \exp\{-(y_i - \mu_y)^2\}$

$\ln P(Y/W) \equiv -\sum_{i=1}^{N} (y_i - w^T \mu - b)^2$ [1]

To find w, we use MLE with constraint $\mu_y^2 = \sigma_y^2$

$$\max_{w} \ln P(Y|\theta) \quad \text{s.t.} \quad \mu_y^2 = \sigma_y^2$$

$$\max_{w} -\sum_i (y_i - w^T\mu - b)^2 \quad \text{s.t.} \quad (w^T\mu + b)^2 = w^T S w$$ [1]

Using Lagrange multiplier λ

$$\max_{w} -\sum_i \left\{ (y_i - b)^2 - 2(y_i - b)w^T\mu + w^T\mu\mu^T w \right\}$$
$$- \lambda \left[w^T\mu\mu^T w + 2w^T\mu b + b^2 - w^T S w \right]$$ [0.5]

Taking derivative w.r.t w

$$-\sum_i \left\{ 0 - 2(y_i - b)\mu + 2\mu\mu^T w \right\} - \lambda\left[2\mu\mu^T w + 2\mu b - 2Sw\right]$$

$$\sum_i (y_i - b)\mu - b\mu\lambda = \left(\sum_i \mu\mu^T + \lambda\mu\mu^T - \lambda S\right) w$$

$$w = \left[(N+\lambda)\mu\mu^T - \lambda S\right]^{-1} \left\{ -\lambda b\mu + \sum_i (y_i - b)\mu \right\}$$

[0.5]

Q2. X_1, X_2 data matrices.

$y_1 = w^T X_1 \mid y_2 = w^T X_2$

$y = [y_1 \; y_2] \in \{0,1\}^N$

$$P(y|\theta) = \prod_{i=1}^{N} \theta^{y_i} (1-\theta)^{1-y_i}$$

[1] $$\ln P(y|\theta) = \sum_i y_i \theta + (1-y_i) \ln(1-\theta)$$

$$= \sum_i w^T x_i \theta + (1 - w^T x_i) \ln(1-\theta)$$

x_i is i^{th} column of $X = [X_1, X_2]$

Prior on $w \to$ α

$$P(w) \propto e^{-w^T w/2}$$

For MAP, & using condition.

$$\max_w \sum_i w^T x_i \theta + (1 - w^T x_i) \ln(1-\theta) - \frac{1}{2} w^T w \quad \text{st. } w^T \mu_1 = w^T \mu_2$$

[2]

Forming Lagrang & taking derivative

$$\sum_i \{\theta x_i - x_i \ln(1-\theta)\} - w - \lambda(\mu_1 - \mu_2) = 0$$

$$W = \sum_i x_i \ln \frac{\theta}{1-\theta} - \lambda(\mu_1 - \mu_2)$$

$$\sum_i x_i = \frac{N}{2}(\mu_1 + \mu_2) \qquad [1]$$

Q3) $e = y - W^T x = q$

$q \sim N(\mu, S)$

$$P(q) = P(e) = \frac{1}{(2\pi)^{d/2}|S|^{1/2}} e^{-\{(q-\mu)^T S^{-1}(q-\mu)\}}$$

For MAP, $\max_{\mu} \sum_i -(q_i-\mu)^T S^{-1}(q_i-\mu) - \mu^T\mu$ [1]

$$\frac{\partial}{\partial\mu}\left[\left\{\sum_i\left(q_i^T S^{-1} q_i - q_i^T S^{-1}\mu - \mu^T S^{-1} q_i + \mu^T S^{-1}\mu\right)\right\} - \mu^T\mu\right]$$

$$= \sum_i 2S^{-1} q_i - N2S^{-1}\mu - 2\mu = 0$$ [1]

$$\sum_i q_i - N\mu - S\mu = 0$$

$$(S + NI)\mu = \sum_i Y_i - \sum_i W^T x_i$$

$$(S+NI)\mu = N\mu_y - W^T N\mu_x$$

$$\mu = (S+NI)^{-1}\left[\mu_y - W^T\mu_x\right]N$$

[1]

Q4. $P(y|\omega_1) = \frac{1}{Z} e^{-\frac{1}{2} y^T y}$

$$P(y|\omega_2) = \frac{1}{Z} e^{-\frac{1}{2}(y-\mu^T u)^T (y-\mu^T u)}$$

$$P(y|\omega_3) = \frac{1}{Z} e^{-\frac{1}{2}(y+\mu^T u)^T (y-\mu^T u)}$$

$$\max_u P(y|\omega_1) P(y|\omega_2) P(y|\omega_3)$$

$$= \max_u -\frac{1}{2} y^T y - \frac{1}{2}(y-\mu^T u)^T (y-\mu^T u) - \frac{1}{2}(y+\mu^T u)^T (y+\mu^T u) \quad [1]$$

interpreting FDA.

$$\max_u -\frac{1}{2} y^T y - \frac{1}{2}\left[y^T y - y^T \mu^T u - u^T \mu y + u^T \mu \mu^T u\right] - \frac{1}{2}\left[y^T y + y^T \mu^T u + u^T \mu y + u^T \mu \mu^T u\right] + u^T S_B u - \lambda(u^T S_w u - 1) \quad [1]$$

$$= \max_u u^T \cancel{\mu} y - u^T \mu \mu^T u + u^T S_B u - \lambda(u^T S_w u - 1)$$

$$\max_u u^T (S_B - \mu\mu^T) u - \lambda(u^T S_w u - 1) \quad [1]$$

Derivative.

$$2(S_B - \mu\mu^T) u - 2\lambda S_w u = 0$$

u - eigenvector of

$$S_w^{-1}\left[S - S_w - \mu\mu^T\right]$$

Scatter S_W = $\frac{\text{Cov. Matrix}}{N-1}$ $\quad [1]$

$$S_1 = \frac{I}{N-1} = S_2 = S_3$$

$$S_W = 3(N-1)\, I$$