PS1

Q2 given $p(y^{(i)}=1 \mid t^{(i)}=1, x) = p(y^{(i)}=1 \mid t^{(i)}=1)$

$$P(t^{(i)}=1 \mid x^{(i)}) = \frac{P(x^{(i)} \mid t^{(i)}=1) \times P(x^{(i)})}{P(t^{(i)}=1)}$$

$$= \frac{P(x^{(i)})}{P(t^{(i)}=1)} \times P(x^{(i)} \mid t^{(i)}=1, y^{(i)})$$

$$= \frac{P(x^{(i)})}{P(t^{(i)}=1)} \times \frac{P(y^{(i)} \mid t^{(i)}=1, x^{(i)}) \times P(y^{(i)})}{P(x^{(i)})}$$

$$=$$

$$P(y^{(i)}=1, t^{(i)}=1, x^{(i)}) = P(y^{(i)}=1 \mid t^{(i)}=1, x^{(i)}) \times P(t^{(i)}=1 \mid x^{(i)}) \times P(x^{(i)})$$

$$= P(t^{(i)}=1 \mid y^{(i)}=1, x^{(i)}) \times P(y^{(i)}=1 \mid x^{(i)}) \times P(x^{(i)})$$

$$\Rightarrow P(t^{(i)}=1 \mid x^{(i)}) = P(y^{(i)}=1 \mid x^{(i)}) \times \frac{\overset{1}{P(t^{(i)}=1 \mid y^{(i)}=1, x^{(i)})}}{P(y^{(i)}=1 \mid t^{(i)}=1, x^{(i)})}$$

$$= \frac{P(y^{(i)}=1 \mid x^{(i)})}{\underbrace{\boxed{P(y^{(i)}=1 \mid t^{(i)}=1)}}_{\alpha}}$$

PAGE No 26
DATE / /

(b) ~~for $x' \in V^+ \Rightarrow P(t^{(i)}=1 | x$~~

for all $x^{(i)} \in V^+$,

$$P(t^{(i)}=1)=1$$
$$P(y^{(i)}=1)=1$$

now, $h(x^{(i)}) \approx p(y^{(i)}=1 | x^{(i)})$ for all $x^{(i)}$

clearly, $\alpha = \dfrac{P(y^{(i)}=1 | x^{(i)})}{P(t^{(i)}=1 | x^{(i)})}$

for all $x^{(i)} \in V_+$, $h(x^{(i)}) \approx P(y^{(i)}=1 | x^{(i)})$

$$\Rightarrow h(x^{(i)}) \approx \alpha.$$

(c) Code
(d) Code
(e) Code.

Q3 (a) $p(y;\lambda) = \dfrac{e^{-\lambda}\lambda^y}{y!}$

$$= \frac{1}{y!} e^{-\lambda + y\log\lambda} = \frac{1}{y!} e^{(\log\lambda)y - \lambda}$$

$\Rightarrow b(y) = 1/(y!)$
$\eta = \log\lambda \Rightarrow \lambda = e^{\eta}$
$T(y) = y$
$a(\eta) = e^{\eta}$

(b) $\mu = E[y;\eta] = g(\eta) = e^{\eta}$ (λ)

$= e^{\theta^T x}$

$\Rightarrow g(\cdot) = e^{\cdot}$

canonical response f^n is exponentiation.

(c) $h_\theta(x) = g(\theta^T x) = e^{\eta = \theta^T x} \Rightarrow p(y^{(i)} \mid x^{(i)}; \theta) = \dfrac{e^{-e^{\theta^T x}} (e^{\theta^T x})^y}{y!}$

$$\mathcal{L}(\theta) = p(y \mid x; \theta)$$

$$= \prod_{i=1}^{m} p(y^{(i)} \mid x^{(i)}; \theta)$$

$$= \prod_{i=1}^{m} \frac{e^{-g(\eta)} \times (g(\eta))^{y^{(i)}}}{y!}$$

$$\cancel{\log}\ l(\theta) = \log \mathcal{L}(\theta) = \sum_{i=1}^{m} \left\{ -g(\eta) + y^{(i)} \log g(\eta) + \log y! \right\}$$

$$\Rightarrow \frac{\partial l(\theta)}{\partial \theta_j} = \sum_{i=1}^{m} \left\{ -\frac{\partial g(\eta)}{\partial \eta} \times \frac{\partial \eta}{\partial \theta_j} + \frac{y^{(i)}}{g(\eta)} \times \frac{\partial g(\eta)}{\partial \eta} \times \frac{\partial \eta}{\partial \theta_j} \right\}$$

$$\cancel{= \sum_{i=1}^{m} \left\{ -\frac{\partial(\theta^T x^{(i)})}{\partial \theta_j} + \frac{y^{(i)}}{g(\eta)} \times \partial(\theta^T x^{(i)}) \right.}$$

$$= \sum_{i=1}^{m} \left\{ -1 + \frac{y^{(i)}}{g(\eta)} \right\} \times \frac{\partial g(\eta)}{\partial \eta} \times \frac{\partial \eta}{\partial \theta_j}$$

$$= \sum_{i=1}^{m} \frac{(y^{(i)} - h_\theta(x^{(i)}))}{\cancel{g(\eta)}} \times \cancel{g(\eta)} \times x_j^{(i)}$$

$$= \sum_{i=1}^{m} (y^{(i)} - h_\theta(x^{(i)}))\, x_j^{(i)}$$

$\left\{ g(\eta) = h_\theta(x^{(i)}) \right.$
$\dfrac{\cancel{\partial \log g(\eta)}}{\partial \eta} = g(\eta)$
$\because g(\eta) = e^{\eta}$

$\Rightarrow$ update rule, (gradient ascent)

$$\theta_j := \theta_j + \alpha \sum_{i=1}^{m} (y^{(i)} - h_\theta(x^{(i)}))\, x_j^{(i)}$$

in practice divide by m

PS1 contd.

Q4 modified exponential -

$$p(y;\eta) = b(y)\exp(\eta y - a(\eta))$$

(a) $E[Y|X;\theta] = \int_Y Y\, p(Y|X;\theta)\, dY$

$$E[Y|X;\theta] = \int_y y\, p(y|x;\theta)\, dy$$

$$\& \quad \eta = \theta^T x$$

$$= \int_y \frac{y\, b(y)\, e^{\eta y}}{e^{a(\eta)}}\, dy = \int_y y\, b(y)\, e^{\eta y - a(\eta)}\, dy$$

$$\frac{\partial E[Y|X;\theta]}{\partial \eta} = \int_y y\, b(y)\, e^{\eta y - a(\eta)} \times \left(y - \frac{\partial a(\eta)}{\partial \eta}\right) dy$$

$$\frac{\partial}{\partial \eta} \int_y p(y|x;\theta)\, dy = \int_y b(y)\, e^{\eta y - a(\eta)} \times \left(y - \frac{\partial a(\eta)}{\partial \eta}\right) dy$$

$$0 = \int_y y\, b(y)\, e^{\eta y - a(\eta)}\, dy - \frac{\partial a(\eta)}{\partial \eta} \underbrace{\int p(y|x;\theta)\, dy}_{1}$$

$$\Rightarrow \boxed{\frac{\partial a(\eta)}{\partial \eta} = E[Y|X;\theta]}$$

(b) $$Var[Y|X;\theta] = \int_y y^2 p(y|x;\theta)\,dy - (E[Y|X;\theta])^2$$

now,

$$\frac{\partial^2 a(\eta)}{\partial \eta^2} = \frac{\partial E[Y|X;\theta]}{\partial \eta}$$

$$= \int_y y\, b(y)\, e^{\eta y - a(\eta)} \left(y - \frac{\partial a(\eta)}{\partial \eta}\right) dy$$

$$= \int_y y^2 p(y|x;\theta)\,dy - (E[Y|X;\theta])^2$$

$$\Rightarrow \boxed{\frac{\partial^2 a(\eta)}{\partial \eta^2} = Var(Y|X;\theta)}$$

(c) $$\mathcal{L}(\theta) = \prod_{i=1}^{m} p(y^{(i)}|x^{(i)};\theta)$$

$$l(\theta) = \log \mathcal{L}(\theta)$$

$$l(\theta) = \sum_{i=1}^{m} \log p(y^{(i)}|x^{(i)};\theta)$$

$$= \sum_{i=1}^{m} \{\log b(y^{(i)}) + \eta y^{(i)} - a(\eta)\}$$

$$\Rightarrow l(\theta) = \sum_{i=1}^{m} \{\log b(y^{(i)}) + (\theta^T x^{(i)}) y^{(i)} - a(\theta^T x^{(i)})\}$$

$$NLL = -l(\theta)$$

$$loss = \sum_{i=1}^{m} \{(\theta^T x^{(i)}) y^{(i)} - a(\theta^T x^{(i)})\}$$

$$\frac{\partial\, loss}{\partial \theta_j} = \sum_{i=1}^{m} \{ \cancel{x_j^{(i)}}\, y^{(i)} - E[y^{(i)} | x^{(i)}; \theta] \} x_j^{(i)}$$

$$\cancel{= \sum_{i=1}^{m}}$$

$$H_{jk} = \frac{\partial\, loss}{\partial \theta_k \partial \theta_j}$$

$$= \sum_{i=1}^{m} a''(\eta)\, x_k^{(i)} x_j^{(i)}$$

$$\Rightarrow z^T H z = \sum_{j=1}^{n} \left(\sum_{k=1}^{n} \left(\sum_{i=1}^{m} a''(\eta)\, x_k^{(i)} x_j^{(i)} \right) z_k \right) z_j$$

$\because$ summation order does not matter,

$$z^T H z = \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{n} a''(\eta) (x_k^{(i)} z_k)(x_j^{(i)} z_j)$$

$$= \sum_{i=1}^{m} \underbrace{a''(\eta)}_{\hookrightarrow Var[Y|X;\theta]} ((x^{(i)})^T \cdot z)^2$$

≥ 0

$\Rightarrow$ PSD.

Q5 (a) $J(\theta) = \frac{1}{2}\sum_{i=1}^{m} w^{(i)} (\theta^T x^{(i)} - y^{(i)})^2$

i) $X \in \mathbb{R}^{m\times n}$

$\theta \in \mathbb{R}^n$

$$\Rightarrow X\theta \in \mathbb{R}^m \equiv \begin{bmatrix} - x^{(1)T}\theta - \\ - x^{(2)T}\theta - \\ \vdots \\ - x^{(m)T}\theta - \end{bmatrix}_{m\times 1} \equiv \begin{bmatrix} - \theta^T x^{(1)} - \\ - \theta^T x^{(2)} - \\ - \theta^T x^{(3)} - \\ \vdots \\ - \theta^T x^{(m)} - \end{bmatrix}_{m\times 1}$$

$$\Rightarrow X\theta - y = \begin{bmatrix} \theta^T x^{(1)} - y^{(1)} \\ \theta^T x^{(2)} - y^{(2)} \\ \vdots \\ \theta^T x^{(m)} - y^{(m)} \end{bmatrix}$$

$$W(X\theta - y) = W \begin{bmatrix} \theta^T x^{(1)} - y^{(1)} \\ \vdots \\ \theta^T x^{(m)} - y^{(m)} \end{bmatrix}$$

$W = \frac{1}{2}\mathrm{diag}(w^{(1)} \quad w^{(2)} \ldots\ldots w^{(m)}]$, $W \in \mathbb{R}^{m\times m}$

$$\Rightarrow W(X\theta - y) = \begin{bmatrix} \frac{1}{2} w^{(1)} (\theta^T x^{(1)} - y^{(1)}) \\ \vdots \\ \frac{1}{2} w^{(m)} (\theta^T x^{(m)} - y^{(m)}) \end{bmatrix}$$

clearly, $(X\theta - y)^T W (X\theta - y) = \frac{1}{2}\sum_{i=1}^{m} w^{(i)} (\theta^T x^{(i)} - y^{(i)})^2$ $= J(\theta)$

(ii) $J(\theta) = (\theta^T X^T - y^T) W (X\theta - y)$

$= (\theta^T X^T W - y^T W)(X\theta - y)$

$= \theta^T (X^T W X)\theta - \theta^T X^T W y - y^T W X\theta + y^T y$

note that $W = W^T$ $\because$ W is diagonal + square

$\Rightarrow (X^T W X)^T = X^T W^T X$

$= X^T W X$

$\Rightarrow X^T W X$ is symmetric.

also, $y^T W X \theta \in \mathbb{R}$.

$\Rightarrow$ $\theta^T X^T W y \in \mathbb{R}$ & are the same.

$$J(\theta) = \theta^T (X^T W X)\theta - 2(y^T W X)\theta + y^T y$$

$$\nabla_\theta J(\theta) = 2(X^T W X)\theta - 2X^T W y$$

$$\nabla_\theta J(\theta) = 0 \Rightarrow (X^T W X)\theta = X^T W y$$

$$\boxed{\theta = (X^T W X)^{-1} X^T W y}$$

iii) $$p(y^{(i)} \mid x^{(i)}; \theta) = \frac{1}{\sqrt{2\pi (\sigma^{(i)})^2}} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right)$$

$$\mathcal{L}(\theta) = p(y \mid x; \theta) \cancel{\left(= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\,\sigma^{(i)}}\right)} = \prod_{i=1}^{m} p(y^{(i)} \mid x^{(i)}; \theta)$$

$$= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\,(\sigma^{(i)})} \exp\left(-\frac{(y^{(i)} - \theta^T(x^{(i)}))^2}{2(\sigma^{(i)})^2}\right)$$

$$\ell(\theta) = \log \mathcal{L}(\theta)$$

$$l(\theta) = \sum_{i=1}^{m} \left\{ \frac{-1}{2} \log\left(2\pi(\sigma^{(i)})^2\right) - \frac{1}{2(\sigma^{(i)})^2} \times (y^{(i)} - \theta^T x^{(i)})^2 \right\}$$

$\therefore$ maximize $l(\theta) \Rightarrow$ minimize $(-l(\theta))$

$$\Rightarrow \min_{\theta} \sum_{i=1}^{m} \frac{1}{2(\sigma^{(i)})^2} (y^{(i)} - \theta^T x^{(i)})^2$$

$$\therefore w^{(i)} = \frac{1}{(\sigma^{(i)})^2}$$

(b) Code

(c) Code

(d) $J(\theta) = \frac{-1}{m} \sum_{i=1}^{m} \left\{ y^{(i)} \log(h_\theta(x^{(i)})) + (1-y^{(i)}) \log(1-h_\theta(x^{(i)})) \right\}$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{-1}{m} \sum_{i=1}^{m} \left\{ y^{(i)} \times \frac{1}{h_\theta(x^{(i)})} \times x_j^{(i)} + (1-y^{(i)}) \times \frac{1}{1-h_\theta(x^{(i)})} \times -x_j^{(i)} \right\} \times h_\theta(x^{(i)}) \times (1-h_\theta(x^{(i)}))$$

$$= \frac{-1}{m} \sum_{i=1}^{m} \left\{ y^{(i)} - \cancel{y^{(i)} h_\theta(x^{(i)})} - h_\theta(x^{(i)}) + \cancel{y^{(i)} h_\theta(x^{(i)})} \right\} x_j^{(i)}$$

$$= \frac{-1}{m} \sum_{i=1}^{m} \left\{ y^{(i)} - h_\theta(x^{(i)}) \right\} x_j^{(i)}$$

$$\frac{\partial J(\theta)}{\partial \theta_k \partial \theta_j} = \frac{-1}{m} \sum_{i=1}^{m} -x_k^{(i)} x_j^{(i)} h_\theta(x^{(i)}) (1-h_\theta(x^{(i)}))$$

$$H_{jk} = \frac{1}{m} \sum_{i=1}^{m} h_\theta(x^{(i)}) (1-h_\theta(x^{(i)})) x_j^{(i)} x_k^{(i)}$$

$$\Rightarrow z^T H z = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{n} h_\theta(x^{(i)}) (1-h_\theta(x^{(i)})) (x_j^{(i)} z_j)(x_k^{(i)} z_k)$$

$$\geq 0 \quad \because \cancel{\epsilon_j} \; \Sigma_i \Sigma_j x_j z_j x_k z_k \equiv (x^T z)^2 \geq 0$$

$$\& \; h_\theta(x) \in [0,1)$$

$$\Rightarrow h_\theta(x)(1-h_\theta(x)) \geq 0 \;\; \forall x.$$

$\therefore$ H is PSD.

(b) Code.

(c) ~~P(y|x)~~

$$P(y=1 \mid x; \phi, \mu_0, \mu_1, \Sigma) = \frac{P(x \mid y=1; \mu_0, \mu_1, \Sigma) \times P(y=1)}{P(x; \mu_0, \mu_1, \Sigma)}$$

$$\hookrightarrow = p(x \mid y=1; \dots) P(y=1) + P(x \mid y=0; \dots) P(y=0)$$

$$= \frac{\exp((-1/2)(x-\mu_1)^T \Sigma^{-1} (x-\mu_1)) \times \phi}{\exp((-1/2)(x-\mu_0)^T \Sigma^{-1} (x-\mu_0))(1-\phi) + \exp((-1/2)(x-\mu_1)^T \Sigma^{-1} (x-\mu_1))(\phi)}$$

$$= \frac{\phi}{\phi + e^{\frac{+1}{2}\left\{ (x-\mu_1)^T \Sigma^{-1} (x-\mu_1) - (x-\mu_0)^T \Sigma^{-1} (x-\mu_0) \right\}} (1-\phi)}$$

$$(x-\mu_0)^T\Sigma^{-1}(x-\mu_0) - (x-\mu_1)^T\Sigma^{-1}(x-\mu_1)$$

$$= (x^T-\mu_0^T)\Sigma^{-1}(x-\mu_0) - (x^T-\mu_1^T)\Sigma^{-1}(x-\mu_1)$$

$$= \begin{pmatrix} \cancel{x^T\Sigma^{-1}x} - x^T\Sigma^{-1}\mu_0 \\ -\mu_0^T\Sigma^{-1}x + \mu_0^T\Sigma^{-1}\mu_0 \end{pmatrix} - \begin{pmatrix} \cancel{x^T\Sigma^{-1}x} - x^T\Sigma^{-1}\mu_1 \\ -\mu_1^T\Sigma^{-1}x + \mu_1^T\Sigma^{-1}\mu_1 \end{pmatrix}$$

consider $x \in \mathbb{R}^n$ not the full $X \in \mathbb{R}^{m\times n}$!

$$= 2(\mu_1-\mu_0)^T\Sigma^{-1}x + (\mu_0^T\Sigma^{-1}\mu_0 - \mu_1^T\Sigma^{-1}\mu_1)$$

$$\Rightarrow P(y=1|x;\phi,\mu_0,\mu_1,\Sigma) = \frac{1}{1+e^{(\mu_1-\mu_0)^T\Sigma^{-1}x + (\mu_0^T\Sigma^{-1}\mu_0 - \mu_1^T\Sigma^{-1}\mu_1) + \ln\left(\frac{1-\phi}{\phi}\right)}}$$

$$\Rightarrow \theta^T = (\mu_1-\mu_0)^T\Sigma^{-1} \Rightarrow \theta = \underline{\Sigma^{-1}(\mu_1-\mu_0)}$$

$$\& \ \theta_0 = \frac{1}{2}\left(\mu_0^T\Sigma^{-1}\mu_0 - \mu_1^T\Sigma^{-1}\mu_1 + \ln\left(\frac{1-\phi}{\phi}\right)\right)$$

~~don't~~ invert all signs ~~∴ (-ve) outside~~.
∵ by mistake I had taken μ_0's dist. in numerator

(d) $x \in \mathbb{R}^1$ $(X \in \mathbb{R}^{m\times 1})$
$\mu_0 \in \mathbb{R}^1$
$\mu_1 \in \mathbb{R}^1$
$\Sigma \in \mathbb{R}^{1\times 1} \in \mathbb{R}^1$

$$l(\phi,\mu_0,\mu_1,\Sigma) = \sum_{i=1}^{m}\left\{\log p(x^{(i)}|y^{(i)};\phi,\mu_0,\mu_1,\Sigma) + \log p(y^{(i)};\phi)\right\}$$

now, $$p(x^{(i)}|y^{(i)};\phi,\mu_0,\mu_1,\Sigma) = \left\{\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)\right\}^{y^{(i)}} \times \left\{\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right)\right\}^{1-y^{(i)}}$$

$\Rightarrow$ in our case, $p(x^{(i)} | y^{(i)}; \phi, \mu_0, \mu_1, \Sigma) = \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x^{(i)}-\mu_1)^2}{2\sigma^2}} \right\}^{y^{(i)}}$

$$\times \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x^{(i)}-\mu_0)^2}{2\sigma^2}} \right\}^{1-y^{(i)}}$$

also, $p(y^{(i)}; \phi) = (\phi)^{y^{(i)}} (1-\phi)^{1-y^{(i)}}$

$$\log p(x^{(i)} | y^{(i)}; \phi, \mu_0, \mu_1, \Sigma) = y^{(i)} \left(\frac{-1}{2} \log(2\pi\sigma^2) - \frac{(x^{(i)}-\mu_1)^2}{2\sigma^2} \right) + (1-y^{(i)}) \left(\frac{-1}{2} \log(2\pi\sigma^2) - \frac{(x^{(i)}-\mu_0)^2}{2\sigma^2} \right)$$

$$\log p(y^{(i)}; \phi) = y^{(i)} \log\phi + (1-y^{(i)}) \log(1-\phi)$$

$$\therefore l(\phi, \mu_0, \mu_1, \Sigma) = \sum_{i=1}^{m} \left\{ -\frac{y^{(i)}}{2} \times \frac{(x^{(i)}-\mu_1)^2}{\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2) - \frac{(1-y^{(i)})}{2} \frac{(x^{(i)}-\mu_0)^2}{\sigma^2} + y^{(i)} \log\phi + (1-y^{(i)}) \log(1-\phi) \right\}$$

$$\frac{\partial l(\phi, \mu_0, \mu_1, \Sigma)}{\partial \phi} = \sum_{i=1}^{m} \left(\frac{y^{(i)}}{\phi} + \frac{(1-y^{(i)})}{1-\phi}(-1) \right)$$

$$= \sum_{i=1}^{m} \left(\frac{y^{(i)} - \cancel{y^{(i)}\phi} - \phi + \cancel{y^{(i)}\phi}}{\phi(1-\phi)} \right)$$

setting to 0, we get, $\sum_{i=1}^{m} (y^{(i)} - \phi) = 0$

$$\phi = \frac{1}{m} \sum_{i=1}^{m} y^{(i)}$$

$$\boxed{\Rightarrow \phi = \frac{1}{m} \sum_{i=1}^{m} I\{y^{(i)} = 1\}}$$

$$\frac{\partial l(\phi,\mu_0,\mu_1,\Sigma)}{\partial \mu_0} = -\sum_{i=1}^{m} \frac{(1-y^{(i)})}{\cancel{2}} \times \frac{\cancel{2}(x^{(i)}-\mu_0)}{\sigma^2}$$

setting to 0,

$$\sum_{i=1}^{m}(1-y^{(i)})(x^{(i)}-\mu_0) = 0$$

$$\mu_0 = \frac{\sum_{i=1}^{m} x^{(i)}(1-y^{(i)})}{\sum_{i=1}^{m}(1-y^{(i)})}$$

$$\cancel{\sum_{i=1}^{m} I\{x^{(i)}}$$

$$\boxed{\mu_0 = \frac{\sum_{i=1}^{m} I\{y^{(i)}=0\}x^{(i)}}{\sum_{i=1}^{m} I\{y^{(i)}=0\}}}$$

similarly, $\boxed{\mu_1 = \frac{\sum_{i=1}^{m} I\{y^{(i)}=1\}x^{(i)}}{\sum_{i=1}^{m} I\{y^{(i)}=1\}}}$

$$\frac{\partial l(\phi,\mu_0,\mu_1,\Sigma)}{\partial \Sigma} = \sum_{i=1}^{m}\left\{ \frac{+y^{(i)}}{2}\frac{(x^{(i)}-\mu_1)^2}{\sigma^4} - \frac{1}{2} \times \frac{\cancel{2\pi}}{\cancel{2\pi}\sigma^2} + \frac{(1-y^{(i)})}{2}\frac{(x^{(i)}-\mu_0)^2}{\sigma^4}\right\}$$

setting to 0,

$$\sum_{i=1}^{m}\left\{ y^{(i)}\frac{(x^{(i)}-\mu_1)^2}{\sigma^2} - 1 + \frac{(1-y^{(i)})(x^{(i)}-\mu_0)^2}{\sigma^2}\right\} = 0$$

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)}(x^{(i)}-\mu_1)^2 + (1-y^{(i)})(x^{(i)}-\mu_0)^2\right)$$

$$\therefore \sigma^2 = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)}-\mu_{y^{(i)}})(x^{(i)}-\mu_{y^{(i)}})^T$$

(e) Code

(f) Code

(g) Dataset 1! $\because$ maybe the underlying dist. is not normal.

(h) Box Cox Transformation!

at $\lambda = 0.8$, we again have equal performance to LogR.