HW ①-ⓐ.

$$H_{ij} = \frac{\partial^2 J(\theta)}{\partial\theta_i \partial\theta_j} \quad J(\theta) = \frac{\partial J(\theta)}{\partial\theta_j} \quad -\frac{1}{m}\sum_{k=1}^{m}\left(y^{(k)}\frac{1}{g(z)} + (1-y^{(k)})\frac{1}{1-g(z)}(-1)\right)g(z)(1-g(z))\,x_i^k$$

$$= \frac{\partial J(\theta)}{\partial\theta_j} \quad -\frac{1}{m}\sum_{k=1}^{m} x_i^k\left(y^{(k)}(1-g(z)) - (1-y^{(k)})\,g(z)\right)$$

$$= \frac{\partial J(\theta)}{\partial\theta_j} - \frac{1}{m}\sum_{k=1}^{m} x_i^k\, y^{(k)} - x_i^k\, g(z)$$

$$= -\frac{1}{m}\sum_{k=1}^{m} - x_i^k x_j^k\, g(z)\cdot(1-g(z)) = \frac{1}{m}\sum_{k=1}^{m} x_i^k x_j^k\, g(z)\cdot(1-g(z))$$

Since $g(z) > 0$ $\quad (1-g(z)) > 0 \quad \frac{1}{m} > 0.$

The sign of H_{ij} is determined by x_i^k, and x_j^k.

$$z^T H z = \sum_{i=1}\sum_{j=1} z_i \sum_{k=1}^{m}\frac{1}{m} x_i^k x_j^k\, g(z)\cdot(1-g(z))\, z_j = \frac{1}{m}\sum_{k=1}^{m}\sum_{i=1}\sum_{j=1} z_i x_i^k x_j^k z_j\,(g(z)\cdot(1-g(z)))$$

$$= \frac{1}{m}\sum_{k=1}^{m} g(z)(1-g(z)) \sum_{i=1}\sum_{j=1} z_i x_i^k x_j^k z_j$$

$$= \frac{1}{m}\sum_{k=1}^{m} g(z)(1-g(z)) \sum_{i=1} z_i x_i^k \sum_{j=1} x_j^k z_j$$

$$= \frac{1}{m}\sum_{k=1}^{m} g(z)(1-g(z)) \left((x^k)^T z\right)^2$$

since $(x^{kT} z)^2 \geq 0$. $\quad z^T H z \geq 0$. $\quad$ so, H is positive semidefinite.

1-c.

$$P(y=1 \mid x; \phi, \mu_0, \mu_1, \Sigma) = \frac{P(x \mid y=1)\, P(y=1)}{P(x \mid y=1)\, P(y=1) + P(x \mid y=0)\, P(y=0)}$$

$$= \frac{1}{1 + \frac{P(x \mid y=0)\, P(y=0)}{P(x \mid y=1)\, P(y=1)}} \qquad \text{Let } \lambda = \frac{P(x \mid y=0)\, P(y=0)}{P(x \mid y=1)\, P(y=1)}.$$

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

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

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

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

Since Σ is symmetric, Σ^{-1} is symmetric.

$$\text{Then } \lambda = \exp\left(\cancel{\tfrac{1}{2}} (\mu_0 - \mu_1)^T \Sigma^{-1} x + \frac{1}{2} (\mu_1^T \Sigma^{-1} \mu_1 - \mu_0^T \Sigma^{-1} \mu_0) \right) \cdot \frac{1-\phi}{\phi}$$

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

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

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

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

HW. ① - ⓓ.

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

ϕ. $\ell(\phi) = \log \prod_{i=1}^{m} P(x^{(i)}|y^{(i)}) \cdot P(y^{(i)})$

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

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

$$\frac{\sum y^{(i)}}{\phi} = \frac{\sum 1-y^{(i)}}{1-\phi}$$

$$\sum y^{(i)} \cancel{- \phi \sum y^{(i)}} = \phi m \cancel{- \phi \sum y^{(i)}}$$

$\boxed{\text{second derivative test is omitted.}}$

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

μ_0: $\ell(\mu_0) = \log \prod_{i=1}^{m} P(x^{(i)}|y^{(i)}) P(y^{(i)})$

μ_1 same idea

$$= \sum_{i=1}^{m} \log(P(x^{(i)}|y^{(i)})) + \log(P(y^{(i)})).$$

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

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

$$\frac{\partial \ell}{\partial \mu_0} = \sum_{i=1}^{m} x - \mu_0 = 0.$$

for $y=0$

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

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

HW 1 - d continue page 2.

$$\Sigma : \ l(\Sigma) = \log \prod_{i=1}^{m} P(x^{(i)} \mid y^{(i)})\, P(y^{(i)})$$

$$= \sum_{i=1}^{m} \log \left(\frac{1}{(2\pi)^{n/2}\, \Sigma^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_y)^2 \Sigma^{-1}\right) + \log(P(y^{(i)})\right).$$

$$\frac{\partial l(\Sigma)}{\partial \Sigma} = \sum_{i=1}^{m} \frac{\partial}{\partial \Sigma} \log\left(\frac{1}{(2\pi)^{n/2}\, \Sigma^{1/2}}\right) + \frac{\partial}{\partial \Sigma}\left(-\frac{1}{2}(x-\mu_y)^2 \Sigma^{-1}\right)$$

$$= \sum_{i=1}^{m} \cancel{(2\pi)^{n/2}}\, \cancel{\Sigma^{1/2}} \cdot (-1) \cdot \frac{1}{\cancel{(2\pi)^{n}}\, \Sigma} \cdot \cancel{(2\pi)^{n/2}} \cdot \frac{1}{2}\, \cancel{\Sigma^{-1/2}} + \left(\frac{1}{2}(x-\mu_y)^2 \Sigma^{-2}\right).$$

$$= \sum_{i=1}^{m} -\frac{1}{\cancel{2}\Sigma} + \frac{1}{\cancel{2}}(x-\mu_y)^2 \Sigma^{-2} = 0.$$

$$\Sigma^{-2} \sum_{i=1}^{m} (x-\mu_y)^2 = \frac{m}{\Sigma}$$

$$\sum_{i=1}^{m} (x-\mu_y)^2 = m\Sigma$$

$$\Sigma = \frac{1}{m} \sum_{i=1}^{m} (x-\mu_y)^2.$$

HW 2 - a.

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

because $P(y^{(i)}=1) + P(y^{(i)}=0) = 1$.

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

$$P(t^{(i)}=1 \mid x^{(i)}) = \frac{P(t^{(i)}=1 \mid y^{(i)}=1, x^{(i)}) \cdot P(y^{(i)}=1, x^{i})}{P(x^{(i)})} + \frac{P(y^{(i)}=0 \mid t^{(i)}=1, x^{(i)}) \cdot P(t=1, x_i)}{P(x^{(i)})}$$

since $P(t^{(i)}=1 \mid y^{(i)}=1, x^{(i)}) = 1$, $P(y^{(i)}=0 \mid t^{(i)}=1, x^{(i)}) = P(y^{(i)}=0 \mid t^{(i)}=1, x^{(i)}) = P(y^{(i)}=0 \mid t=$

$$= \frac{P(y^{(i)}=1, x^{i})}{P(x^{(i)})} + P(y^{(i)}=0 \mid t^{(i)}=1) \cdot \frac{P(t^{(i)}=1, x_i)}{P(x^{(i)})}$$

$$= P(y^{(i)}=1 \mid x^{(i)}) + P(y^{(i)}=0 \mid t^{(i)}=1) \cdot P(t^{(i)}=1 \mid x^{(i)})$$

$$P(t^{(i)}=1 \mid x^{(i)}) - P(y^{(i)}=0 \mid t^{(i)}=1) \cdot P(t^{(i)}=1 \mid x^{(i)}) = P(y^{(i)}=1 \mid x^{(i)}).$$

$$P(t^{(i)}=1 \mid x^{(i)}) \cdot (1 - P(y^{(i)}=0 \mid t^{(i)}=1)) = P(y^{(i)}=1 \mid x^{(i)})$$

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

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

---

2 - b.

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

$$\approx P(t^{(i)}=1 \mid x^{(i)}) \cdot \alpha$$

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

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

HW ③ - ⓐ.

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

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

$b(y) = \frac{1}{y!}$ $\quad \eta = \log(\lambda)$ $\quad a(\eta) = e^\eta$

$T(y) = y$ $\quad \lambda = e^\eta$

ⓑ. $h_\theta(x) = E[y|x;\theta]$

$$= \lambda$$

$$= e^\eta$$

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

ⓒ. ~~$L(\theta) = p(\vec{y}|X;\theta)$~~

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

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

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

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

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

$$= x_j(y - h_\theta(x))$$

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

HW 4 a.

we know $\int_{-\infty}^{\infty} p(y;\eta)\,dy = 1$

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

$$\int \frac{\partial}{\partial \eta} b(y)\exp(\eta y - a(\eta))\,dy = 0$$

$$\int b(y)\exp(\eta y - a(\eta))\cdot\left(y - \frac{\partial}{\partial \eta} a(\eta)\right) dy = 0$$

$$\int y\, b(y)\exp(\eta y - a(\eta))\,dy - \int b(y)\exp(\eta y - a(\eta))\cdot \frac{\partial}{\partial \eta} a(\eta)\,dy = 0.$$

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

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

4-b. we know $\int y\, p(y;\eta)\,dy = \frac{\partial}{\partial \eta} a(\eta)$ from 4a.

$$\frac{\partial}{\partial \eta}\int y\, p(y;\eta)\,dy = \frac{\partial^2}{\partial \eta^2} a(\eta)$$

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

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

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

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

HW 4-c.

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

$$= -\sum_{k=1}^{m} \log\left(b(y)\exp(\theta^T x y - a(\theta^T x))\right)$$

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

$$H_{ij} = \frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j} = \frac{\partial}{\partial\theta_j}\left(-\sum_{k=1}^{m} x_i^{(k)} y^{(k)} - \frac{\partial}{\partial\eta} a(\eta)\cdot x_i^{(k)}\right)$$

$$= -\sum_{k=1}^{m} - \frac{\partial^2}{\partial\eta^2} a(\eta)\cdot x_i^k x_j^k$$

$$= \sum_{k=1}^{m} \frac{\partial^2}{\partial\eta^2} a(\eta)\, x_i^k x_j^k$$

Let $z \in \mathbb{R}^n$, then $z^T H z = \sum_{i=1}^{n}\sum_{j=1}^{n} z_i H_{ij} z_j$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} z_i \sum_{k=1}^{m} \frac{\partial^2}{\partial\eta^2} a(\eta)\, x_i^k x_j^k \cdot z_j$$

$$= \sum_{k=1}^{m} \frac{\partial^2}{\partial\eta^2} a(\eta) \sum_{i=1}^{n} z_i x_i^k \sum_{j=1}^{n} x_j^k z_j$$

$$= \sum_{k=1}^{m} \frac{\partial^2}{\partial\eta^2} a(\eta) \left(z^T x^k\right)^2$$

since $Var[Y|X;\theta] \geq 0$. the $\frac{\partial^2}{\partial\eta^2} a(\eta) \geq 0$.

thus $z^T H z \geq 0$. therefore H is PSD.

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

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

$$= \sum_{i=1}^{m}\sum_{j=1}^{m} (\theta^T x^{(i)} - y^{(i)})^T (\tfrac{1}{2} w^{(i)}) (\theta^T x^{(j)} - y^{(j)}) \text{ when } i=j. +$$

$$\sum_{i=1}^{m}\sum_{j=1}^{m} (\theta^T x^{(i)} - y^{(i)})\ 0\ (\theta x^{(j)} - y^{(j)}) \text{ when } i \neq j$$

Hence when $i \neq j$ $W_{ij} = 0$. when $i = j$. $W_{ij} = \frac{1}{2} w^{(i)}$

$$= \sum_{i=1}^{m}\sum_{j=1}^{m} (\theta^T x^{i} - y^{(i)})^T W_{ij} (\theta^T x^{j} - y^{(j)})$$

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

5-b. $\nabla_\theta J(\theta) = \nabla_\theta (X\theta - y)^T W (X\theta - y)$

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

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

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

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

$\cancel{\theta = X^{-1} W^{-1} X^{-T} X^T W y}$

$\boxed{\cancel{\theta = X^{-1} y}}$

HW 5. (a). i.

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

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

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

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

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

ii

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

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

$$2X^T W X\theta = X^T W y + X^T W^T y$$

$$X^T W X\theta = \frac{1}{2}(X^T W y + X^T W^T y)$$

$$\theta = \frac{1}{2}(X^T W X)^{-1}(X^T W y + X^T W^T y)$$

$$= \frac{1}{2}(X^{-1} W^{-1} (X^T)^{-1})(X^T W y + X^T W^T y).$$

$$= \frac{1}{2}(X^{-1} y + X^{-1} W^{-1} W^T y)$$

iii

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

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

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

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

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

since σ is constant and known, $m\log\left(\frac{1}{\sqrt{2\pi}}\right) + \sum_{i=1}^{m} \log\left(\frac{1}{\sigma^i}\right)$ is constant.

To maximize $\ell(\theta)$ is to minimize $\sum_{i=1}^{m} \frac{1}{2(\sigma^i)^2}\cdot(y^{(i)} - \theta^T x^i)^2$

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

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