1. a) $J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log(h_\theta(x^{(i)})) \cancel{+(1-y}$

$+ (1-y^{(i)}) \log(1-h_\theta(x^{(i)}))$,

$y^{(i)} \in \{0,1\} \quad h_\theta(x) = g(\theta^T x) \quad g(z) = 1/(1+e^{-z})$

$$\cancel{\nabla J(\theta)}\ \frac{\partial J(\theta)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j}\Big(-\frac{1}{m}\sum_{i=1}^{m} y^{(i)}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\Big)$$

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

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

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

$$= -\frac{1}{m}\sum_{i=1}^{m}\left(\frac{y^{(i)}}{h_\theta(x^{(i)})}\cdot g'(\theta^T x^{(i)})\cdot x_j^{(i)} + \cancel{\frac{y^{(i)}}{}}\ \frac{y^{(i)}-1}{1-h_\theta(x^{(i)})}\, g'(\theta^T x^{(i)})\cdot x_j^{(i)}\right).$$

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

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

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

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

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

$$\cancel{= \frac{\partial}{\partial\theta_k}\Big(-\frac{1}{m}\sum}$$

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

$$= -\frac{1}{m}\sum_{i=1}^{m}x_j^{(i)}\cdot\frac{\partial}{\partial\theta_k}(-h_\theta(x^{(i)})) = \frac{1}{m}\sum_{i=1}^{m}g'(\theta^T x^{(i)})\cdot x_j^{(i)}x_k^{(i)}.$$

$$\therefore H_{k,j} = \frac{1}{m}\sum_{i=1}^{m} (1-h_\theta(x^{(i)}))h_\theta(x^{(i)}) \cdot x_j^{(i)} \cdot x_k^{(i)}.$$

$$z^T H z = (z_1, z_2, \cdots, z_n)\begin{pmatrix} h_1^T \\ h_2^T \\ \vdots \\ h_n^T \end{pmatrix} z.$$

$$= z_1 h_1^T z + z_2 h_2^T z + \cdots + z_n h_n^T z.$$

$$= z_1 (h_{11}\ h_{12} \cdots h_{1n})\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} + \cdots + z_n h_n^T z.$$

$$= z_1 \sum_{i=1}^{n} h_{1i} z_i + \cdots + z_n \sum_{i=1}^{n} h_{ni} z_i$$

$$= \sum_{j=1}^{n} z_j \sum_{i=1}^{n} h_{ji} z_i$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{n} z_j \cdot h_{ji} z_i$$

$$= \sum_{j=1}^{n}\sum_{k=1}^{n} z_j h_{jk} z_k$$

$$= \sum_{j=1}^{n}\sum_{k=1}^{n} z_j \left(\frac{1}{n}\sum_{i=1}^{n} (1-h_\theta(x^{(i)}))h_\theta(x^{(i)}) \cdot x_j^{(i)} x_k^{(i)}\right) \cdot z_k$$

$$= \sum_{j=1}^{n}\sum_{k=1}^{n}\frac{1}{n}\sum_{i=1}^{n}\left[(1-h_\theta(x^{(i)}))h_\theta(x^{(i)})\, z_j x_j^{(i)} x_k^{(i)} z_k\right]$$

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

$$= \frac{1}{n}\sum_{i=1}^{n}(1-h_\theta(x^{(i)}))h_\theta(x^{(i)})(z^T x^{(i)})^2.$$

$$\because h_\theta(x^{(i)}) \in (0,1)$$

$$\therefore z^T H z \geq 0$$

(c) $p(y) = \begin{cases} \phi & \text{if } y=1 \\ 1-\phi & \text{if } y=0. \end{cases}$

$$p(x|y=0) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right)$$

$$p(x|y=1) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)$$

$$p(x) = p(x|y=0)\,p(y=0) + p(x|y=1)\,p(y=1)$$

$$p(y=1|x) = \frac{p(y=1,x)}{p(x)} = \frac{p(x|y=1)\,p(y=1)}{p(x)}$$

$$= \frac{p(x|y=1)\,p(y=1)}{p(x|y=1)\,p(y=1) + p(x|y=0)\,p(y=0)}$$

$$= \frac{1}{1 + \frac{p(x|y=0)\,p(y=0)}{p(x|y=1)\,p(y=1)}}$$

$$\frac{p(x|y=0)\,p(y=0)}{p(x|y=1)\,p(y=1)} = \frac{p(y=0)}{p(y=1)} \cdot \frac{p(x|y=0)}{p(x|y=1)}$$

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

$$= \frac{1-\phi}{\phi} \cdot \exp\left[-\frac{1}{2}\left((x-\mu_0)^T\Sigma^{-1}(x-\mu_0) - (x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)\right]$$

since Σ is symmetric $\Rightarrow$ Σ^{-1} is symmetric

$\Sigma^{-1} = U^T \Lambda U$ where Λ is a diagonal matrix.

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

$$= (x-\mu_0)^T \cdot U^T \cdot \Lambda \cdot U(x-\mu_0) - (x-\mu_1)^T U^T \Lambda U (x-\mu_1)$$

$$= (U(x-\mu_0))^T \Lambda U(x-\mu_0) - ((U(x-\mu_1))^T \Lambda U(x-\mu_1))$$

let $\mathrm{diag}(v) = \Lambda$

$w_0 = U(x-\mu_0)$

$w_1 = U(x-\mu_1)$

Such that $(U(x-\mu_0))^T \Lambda U(x-\mu_0) - ((U(x-\mu_1))^T \Lambda U(x-\mu_1))$

$$= w_0^T \,\mathrm{diag}(v)\, w_0 - w_1^T \,\mathrm{diag}(v)\, w_1$$

$$= \sum_{i=1}^m v_i (w_0^{(i)})^2 - \sum_{i=1}^m v_i (w_1^{(i)})^2$$

$$= \sum_{i=1}^m v_i \left(w_0^{(i)} - w_1^{(i)}\right)\left(w_0^{(i)} + w_1^{(i)}\right)$$

$$= \begin{pmatrix} v_1(w_0^1 - w_1^1) \\ v_2(w_0^2 - w_1^2) \\ \vdots \\ v_m(w_0^m - w_1^m) \end{pmatrix}^T \cdot \begin{pmatrix} w_0^1 + w_1^1 \\ w_0^2 + w_1^2 \\ \vdots \\ w_0^m + w_1^m \end{pmatrix}$$

$$= \begin{pmatrix} w_0^1 - w_1^1 \\ w_0^2 - w_1^2 \\ \vdots \\ w_0^m - w_1^m \end{pmatrix}^T \cdot \mathrm{diag}(v) \cdot (w_0 + w_1)$$

$$= (w_0 - w_1)^T \cdot \mathrm{diag}(v) \cdot (w_0 + w_1)$$

$$w_0 - w_1 = U(x-\mu_0) - U(x-\mu_1) = U(\mu_1 - \mu_0)$$

$$w_0 + w_1 = U(2x - \mu_0 - \mu_1)$$

hence,

$$(w_0 - w_1)^T \cdot \mathrm{diag}(v) \cdot (w_0 + w_1)$$

$$= (U(\mu_1 - \mu_0))^T \cdot \Lambda \cdot U(2x - \mu_0 - \mu_1)$$

$$= (\mu_1 - \mu_0)^T U^T \Lambda U (2x - \mu_0 - \mu_1)$$

$$= (\mu_1 - \mu_0)^T A (2x - \mu_0 - \mu_1)$$

$$= 2(\mu_1 - \mu_0)^T A x - (\mu_1 - \mu_0)^T A (\mu_0 + \mu_1)$$

hence

$$p(y=1|x) = \frac{1}{1 + \frac{1-\phi}{\phi}\exp\left(-\frac{1}{2}\left[2(\mu_1-\mu_0)^T A x - (\mu_1-\mu_0)^T A(\mu_0+\mu_1)\right]\right)}$$

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

let $\theta^T = (\mu_1 - \mu_0)^T A$

$\theta_0 = -\log\frac{1-\phi}{\phi} - \frac{1}{2}(\mu_1-\mu_0)^T A(\mu_0+\mu_1)$

where $A = \Sigma^{-1}$

$$p(y=1|x) = \frac{1}{1 + \exp(-(\theta^T x + \theta_0))}$$

(d) $l(\phi,\mu_0,\mu_1,\Sigma) = \log \prod_{i=1}^{m} p(x^{(i)}, y^{(i)}; \phi, \mu_0, \mu_1, \Sigma)$

$= \log \prod_{i=1}^{m} p(x^{(i)} | y^{(i)}; \phi, \mu_0, \mu_1, \Sigma) \cdot p(y^{(i)}; \phi)$.

$= \log \prod_{i=1}^{m} \Big[\Big(\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \exp\big(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\big)\Big)^{(1-y^{(i)})}$

$\cdot \Big(\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \exp\big(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\big)\Big)^{y^{(i)}}\Big]$.

$\cdot\, \phi^{y^{(i)}}(1-\phi)^{(1-y^{(i)})}$.

$= \sum_{i=1}^{m} \log \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\Big[\exp\big(-\frac{1}{2}(x^{(i)}-\mu_0)^T\Sigma^{-1}(x^{(i)}-\mu_0)\big)^{(1-y^{(i)})}$

$\cdot \exp\big(-\frac{1}{2}(x^{(i)}-\mu_1)^T\Sigma^{-1}(x^{(i)}-\mu_1)\big)^{y^{(i)}}\Big] \cdot \phi^{y^{(i)}}(1-\phi)^{(1-y^{(i)})}$

$= \sum_{i=1}^{m} \log \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} + \sum_{i=1}^{m} \log\Big(\exp\big(-\frac{1}{2}(x^{(i)}-\mu_0)^T\Sigma^{-1}(x^{(i)}-\mu_0)\big)^{(1-y^{(i)})}$

$\cdot \exp\big(-\frac{1}{2}(x^{(i)}-\mu_1)^T\Sigma^{-1}(x^{(i)}-\mu_1)\big)^{y^{(i)}}\Big]$.

$+ \sum_{i=1}^{m} \log \phi^{y^{(i)}}(1-\phi)^{(1-y^{(i)})}.)$.

$\frac{\partial l}{\partial \phi} = \frac{\partial}{\partial \phi}\Big(\sum_{i=1}^{m} \log \phi^{y^{(i)}}(1-\phi)^{(1-y^{(i)})}\Big)$

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

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

$= \sum_{i=1}^{m}\Big(y^{(i)}\big(\frac{1}{\phi}+\frac{1}{1-\phi}\big) - \frac{1}{1-\phi}\Big)$ $\quad y^{(i)} \in \{0,1\}$.

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

let $\frac{\partial l}{\partial \phi} = 0$.

$\therefore \frac{1}{\phi}\sum_{i=1}^{m} 1\{y^{(i)}=1\} - \frac{1}{1-\phi}\Big(m - \sum_{i=1}^{m} 1(y^{(i)}=1)\Big) = 0$.

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

$\cancel{\frac{\partial l}{\partial \mu_0} = \frac{\partial}{\partial \mu_0}\Big(\sum_{i=1}^{m}\log\big(\exp(-\frac{1}{2}(x^{(i)}-\mu_0)^T\Sigma^{-1}(x^{(i)}-\mu_0))^{(1-y^{(i)})} \cdot \exp(-\frac{1}{2}(x^{(i)}-\mu_1)^T\Sigma^{-1}(x^{(i)}-\mu_1))^{y^{(i)}}\big)\Big)}$

$\cancel{= \frac{\partial}{\partial \mu_0}\sum_{i=1}^{m}[}$

$l(\phi,\mu_0,\mu_1,\Sigma)$

$= \sum_{i=1}^{m} \log \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}$

$+ \sum_{i=1}^{m}\Big[(1-y^{(i)})\big(-\frac{1}{2}(x^{(i)}-\mu_0)^T\Sigma^{-1}(x^{(i)}-\mu_0)\big)$

$+ y^{(i)}\big(-\frac{1}{2}(x^{(i)}-\mu_1)^T\Sigma^{-1}(x^{(i)}-\mu_1)\big)\Big]$

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

$= \sum_{i=1}^{m} \log \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}$

$+ \sum_{i=1}^{m}(1-y^{(i)})\big(-\frac{1}{2}(x^{(i)}-\mu_0)^T\Sigma^{-1}(x^{(i)}-\mu_0)\big)$

$+ \sum_{i=1}^{m} y^{(i)}\big(-\frac{1}{2}(x^{(i)}-\mu_1)^T\Sigma^{-1}(x^{(i)}-\mu_1)\big)$

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

when $n=1$

$\frac{\partial l}{\partial \mu_0} = \frac{\partial}{\partial \mu_0}\Big(\sum_{i=1}^{m}(1-y^{(i)})\big(-\frac{1}{2}(x^{(i)}-\mu_0)^T\Sigma^{-1}(x^{(i)}-\mu_0)\big)\Big)$

$= \sum_{i=1}^{m}(1-y^{(i)})\frac{\partial}{\partial \mu_0}\Big(-\frac{1}{2}\frac{1}{\sigma^2}(x^{(i)}-\mu_0)^2\Big)$

$= \sum_{i=1}^{m}(1-y^{(i)})\Big(-\frac{1}{2\sigma^2}\Big)2(x^{(i)}-\mu_0)(-1)$

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

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

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

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

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

Similarly, $\mu_1 = \frac{\sum_{i=1}^{m} 1\{y^{(i)}=1\}x^{(i)}}{\sum_{i=1}^{m} 1\{y^{(i)}=1\}}$.

when $n=1$ $\quad \Sigma = (\sigma^2)$

$\frac{\partial l}{\partial \Sigma} = \frac{\partial l}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\Big(\sum_{i=1}^{m}\log\frac{1}{(2\pi)^{\frac{1}{2}}(\sigma^2)^{\frac{1}{2}}} + \sum_{i=1}^{m}\Big[(1-y^{(i)})\big(-\frac{1}{2\sigma^2}(x^{(i)}-\mu_0)^2\big) + y^{(i)}\big(-\frac{1}{2\sigma^2}(x^{(i)}-\mu_1)^2\big)\Big]\Big)$

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

$$l(\phi,\mu_0,\mu_1,\Sigma)=\log\prod_{i=1}^{m}p(x^{(i)},y^{(i)};\phi,\mu_0,\mu_1,\Sigma)$$

$$=\log\prod_{i=1}^{m}p(x^{(i)}|y^{(i)};\phi,\mu_0,\mu_1,\Sigma)\,p(y^{(i)};\phi)$$

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

$$p(x^{(i)}|y^{(i)}=0;\phi,\mu_0,\mu_1,\Sigma)=\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right)$$

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

$$\cdot\left(\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right)\right)^{(1-y^{(i)})}.$$

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

$$\therefore\ l=\log\prod_{i=1}^{m}\left(\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\right)\left(\exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)\right)^{y^{(i)}}\cdot\left(\exp\left(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right)\right)^{(1-y^{(i)})}$$

$$\cdot\,\phi^{y^{(i)}}(1-\phi)^{(1-y^{(i)})}$$

$$=m\cdot\log\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}+\sum_{i=1}^{m}y^{(i)}\log\exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)$$

$$+\sum_{i=1}^{m}(1-y^{(i)})\log\exp\left(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right)$$

$$+\cancel{m\cdot\log\phi^{y^{(i)}}(1-\phi)^{(1-y^{(i)})}}\sum_{i=1}^{m}\log\phi^{y^{(i)}}(1-\phi)^{(1-y^{(i)})}.$$

$$=m\cdot\log\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}+\sum_{i=1}^{m}y^{(i)}\left(-\frac{1}{2}(x^{(i)}-\mu_1)^T\Sigma^{-1}(x^{(i)}-\mu_1)\right)$$

$$+\sum_{i=1}^{m}(1-y^{(i)})\left(-\frac{1}{2}(x^{(i)}-\mu_0)^T\Sigma^{-1}(x^{(i)}-\mu_0)\right)+\sum_{i=1}^{m}\log\phi^{y^{(i)}}(1-\phi)^{(1-y^{(i)})}$$

when $n=1$

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

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

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

let $t=\sigma^2$

$$\frac{\partial l}{\partial t}=-\frac{m}{2t}+\sum_{i=1}^{m}y^{(i)}\left(-\frac{1}{2}\cdot\left(-\frac{1}{t^2}\right)(x^{(i)}-\mu_1)^2\right)$$

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

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

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

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

$$\Sigma=t=\frac{1}{m}\sum_{i=1}^{m}(x^{(i)}-\mu_{y^{(i)}})(x^{(i)}-\mu_{y^{(i)}})^T.$$

(g) GDA works much better on dataset 2, which seems much more likely follow the Gaussian distribution.

2. a) $p(t^{(i)}=1 \mid y^{(i)}=1)=1$

$$\frac{p(t^{(i)}=1, y^{(i)}=1)}{p(y^{(i)}=1)}=1$$

$$p(t^{(i)}=1, y^{(i)}=1)=p(y^{(i)}=1) \quad ①$$

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

$$\frac{p(y^{(i)}=1, t^{(i)}=1, x^{(i)})}{p(t^{(i)}=1, x^{(i)})}=\frac{p(y^{(i)}=1, t^{(i)}=1)}{p(t^{(i)}=1)} \quad ②$$

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

$$=\frac{p(y^{(i)}=1, x^{(i)}, t^{(i)}=1)+p(y^{(i)}=1, x^{(i)}, t^{(i)}=0)}{p(t^{(i)}=1, x^{(i)})}$$

$$=\frac{p(y^{(i)}=1, x^{(i)}, t^{(i)}=1)}{p(t^{(i)}=1, x^{(i)})}. \quad \text{according to ②}$$

$$=\frac{p(y^{(i)}=1, t^{(i)}=1)}{p(t^{(i)}=1)}=\frac{p(y^{(i)}=1)}{p(t^{(i)}=1)}. \quad \text{according to ①}$$

hence α is a const number.

b) Since $h(x^{(i)}) \approx p(y^{(i)}=1 \mid x^{(i)})$ $\forall x^{(i)} \in V$

$\Rightarrow \forall x^{(i)} \in V_+ \quad h(x^{(i)}) \approx p(y^{(i)}=1 \mid x^{(i)})$.

So, I'll prove

$$\forall x^{(i)} \in V_+ \quad p(y^{(i)}=1 \mid x^{(i)}) \approx \alpha.$$

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

equalivantly,

$$\frac{p(y^{(i)}=1, x^{(i)})}{p(x^{(i)})}=\frac{p(y^{(i)}=1, x^{(i)})}{p(t^{(i)}=1, x^{(i)})}$$

equalivantly $\quad p(x^{(i)})=p(t^{(i)}=1, x^{(i)})$

for $x^{(i)} \in V_+$

Since I can assume that

$p(t^{(i)}=1 \mid x^{(i)}) \approx 1$ when $x^{(i)} \in V_+$.

$p(t^{(i)}=1, x^{(i)})=p(x^{(i)})$

~~for~~

for $x^{(i)} \in V_+$

c).

扫描全能王 创建

3. a) $p(y;\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}$

$$= \frac{1}{y!}\exp(-\lambda + y\ln\lambda)$$

$\therefore b(y) = \frac{1}{y!}$ $\quad y \in N^*$

$\eta^T = (\ln\lambda)$

$a(\eta) = \lambda$

$a(\eta) = e^{\eta}$

$\therefore p(y;\eta) = \frac{1}{y!}\exp(\eta \cdot y - e^{\eta})$

where $\eta = \ln\lambda$

b). $E(y|x^{(i)};\theta)$.

$$p(y|x^{(i)};\theta) = \frac{e^{-e^{\theta^T x^{(i)}}}\left(e^{\theta^T x^{(i)}}\right)^{y}}{y!}$$

$$E(y|x^{(i)};\theta) = \lambda = e^{\theta^T x^{(i)}}$$

c). $L(\theta) = \prod_{i=1}^{m} p(y=y_{(i)}|x_{(i)};\theta)$

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

$$= \sum_{i=1}^{m} \log\left[b(y_{(i)})\exp(\eta_{(i)} y_{(i)} - a(\eta_{(i)}))\right].$$

$$= \sum_{i=1}^{m} \left[\log b(y_{(i)}) + \eta_{(i)} y_{(i)} - a(\eta_{(i)})\right].$$

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

$$= \sum_{i=1}^{m} \log b(y_{(i)}) + \sum_{i=1}^{m} \theta^T x_{(i)} y_{(i)} - \sum_{i=1}^{m} e^{\theta^T x_{(i)}}$$

$$\frac{\partial l(\theta)}{\partial \theta_j} = \sum_{i=1}^{m} x_j^{(i)} y_{(i)} - \sum_{i=1}^{m} e^{\theta^T x^{(i)}} \cdot x_j^{(i)}$$

$$= \sum_{i=1}^{m} \left(y_{(i)} - e^{\theta^T x^{(i)}}\right) x_j^{(i)}$$

$$= \begin{pmatrix} y_1 - e^{\theta^T x^1} \\ y_2 - e^{\theta^T x^2} \\ \vdots \\ y^m - e^{\theta^T x^m} \end{pmatrix}^T \cdot \begin{pmatrix} x_j^1 \\ x_j^2 \\ \vdots \\ x_j^m \end{pmatrix}$$

~~let f(x) = e~~

$$= (y - e^{\theta^T x})^T \cdot x_j$$

$$\nabla l(\theta) = \begin{pmatrix} (y - e^{\theta^T x})^T \cdot x_1 \\ (y - e^{\theta^T x})^T \cdot x_2 \\ \vdots \\ (y - e^{\theta^T x})^T \cdot x_n \end{pmatrix}$$

$$= \left((y - e^{\theta^T x})^T \cdot x\right)^T$$

$$= x^T \cdot (y - e^{\theta^T x}).$$

Since we want to maximize $l(\theta)$, chose $J(\theta) = -l(\theta)$

~~so we want to satisfy~~ $\cancel{\nabla l(\theta) = 0}$.

So got the stochastic gradient ascent update rule:

$\cancel{\theta' := \theta + \alpha}$

$$\theta_j' := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

$$\theta' := \theta + \alpha \cdot \nabla l(\theta)$$

$$\theta' := \theta + \alpha \cdot x^T \cdot (y - e^{\theta^T x})$$

d)

4. a) $\frac{\partial}{\partial\eta}\int p(y;\eta)dy = \int \frac{\partial}{\partial\eta} p(y;\eta)dy.$

$$= \int \frac{\partial}{\partial\eta} b(y)\exp(\eta y - a(\eta))dy$$

$$= \int b(y)\exp(\eta y - a(\eta))(y - a'(\eta))dy$$

$$= \int y\, b(y)\exp(\eta y - a(\eta))dy - a'(\eta)\int b(y)\exp(\eta y - a(\eta))dy$$

$$\therefore \int y\, b(y)\exp(\eta y - a(\eta))dy$$

$$= \frac{\partial}{\partial\eta}\int b(y)\exp(\eta y - a(\eta))dy + a'(\eta)$$

$$= \frac{\partial}{\partial\eta}\int b(y)\frac{\exp(\eta y)}{\exp(a(\eta))}dy + a'(\eta).$$

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

$$= \exp(-a(\eta)) \int b(y)\exp(\eta y)\,dy.$$

$$\exp(a(\eta)) = \int b(y)\exp(\eta y)\,dy.$$

$$a(\eta) = \log \int b(y)\exp(\eta y)\,dy.$$

$$\frac{\partial}{\partial\eta} a(\eta) = \frac{\int \frac{\partial}{\partial\eta}(b(y)\exp(\eta y))\,dy}{\int b(y)\exp(\eta y)\,dy}$$

$$= \frac{\int y\, b(y)\cdot\exp(\eta y)\,dy}{\int b(y)\exp(\eta y)\,dy}$$

$$= \frac{\int y\, b(y)\exp(\eta y)\,dy}{\exp(a(\eta))}$$

$$= \int y\, b(y)\exp(\eta y - a(\eta))\,dy$$

$$= E(y)$$

$$E(y) = a'(\eta).$$

b) $\mathrm{Var}(Y|x;\theta) = E(Y^2) - E^2(Y)$

$$\frac{\partial}{\partial\eta}\int y\, b(y)\exp(\eta y - a(\eta))\,dy$$

$$= \int \frac{\partial}{\partial\eta}(y\, b(y)\exp(\eta y - a(\eta)))\,dy.$$

$$= \int y\, b(y)\exp(\eta y - a(\eta))\cdot(y - a'(\eta))\,dy.$$

$$= \int y^2 b(y)\exp(\eta y - a(\eta))\,dy$$

$$- a'(\eta)\int y\, b(y)\exp(\eta y - a(\eta))\,dy$$

$$= \int y^2 b(y)\exp(\eta y - a(\eta))\,dy - a'(\eta)\cdot a'(\eta)$$

$\cancel{= a''(\eta) +}$

$$\therefore \int y^2 b(y)\exp(\eta y - a(\eta))\,dy = a''(\eta) + a'^2(\eta)$$

$$E(Y^2) = a''(\eta) + a'^2(\eta)$$

$\cancel{\therefore V(Y)}$

$$\therefore V(Y|x;\theta) = E(Y^2) - E^2(Y) = a''(\eta)$$

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

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

$$= \sum_{i=1}^{m} \log(b(y_{(i)})\exp(\eta_{(i)} y_{(i)} - a(\eta_{(i)})))$$

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

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

$$NLL(\theta) = -\sum_{i=1}^{m}\log b(y^{(i)}) - \sum_{i=1}^{m}\theta^T x^{(i)} y^{(i)}$$

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

$$\frac{\partial}{\partial\theta_j} NLL(\theta) = -\sum_{i=1}^{m} x_j^{(i)} y^{(i)}$$

$$+ \sum_{i=1}^{m} a'(\theta^T x^{(i)})\cdot x_j^{(i)}$$

$$= \sum_{i=1}^{m} (a'(\theta^T x^{(i)}) - y^{(i)})\cdot x_j^{(i)}$$

$$= \begin{pmatrix} a'(\theta^T x^{(1)}) - y^{(1)} \\ a'(\theta^T x^{(2)}) - y^{(2)} \\ \vdots \\ a'(\theta^T x^{(m)}) - y^{(m)} \end{pmatrix}^T \cdot \begin{pmatrix} x_j^1 \\ x_j^2 \\ \vdots \\ x_j^m \end{pmatrix}$$

$$= (a'(\theta X) - y)^T \cdot x_j$$

$$\nabla NLL(\theta) = \begin{pmatrix} (a'(\theta^T x) - y)^T\cdot x_1 \\ (a'(\theta^T x) - y)^T\cdot x_2 \\ \vdots \\ (a'(\theta^T x) - y)^T\cdot x_m \end{pmatrix}$$

$$= \begin{pmatrix} x_1^T\cdot(a'(\theta^T x) - y) \\ x_2^T\cdot(a'(\theta^T x) - y) \\ \vdots \\ x_m^T\cdot(a'(\theta^T x) - y) \end{pmatrix}$$

$$= X^T\cdot(a'(\theta^T x) - y)$$

$$\frac{\partial}{\partial\theta_k}\left(\frac{\partial}{\partial\theta_j} NLL(\theta)\right) = \frac{\partial}{\partial\theta_k}\sum_{i=1}^{m}(a'(\theta^T x^{(i)}) - y^{(i)})\cdot x_j^{(i)}$$

$$= \sum_{i=1}^{m} x_j^{(i)}\frac{\partial}{\partial\theta_k}(a'(\theta^T x^{(i)}) - y^{(i)}).$$

$$= \sum_{i=1}^{m} x_j^{(i)}\cdot a''(\theta^T x^{(i)})\cdot x_k^{(i)}$$

$$H(\theta)_{j,k} = \sum_{i=1}^{m} x_k^{(i)}\cdot a''(\theta^T x^{(i)})\, x_j^{(i)}$$

$$H(\theta) = X^T\,\mathrm{diag}(a''(\theta^T x))\,X$$

~~$\forall x \in R^n$~~ ~~$x^T H x = \sum_{j=1}^{n}\sum_{k=1}^{n} x_j H_{jk} x_k$~~

$\forall\ v \in R^n$

$v^T H v = v^T X^T \operatorname{diag}(a''(\theta^T x)) X v$

$u = Xv$

$v^T H v = (Xv)^T \cdot \operatorname{diag}(a''(\theta^T x))(Xv)$

$= u^T \operatorname{diag}(a''(\theta^T x))\, u$

Since $a''(\theta^T x) = \begin{pmatrix} a''(\theta^T x^{(1)}) \\ a''(\theta^T x^{(2)}) \\ \vdots \\ a''(\theta^T x^{(m)}) \end{pmatrix}$

$= \begin{pmatrix} a''(\eta_{(1)}) \\ a''(\eta_{(2)}) \\ \vdots \\ a''(\eta_{(m)}) \end{pmatrix}$

$\because a''(\eta_i) = Var(Y|X^{(i)};\theta) \geq 0$

$\therefore \operatorname{diag}(a''(\theta^T x))$ is PSD

$\therefore v^T H v = u^T \operatorname{diag}(a''(\theta^T x))\, u \geq 0$

$\therefore H$ is PSD.

5. a)

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

$= \begin{pmatrix} \theta^T x^{(1)} - y^{(1)} \\ \theta^T x^{(2)} - y^{(2)} \\ \vdots \\ \theta^T x^{(m)} - y^{(m)} \end{pmatrix}^T \cdot \begin{pmatrix} \frac{1}{2}w^1 & & & \\ & \frac{1}{2}w^2 & & \\ & & \ddots & \\ & & & \frac{1}{2}w^m \end{pmatrix} \cdot \begin{pmatrix} \theta^T x^{(1)} - y^{(1)} \\ \theta^T x^{(2)} - y^{(2)} \\ \vdots \\ \theta^T x^{(m)} - y^{(m)} \end{pmatrix}$

$= (X\theta - y)^T \cdot \overset{W}{\cancel{\operatorname{diag}(\tfrac{1}{2} w^{(i)})}}\, (X\theta - y)$

$W_{ij} = \begin{cases} \frac{1}{2} w^{(i)} & i = j \\ 0 & \text{else.} \end{cases}$

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

~~$\nabla J(\theta) = 2W(X\theta - y) \cdot X^T$~~

~~$J(\theta) = (x_1\theta - y, x_2\theta - y, \ldots, x_n\theta - y)^T W($~~

$J(\theta) = \begin{pmatrix} x^{(1)T}\theta - y^{(1)} \\ x^{(2)T}\theta - y^{(2)} \\ \vdots \\ x^{(m)T}\theta - y^{(m)} \end{pmatrix}^T W \begin{pmatrix} x^{(1)T}\theta - y^{(1)} \\ x^{(2)T}\theta - y^{(2)} \\ \vdots \\ x^{(m)T}\theta - y^{(m)} \end{pmatrix}$

where W is a diagonal matrix.

hence.

~~$J(\theta) = \sum_{i=1}^{m} (x^{(i)T}\theta - y^{(i)})^T W^{(i)} (x^{(i)T}\theta - y^{(i)})$~~

~~$\nabla J(\theta) = \sum_{i=1}^{m} \nabla_\theta (x^{(i)T}\theta - y^{(i)})^T W^{(i)} (x^{(i)T}\theta - y^{(i)})$~~

~~$= \sum_{i=1}^{m} (\frac{1}{2} w^{(i)} \cdot 2I (x^{(i)T}\theta - y^{(i)})$~~

~~$\cdot \operatorname{diag}(x^{(i)}))$~~

~~$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{m} (\frac{1}{2} w^{(i)} (x_j^{(i)T}$~~

~~$\nabla J(\theta) = 2W(x\theta - y) \operatorname{diag}(x^{(i)}) \nabla_\theta (x\theta - y)$~~

~~$\nabla J(\theta) =$~~ let $v = x\theta - y$.

$\frac{\partial J}{\partial \theta_i} = \sum_{j=1}^{n} \frac{\partial J}{\partial v_j} \frac{\partial v_j}{\partial \theta_i}$

~~$\therefore \nabla J(\theta) =$~~ $\frac{\partial v_i}{\partial \theta_j} = x_j^{(i)}$

$\therefore \nabla(\theta) = X^T \cdot 2W(X\theta - y)$.

$\nabla(\theta) = 0$

$X^T \cdot 2W(X\theta - y) = 0$

$X^T \cdot W X\theta = X^T W y$

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

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

$\cancel{\frac{1}{\sqrt{2\pi}}\exp(}$

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

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

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

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

$\therefore\ b(y)=\frac{1}{\sqrt{2\pi}\,\sigma^{(i)}}\exp\left(-\frac{(y^{(i)})^2}{2(\sigma^{(i)})^2}\right)$

$\eta=\frac{\theta^T x^{(i)}}{(\sigma^{(i)})^2}$

$a(\eta)=\frac{(\theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}=\frac{(\sigma^{(i)})^2}{2}\eta^2$

$\cancel{J(\theta)=-l(\theta)}$

$\cancel{\nabla J(\theta)=X^T(a'(\theta^T x)-y)}$

$h_\theta(x)=a'(\eta)=(\sigma^{(i)})^2\eta=(\sigma^{(i)})^2\frac{\theta^T x}{(\sigma^{(i)})^2}=\theta^T x.$

$J(\theta)=-l(\theta)=-\sum_{i=1}^m\log b(y^{(i)})-\sum_{i=1}^m\eta^{(i)}y^{(i)}$

$+\sum_{i=1}^m a(\eta^{(i)}).$

$\frac{\partial}{\partial\theta_j}J(\theta)=\frac{\partial}{\partial\theta_j}\sum_{i=1}^m\left(a(\eta^{(i)})-\eta^{(i)}y^{(i)}\right)$

$=\sum_{i=1}^m\frac{\partial}{\partial\theta_j}\left(\frac{(\theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}-\frac{\theta^T x^{(i)}}{(\sigma^{(i)})^2}y^{(i)}\right).$

$=\sum_{i=1}^m\frac{\partial}{\partial\theta_j}\cdot\frac{1}{2}\cdot\frac{1}{(\sigma^{(i)})^2}\cdot\left((\theta^T x^{(i)})^2-2\theta^T x^{(i)}y^{(i)}+(y^{(i)})^2\right).$

$=\sum_{i=1}^m\frac{1}{2}\frac{1}{(\sigma^{(i)})^2}\cdot\frac{\partial}{\partial\theta_j}\left(((\theta^T x^{(i)})-y^{(i)})\cdot((\theta^T x^{(i)})-y^{(i)})\right).$

$=\frac{\partial}{\partial\theta_j}\frac{1}{2}\cdot\frac{1}{(\sigma^{(i)})^2}\cdot\sum_{i=1}^m\left(((\theta^T x^{(i)})-y^{(i)})\cdot((\theta^T x^{(i)})-y^{(i)})\right)$

let $f(\theta)=\frac{1}{2}\cdot\sum_{i=1}^m\frac{1}{(\sigma^{(i)})^2}\left((\theta^T x^{(i)}-y^{(i)})\cdot((\theta^T x^{(i)})-y^{(i)})\right)$

$\therefore\ J(\theta)-f(\theta)=-\sum_{i=1}^m\log b(y^{(i)})-\sum_{i=1}^m(y^{(i)})^2\cdot\frac{1}{2}\cdot\frac{1}{(\sigma^{(i)})^2}$

which is a const, note as C.

$J(\theta)=f(\theta)+C$

$\therefore$ if we want to minimize $J(\theta)$, we just need to minimize $f(\theta)$.

According to ii)

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

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

(b) The model seem to be underfitting.