1. (a)

$$\frac{\partial J(\theta)}{\partial \theta_j} = -\frac{1}{m}\frac{\partial \sum_{i=1}^{m} y^{(i)} \log g(\theta^T x^{(i)}) + (1-y^{(i)}) \log(1-g(\theta^T x^{(i)}))}{\partial \theta_j}$$

$$= -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \frac{1}{g(\theta^T x^{(i)})} \cdot g(\theta^T x)(1-g(\theta^T x)) \cdot x_j^{(i)} + (1-y^{(i)}) [g(\theta^T x)(1-g(\theta^T x)) x_j^{(i)}$$

$$\frac{\partial g(\theta^T x)}{\partial \theta_j} = g(\theta^T x)\cdot(1-g(\theta^T x))\cdot x_j^{(i)}$$

$$= -\frac{1}{m}\sum_{i=1}^{m} y^{(i)}(1-g(\theta^T x^{(i)}))\cdot x_j^{(i)} + (y^{(i)}-1)\, g(\theta^T x^{(i)})\, x_j^{(i)}$$

$$= -\frac{1}{m}\sum_{i=1}^{m} x_j^{(i)} y^{(i)} - x_j^{(i)} y^{(i)} g(\theta^T x^{(i)}) - x_j^{(i)} g(\theta^T x^{(i)}) + x_j^{(i)} y^{(i)} g(\theta^T x^{(i)})$$

$$= \frac{1}{m}\sum_{i=1}^{m} (g(\theta^T x^{(i)}) - y^{(i)})(x_j^{(i)}) = \varphi(\theta) \qquad \nabla_\theta J(\theta) = \frac{1}{m} X^T (g(X\theta) - Y)$$

$$\therefore H_{jk} = \frac{\partial \varphi(\theta_j)}{\partial \theta_k} = \frac{1}{m}\sum_{i=1}^{m} x_j^{(i)} x_k^{(i)} g(\theta^T x^{(i)})(1-g(\theta^T x^{(i)}))$$

$$\therefore H = \frac{1}{m} X^T W X, \quad (W_{ii} = g(\theta^T x^{(i)})(1-g(\theta^T x^{(i)}))$$

Rest elements $=0$

$$= \frac{1}{m} X^T X W$$

$$\therefore z^T H z = z^T X^T X W z = z^T X^T X z W$$

$\because W_{ii} \geq 0 \qquad \therefore z^T H z \geq 0$

(b) coding problem

(c)

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

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

$$= \frac{1}{1+\exp[-\frac{1}{2}(x-\mu_0)^T \Sigma^{-1}(x-\mu_0) + \frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) + \ln\frac{1-\varphi}{\varphi}}$$

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

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

$$= \frac{1}{1+\exp((\mu_0-\mu_1)^T\Sigma^{-1} + \frac{1}{2}(\mu_1^T\Sigma^{-1}\mu_1 - \mu_0^T\Sigma^{-1}\mu_0) + \ln\frac{1-\varphi}{\varphi}}$$

$\therefore \theta = \Sigma^{-1}(\mu_1 - \mu_0)$

$\theta_0 = \frac{1}{2}(\mu_0^T \Sigma^{-1} \mu_0 - \mu_1^T \Sigma^{-1} \mu_1) + \ln \frac{\varphi}{1-\varphi}$

(d) $p(y=0) = 1-\varphi$ , $p(y=1) = \varphi$

$$\ell = \sum_{i=1}^{m} \ln(p(x^{(i)} | y^{(i)}) + \ln p(y^{(i)}))$$

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

$$\therefore \frac{\partial \ell}{\partial \varphi} = \frac{\partial \sum_{i=1}^{m} \ln(\varphi^{1\{y^{(i)}=1\}} + (1-\varphi)^{1-1\{y^{(i)}=1\}})}{\partial \varphi}$$

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

$$\therefore \varphi = \sum_{i=1}^{m} 1\{y^{(i)}=1\}$$

$$\frac{\partial \ell}{\partial \mu_0} = 0 \Rightarrow \frac{\partial \sum_{i=1}^{m} -(x^{(i)} - \mu^{(i)})^2}{\partial \mu_0} = 0 \Rightarrow \frac{\partial \sum_{i=1}^{m} 1\{y^{(i)}=0\}(-(x^{(i)}-\mu_0)^2) + (1-\{y^{(i)}=0\})(-(x^{(i)}-\mu_1)^2)}{\partial \mu_0}$$

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

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

For the same reason

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

$$\frac{\partial \ell}{\partial \Sigma} = 0 \Rightarrow \frac{\partial \sum_{i=1}^{m} \ln\frac{1}{\sqrt{2\pi\Sigma}} - \frac{(x^{(i)}-\mu^{(i)})^2}{2\Sigma}}{\partial \Sigma} = 0$$

$$\sum_{i=1}^{m} -\frac{1}{\sqrt{\Sigma}} + \frac{(x^{(i)}-\mu^{(i)})^2}{\Sigma^{\frac{3}{2}}} \Rightarrow \Sigma = \frac{1}{m}\sum_{i=1}^{m} (x^{(i)} - \mu^{(i)})^2 = \sigma^2$$

2) (a)

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

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

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

$$\therefore \alpha = P(y^{(i)}=1 \mid t^{(i)}=1)$$

(b)

$\because \alpha = \dfrac{P(y^{(i)}=1 \mid x^{(i)})}{P(t^{(i)}=1 \mid x^{(i)})}$ . when $x^{(i)} \in V^+$, $P(t^{(i)}=1, x^{(i)}) \approx 1$

$\therefore P(y^{(i)}=1 \mid x^{(i)}) \approx \alpha$ when $x^{(i)} \in V^+$

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

3. (a)

$$P(y;\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}$$

$$P(y;\eta) = b(y)\exp(\eta^T T(y) - a(\eta))$$

$$\begin{cases} b(y) = \dfrac{1}{y!} \\ \eta = \ln\lambda \\ T(y) = y \\ a(\eta) = e^{\eta} \end{cases}$$

(b)

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

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

$$\therefore g(\eta) = E(T(y) \mid x;\eta) = \lambda = e^{\theta^T x}$$

(c)

$$l(\theta) = -\lambda + y\ln\lambda - \sum_{i=1}^{y} \ln y = -e^{\theta^T x^{(i)}} + y\theta^T x^{(i)} - \sum_{i=1}^{y} \ln y$$

$$\therefore \frac{\partial l(\theta)}{\partial \theta_j} = -x_j^{(i)} e^{\theta^T x^{(i)}} + y_j x_j^{(i)}$$

$$m\times 1$$

$$(y - e^{(}y - e^{\theta^T x^{(i)}})$$

4. (a) $\int p(y;\eta)\, dy = 1$ ; $\frac{\partial 1}{\partial \eta} = 0$

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

$$0 = E(Y) - a'(\eta) \Rightarrow E(Y) = a'(\eta)$$

(b)

$$\frac{dE(\eta)}{d\eta} = \frac{d}{d\eta}\left(\int y\, p(y;\eta)\ dy\right) = \int y(y - a'(\eta))p(y;\eta)\, dy$$

$$\therefore\ a''(\eta) = E(Y^2) - [E(Y)]^2 = Var(Y|X;\theta)$$

(c)

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

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

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

$$\therefore \nabla_\theta l(\theta) = -X^T(y - a'(X\theta))$$

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

$H = X^T \Sigma X$ $\Sigma$ is a diagonal matrix $m \times m$

$$\Sigma_{ii} = a''(\theta^T x^{(i)}) = Var(y^{(i)}) \geq 0$$

$\therefore H$ PSD

5 (i)

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

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

$\therefore X^T W X \theta = X^T W Y$

$\therefore \theta = (X^T W X)^{-1} X^T W Y$

(iii)

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

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