[PS 1]

#1.

(a) $$\frac{\partial J}{\partial \theta_k} = \frac{1}{m}\sum \frac{e^{-y^i\theta^T x^i}(-y^i x_k^i)}{1+e^{-y^i\theta^T x^i}}$$

$$= -\frac{1}{m}\sum \frac{1}{1+e^{y^i\theta^T x^i}}(-y^i x_k^i) = -\frac{1}{m}\sum h_\theta(-y^i x^i)(-y^i x_k^i)$$

$$\frac{\partial^2 J}{\partial\theta_l\partial\theta_k} = \frac{1}{m}\sum \frac{\partial}{\partial\theta_l} h_\theta(-y^i x^i)(y^i x_k^i)$$

$$= -\frac{1}{m}\sum h_\theta(-y^i x^i)(1-h_\theta(-y^i x^i))(y^i x_k^i)(y^i x_l^i)$$

$$H_{kl} = -\frac{1}{m}\sum h_\theta(x^i)(1-h_\theta(x^i))x_k^i x_l^i$$

$$\therefore H = \frac{1}{m}\sum h_\theta(x^i)(1-h_\theta(x^i))x^i x^{(i)T}$$

$$\left(\begin{array}{l} h_\theta(x^i)(1-h_\theta(x^i)) \\ = h_\theta(-x^i)(1-h_\theta(-x^i)) \end{array} \right)$$

$z^T H z$

$$= z^T\left(\frac{1}{m}\sum h_\theta(x^i)(1-h_\theta(x^i))x^i x^{(i)T}\right)z$$

$$= \frac{1}{m}\sum h_\theta(x^i)(1-h_\theta(x^i))(z^T x^i)^2 \geq 0$$

(b),(c) see ps1-1 bc.ipynb

#2.

(a) $$\frac{e^{-\lambda}\lambda^y}{y!} \leftrightarrow b(y)\exp(\eta^T T(y) - a(\eta))$$

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

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

(b) $$E[T(y);\eta] = E[y;\eta] = \lambda = \boxed{e^\eta}$$

(c) $$p(y^i|x^i;\theta) = \frac{1}{y^i!}\exp(\eta^T y^i - e^\eta)$$

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

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

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

$$\therefore \theta_j \leftarrow \theta_j + \alpha(y^i - e^{\theta^T x^i})x_j^i$$

(d)

$$\ell(\theta) = \log p(y \mid x ; \theta)$$

$$= \log \left[b(y) \exp(\eta^T T(y) - a(\eta)) \right]$$

$$= \log b(y) + x^T \theta y - a(\eta)$$

$$\frac{\partial \ell(\theta)}{\partial \theta_j} = x_j y - \frac{\partial \eta}{\partial \theta_j} \frac{\partial a(\eta)}{\partial \eta}$$

$$= \left(y - \frac{\partial a(\eta)}{\partial \eta}\right) x_j$$

$$\therefore \theta_j \leftarrow \theta_j - \alpha \left(\frac{\partial a(\eta)}{\partial \eta} - y \right) x_j$$

Let's show $\frac{\partial a(\eta)}{\partial \eta} = h_\theta(x)$

$$1 = \int_y b(y) \exp(\eta^T y - a(\eta))\, dy$$

$$\exp(a(\eta)) = \int_y b(y) \exp(\eta^T y)\, dy$$

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

$$\frac{\partial a(\eta)}{\partial \eta} = \int_y y\, \underline{b(y) \exp(\eta^T y - a(\eta))}\, dy$$

$$= \int_y y\, p(y \mid x ; \theta)\, dy$$

$$= E[y \mid x ; \theta]$$

$$= h_\theta(x)$$

$$\therefore \theta_j \leftarrow \theta_j - \alpha (h_\theta(x) - y) x_j$$

#3.

(a)

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

$$= \frac{p(x \mid y=1)\, p(y=1)}{p(x \mid y=1)\, p(y=1) + p(x \mid y=-1)\, p(y=-1)}$$

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

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

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

$y=1$, let $\begin{cases} \theta_0 = \frac{1}{2}(\mu_{-1}^T\Sigma^{-1}\mu_{-1} - \mu_1^T\Sigma^{-1}\mu_1) - \log\frac{1-\phi}{\phi} \\ \theta = \Sigma^{-1}(\mu_1 - \mu_{-1}) \end{cases}$

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

$P(y=-1|x)$ can be derived by same way.

~~(b)~~

(c) $l(\phi, \mu_{-1}, \mu_1, \Sigma) = \log \prod p(x^i | y^i; \phi, \mu_{-1}, \mu_1, \Sigma)\, p(y^i; \phi)$

$$= \Sigma\left[\left(\log\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\right) - \frac{1}{2}(x^i - \mu_{y^i})^T\Sigma^{-1}(x^i - \mu_{y^i})\right] + \Sigma \log \phi^{y^i} + \log(1-\phi)^{1-y^i}$$

$$\nabla_\phi l(\phi, \mu_1, \mu_1, \Sigma) = \Sigma\ \frac{y^i}{\phi} - \frac{1-y^i}{1-\phi}$$

$$= \frac{\Sigma 1\{y^i=1\}}{\phi} - \frac{m - \Sigma 1\{y^i=1\}}{1-\phi} = 0$$

$$\therefore \phi = \frac{1}{m}\Sigma 1\{y^i=1\}$$

$$\nabla_{\mu_{-1}} l(\phi, \mu_{-1}, \mu_1, \Sigma) = -\cancel{\tfrac{1}{2}}\Sigma\ 1\{y^i=-1\}\left(-\cancel{2}\cancel{\Sigma^{-1}}x^i + \cancel{2}\cancel{\Sigma^{-1}}\mu_{-1}\right) = 0$$

$$\therefore \mu_{-1} = \frac{\Sigma 1\{y^i=-1\}x^i}{\Sigma 1\{y^i=-1\}} \quad \left(\text{similarly, } \mu_1 = \frac{\Sigma 1\{y^i=1\}x^i}{\Sigma 1\{y^i=1\}}\right)$$

(let $S=\Sigma^{-1}$) $\nabla_S l(\phi, \mu_{-1}, \mu_1, \Sigma) = -\frac{1}{2}\Sigma\left[-S^{-1} + (x^i-\mu_{y^i})(x^i-\mu_{y^i})^T\right] = 0$

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

#4

$$\theta - \frac{f(\theta)}{f'(\theta)}$$

(a) $\nabla_z g(z) = \begin{bmatrix} \frac{\partial g(z)}{\partial z_1} \\ \vdots \\ \frac{\partial g(z)}{\partial z_n} \end{bmatrix} = \begin{bmatrix} \left(\frac{\partial (Az)}{\partial z_1}\right)^T \frac{\partial f(Az)}{\partial (Az)} \\ \vdots \\ \left(\frac{\partial (Az)}{\partial z_n}\right)^T \frac{\partial f(Az)}{\partial (Az)} \end{bmatrix} = A^T\frac{\partial f(Az)}{\partial (Az)} = A^T \nabla_x f(x)$

$$\left(Az = \begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix}\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = a_1 z_1 + \cdots + a_n z_n\right)$$

$$\nabla_z^2 g(z) = \begin{bmatrix} \frac{\partial^2 g(z)}{\partial z_1^2} & \cdots & \frac{\partial g(z)}{\partial z_1 \partial z_n} \\ \vdots & & \vdots \\ \frac{\partial g(z)}{\partial z_n \partial z_1} & \cdots & \frac{\partial g(z)}{\partial z_n^2} \end{bmatrix} = \begin{bmatrix} \left(\frac{\partial (Az)}{\partial z_1}\right)^T \frac{\partial^2 f(Az)}{\partial (Az)^2} \frac{\partial (Az)}{\partial z_1} & \cdots & \left(\frac{\partial (Az)}{\partial z_1}\right)^T \frac{\partial f(Az)}{\partial (Az)^2} \frac{\partial (Az)}{\partial z_n} \\ \vdots & & \vdots \\ \left(\frac{\partial (Az)}{\partial z_n}\right)^T \frac{\partial^2 f(Az)}{\partial (Az)^2} \frac{\partial (Az)}{\partial z_1} & \cdots & \left(\frac{\partial (Az)}{\partial z_n}\right)^T \frac{\partial^2 f(Az)}{\partial (Az)^2} \frac{\partial (Az)}{\partial z_n} \end{bmatrix}$$

$$= A^T \left[\nabla_x^2 f(x)\right] A$$

let's assume $z^{(i)} = A^{-1} x^{(i)}$

$$z^{(i+1)} = z^{(i)} - H_g(z^{(i)})^{-1} \nabla_z g(z)$$

$$= z^{(i)} - (A^T H_f(x^{(i)}) A)^{-1} A^T \nabla_x f(x^{(i)})$$

$$= z^{(i)} - A^{-1} H_f(x^{(i)})^{-1} \nabla_x f(x^{(i)})$$

$$\underline{Az^{(i+1)}} = \underbrace{Az^{(i)}}_{x^{(i)}} - H_f(x^{(i)})^{-1} \nabla_x f(x^{(i)}) = \underline{x^{(i+1)}} \quad \therefore \sim$$

(b) grad. descent : let's assume $z^{(i)} = A^{-1} x^{(i)}$

$$Az^{(i+1)} = A\left(z^{(i)} - \alpha A^T \nabla_x f(x^{(i)})\right) = x^{(i)} - \alpha AA^T \nabla_x f(x^{(i)})$$

$\therefore$ invariant to lin. reparam. only if $AA^T = I$

#5.

(a) i) W is diagonal matrix, with $W_{ii} = \frac{1}{2} w^{(i)}$

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

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

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

iii) solving mle of θ; $\arg\max_\theta L(\theta) = \arg\max_\theta \prod p(x^i | y^i; \theta)$

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

log $\rightarrow$ $= \arg\max_\theta \left(-\frac{1}{2} \sum \frac{(y^i - \theta^T x^i)^2}{(\sigma^i)^2}\right) = \arg\min_\theta \frac{1}{2} \sum \underbrace{\boxed{\frac{1}{(\sigma^i)^2}}}_{w^i} (y^i - \theta^T x^i)^2$; solving w.l.r

(b), (c) see ps1-5bc.ipynb