PS-1

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

$$h_\theta(x^{(i)}) = \sigma(\theta^T x^{(i)})$$

$$\nabla_\theta J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} \frac{y^{(i)}}{\sigma(\theta^T x^{(i)})}\nabla_\theta \sigma(\theta^T x^{(i)}) + \frac{1-y^{(i)}}{1-\sigma(\theta^T x^{(i)})}\nabla_\theta(1-\sigma(\theta^T x^{(i)}))$$

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

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}(1-\sigma(\theta^T x^{(i)})) - (1-y^{(i)})\sigma(\theta^T x^{(i)})\right]x^{(i)}$$

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

$$\nabla_\theta^2 J(\theta) = \nabla_\theta(\nabla_\theta J(\theta)) = \nabla_\theta -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} - h_\theta(x^{(i)})\right]x^{(i)}$$

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

$$= \boxed{\frac{1}{m}\sum_{i=1}^{m} h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))\, x^{(i)} x^{(i)T} = H}$$

For a particular i,

$$\frac{1}{m}\sum_j \sum_k h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))\, z_j x_j^{(i)} x_k^{(i)} z_k$$

$$= \underbrace{\frac{h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))}{m}}_{k>0} \sum_j \sum_k z_j x_j^{(i)} x_k^{(i)} z_k$$

$$= k(x^T z)^2 \geq 0$$

Generalizing for all i's,

$$\boxed{z^T H z \geq 0}$$

(b) Coded

(c) $$p(y=1|x;\phi,\mu_0,\mu_1,\Sigma) = \frac{p(x|y=1)\,p(y=1)}{p(x)}$$

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

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

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

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

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

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

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

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

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

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

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

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

Hence, $\theta^T = \cancel{2}\mu_1^T\Sigma^{-1} + \mu_0^T\Sigma^{-1} = \Sigma^{-1}(\mu_1-\mu_0)$

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

Page No.:

Date:

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

$$= \log \prod_{i=1}^{m} \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x^{(i)} - \mu_{y^{(i)}})^T \Sigma^{-1} (x^{(i)} - \mu_{y^{(i)}})\right)$$

$$= \sum_{i=1}^{m} \log \frac{1}{\sqrt{2\pi}\,\sigma} - \frac{1}{2}(x^{(i)} - \mu_{y^{(i)}})^T \Sigma^{-1} (x^{(i)} - \mu_{y^{(i)}}) + \log \phi^{y^{(i)}} (1-\phi)^{1-y^{(i)}}$$

$$= \sum_{i=1}^{m} \log \frac{1}{\sqrt{2\pi}\,\sigma} + y^{(i)} \log \phi + (1-y^{(i)}) \log(1-\phi) - \frac{1}{2}(x^{(i)} - \mu_{y^{(i)}})^T \Sigma^{-1} (x^{(i)} - \mu_{y^{(i)}})$$

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

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

$$\Rightarrow \sum_i y^{(i)} - m\phi = 0 \Rightarrow \boxed{\frac{\sum_i y^{(i)}}{m} = \phi}$$

$$\frac{\partial l}{\partial \mu_0} = \frac{\partial}{\partial \mu_0} \sum_{i=1}^{m} -I(y^{(i)}=0) \cdot \frac{1}{2}(x^{(i)} - \mu_0)^T \frac{1}{\sigma} (x^{(i)} - \mu_0)$$

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

$$\Rightarrow \mu_0 \sum_i I(y^{(i)}=0) = \sum_i I(y^{(i)}=0)\, x^{(i)}$$

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

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

$$\frac{\partial l}{\partial \Sigma} = \frac{\partial}{\partial \Sigma} \sum_{i=1}^{m} \frac{-1}{2}(x^{(i)} - \mu_{y^{(i)}})^T \Sigma^{-1} (x^{(i)} - \mu_{y^{(i)}}) + \log\left(\frac{1}{\sqrt{2\pi}\,|\Sigma|^{1/2}}\right)$$

$\Rightarrow$

$$\Rightarrow \frac{\partial l}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2} \sum_{i=1}^{m} -\frac{1}{2}(x^{(i)}-\mu_{y^{(i)}})^T(\sigma^2)^{-1}(x^{(i)}-\mu_{y^{(i)}}) + \cancel{(\sigma^2)^{1/2} \cdot -2\,(\sigma^2)^{-3/2}}\ \log\left(\frac{1}{(\sigma^2)^{1/2}}\right)$$

$$= \sum_{i=1}^{m} +\frac{1}{2}(x^{(i)}-\mu_{y^{(i)}})^T(\sigma^2)^{-2}(x^{(i)}-\mu_{y^{(i)}})^T + (\sigma^2)^{1/2} \times -\frac{1}{2}\sigma^{2^{-3/2}}$$

$$= \sum_{i=1}^{m} \frac{\sigma^{-4}}{2}(x^{(i)}-\mu_{y^{(i)}})(x^{(i)}-\mu_{y^{(i)}})^T \cancel{+} - \cancel{2}\frac{\sigma^{-2}}{2} = 0$$

$$\Rightarrow \boxed{\frac{\sigma^{-2}}{\cancel{2}} = \cancel{\frac{\sigma^{-4}}{2}} \sum_{i=1}^{m} (x^{(i)}-\mu_{y^{(i)}})(x^{(i)}-\mu_{y^{(i)}})^T = \Sigma}$$

(e) Coded

(f) Coded

(g) GDA performed worse on dataset 1. The data does not have a Gaussian distribution.

(h) Transform the data to Gaussian distribution via normalization etc.

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

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

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

$\Rightarrow$ ① = ②

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

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

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

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

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

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

(c) Coded

(d) Coded

(e) Coded

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

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

$$b(y) = \frac{1}{y!},\ \eta = \log\lambda,\ T(y) = y,\ a(\eta) = \boxed{\exp(\eta) = \lambda}.$$

(b) $h_\theta(x) = E[y|x;\theta]$

$$= \lambda = \boxed{\exp(\eta)}$$

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

$\cancel{\ell(\theta) = \log \prod_{i=1}^{m} p(y^{(i)}|x^{(i)};\theta)}$

(c) $\ell(\theta) = \log p(y^{(i)}|x^{(i)};\theta)$

$$= \log b(y^{(i)})\exp(\eta^T T(y^{(i)}) - a(\theta^T x^{(i)}))$$

$$= \log \frac{1}{y!}\exp(\theta^T x^{(i)} y^{(i)} - e^{+\theta^T x^{(i)}})$$

$$= -\log y^{(i)}! + \theta^T x^{(i)} y^{(i)} - e^{\theta^T x^{(i)}}$$

$$\frac{\partial \ell(\theta)}{\partial \theta_j} = y^{(i)} x_j^{(i)} - e^{\theta^T x^{(i)}} x_j^{(i)} = (y^{(i)} - e^{\theta^T x^{(i)}})\, x_j^{(i)}$$

$$\Rightarrow \boxed{\theta_j := \theta_j + \alpha\, x_j^{(i)}(y^{(i)} - e^{\theta^T x^{(i)}})}$$

(d) Coded.

4. (a) $\int_{-\infty}^{\infty} p(y;\eta)\,dy = 1$

$\Rightarrow \frac{\partial}{\partial\eta}\int_{-\infty}^{\infty} p(y;\eta)\,dy = \int_{-\infty}^{\infty} \frac{\partial}{\partial\eta} p(y;\eta)\,dy = \frac{\partial}{\partial\eta} 1 = 0$

$\Rightarrow \int_{-\infty}^{\infty} \frac{\partial}{\partial\eta} b(y)\exp(\eta y - a(\eta))\,dy = 0$

$\Rightarrow \int_{-\infty}^{\infty} b(y)\exp(\eta y - a(\eta))\left(y - \frac{\partial a(\eta)}{\partial\eta}\right)dy = 0$

$\Rightarrow \int_{-\infty}^{\infty} y\,p(y;\eta) - \frac{\partial a(\eta)}{\partial\eta}\int_{-\infty}^{\infty} p(y;\eta)\,dy = 0$

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

(b) $\int_{-\infty}^{\infty} y\,p(y;\eta)\,dy = \frac{\partial a(\eta)}{\partial\eta}$

$\Rightarrow \int_{-\infty}^{\infty} \frac{\partial}{\partial\eta} y\,b(y)\exp(\eta y - a(\eta))\,dy = \frac{\partial^2 a(\eta)}{\partial\eta^2}$

$\Rightarrow \int_{-\infty}^{\infty} y\,b(y)\exp(\eta y - a(\eta))\left(y - \frac{\partial a(\eta)}{\partial\eta}\right)dy = \frac{\partial^2 a(\eta)}{\partial\eta^2}$

$\Rightarrow \int_{-\infty}^{\infty} y^2 p(y;\eta)\,dy - \frac{\partial a(\eta)}{\partial\eta}\int_{-\infty}^{\infty} y\,p(y;\eta)\,dy = \frac{\partial^2 a(\eta)}{\partial\eta^2}$

$\Rightarrow \int_{-\infty}^{\infty} y^2 p(y;\eta)\,dy - \mu^2 = \boxed{Var[Y|X;\theta] = \frac{\partial^2 a(\eta)}{\partial\eta^2}}$

(c) $\ell(\theta) = -\log\prod_{i=1}^{m} p(y^{(i)}|x^{(i)};\theta)$

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

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

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

$$\frac{\partial^2 l}{\partial \theta_j \partial \theta_k} = + \sum_{i=1}^{m} x_j^{(i)} \left(+ a''(\theta^T x^{(i)}) \right) x_k^{(i)} = H$$

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

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{n} \underbrace{a''(\theta^T x^{(i)})}_{\substack{\geq 0 \\ (\because \mathrm{Var})}} \underbrace{z_j^{(i)} x_j^{(i)} \, x_k^{(i)} z_k^{(i)}}_{(x^T z)^2 \geq 0}$$

Hence, H is PSD.

$\underline{5.}$ (a)(i) $J(\theta) = \frac{1}{2}\sum_{i=1}^{m} w^{(i)} (\theta^T x^{(i)} - y^{(i)})^2$

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

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

$W(i,i) = \frac{w^{(i)}}{2}$, rest 0 (diag matrix).

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

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

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

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

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

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

$= \cancel{2X^T W X\theta - X^T y - X^T W y = 0}$

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

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

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

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

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

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

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

$= -m \log \sqrt{2\pi} + \sum_{i=1}^{m} \log \sigma^{(i)} \underbrace{- \frac{1}{2}\sum_{i=1}^{m} w^{(i)} (\theta^T x^{(i)} - y^{(i)})^2}_{\text{weighted linear reg}}$

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

(b) Coded (c) Coded