Q1 (a) $\frac{\partial h_\theta(x^{(i)})}{\partial \theta_k} = \frac{\partial g(\theta^T x^{(i)})}{\partial \theta_k}$ $\quad\because g'(z) = g(z)(1-g(z))$

$\hookrightarrow \quad x_k^{(i)} g(\theta^T x)(1-g(\theta^T x))$

$\frac{\partial}{\partial \theta_k} \log(h_\theta(x^{(i)})) = \cancel{\log} \qquad \log(1-h_\theta(x^{(i)}))$

$= \frac{1}{h_\theta(x^{(i)})} \cdot \frac{\partial h_\theta(x^{(i)})}{\partial \theta_k} = x_k^{(i)}(1-g(\theta^T x^{(i)})) \qquad -x_k^{(i)}(h_\theta(x^{(i)}))$

$\therefore \nabla_k J(\theta) = \frac{-1}{n} \sum_{i=1}^{n} y^{(i)} \quad + (1-y^{(i)})$

$= \frac{-1}{n} \sum_{i=1}^{n} (y^{(i)} - h^{(i)}) x_k^{(i)}$

Hessian $= \nabla_\theta^2 J(\theta) = \frac{\sum_{i=1}^{n} h(1-h)\, x_l^{(i)} x_c^{(i)}}{n}$

$\because$ outer prod ≥ 0

$\frac{1}{n} \sum h(1-h)\, x^{(i)} x^{(i)T}$

$z^T H z \geq 0$

$\therefore \frac{1}{n}(\cdot)(z^T x)(x^T z) \qquad 0 \leq h \leq 1 \quad \because$ logistic

$\therefore 1-h \in [0,1]$

$\longrightarrow \underbrace{(z^T x)^2}_{\geq 0} \qquad \therefore$ PSD

(c) $p(y=1 \mid x) = \dfrac{p(x \mid y=1)\, p(y=1)}{p(x)}$ $\left(p(x) = p(x \mid y=0)p(y=0) + p(x \mid y=1)p(y=1) \right)$

$$= \frac{\exp\left(-\frac{1}{2}(x-\mu_1)^T \Sigma^{-1}(x-\mu_1)\right)\phi}{\text{''}\ \phi + \text{---}\ \mu_0 - (1-\phi)}$$

$$= \frac{a\phi}{a\phi + b(1-\phi)} = \frac{1}{1 + \frac{b(1-\phi)}{a\phi}}$$

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

$$e^{\left(\text{---}\ \cancel{N}-D^{e} + \log\frac{1-\phi}{\phi}\right)}$$

$$N^{e} - D^{e}$$

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

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

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

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

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

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

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

$\begin{bmatrix} 1 \\ m \\ 1 \end{bmatrix}$ $[\leftarrow m \rightarrow]$

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

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

(i) $\frac{\partial \ell}{\partial \phi} = 0 = \sum_{i=1}^{n} \frac{1}{p(y^{(i)}; \phi)} \frac{\partial p(y^{(i)})}{\partial \phi}$

$y\phi + (1-y)(1-\phi)$

$y \cancel{+} (1+y)$

$y\cancel{\phi} + (1-y)(-1)$

$y \quad -1 + y$

$2y - 1$

$0 = \sum_{i=1}^{n} \frac{2y-1}{y(\phi) + (1-y)(1-\phi)}$

$\frac{y\phi + 1 - \phi - y + y\phi}{2y\phi - \phi + 1 - y}$

$\sum_{i=1}^{n} \frac{(2y^{(i)}-1)}{\phi(2y^{(i)}-1) + (1-y^{(i)})}$

$\frac{(y) + (y-1)}{-(y-1)}$

$\phi(y-1) + \phi(y)$

$y(2\phi - 1) - 1(\phi - 1)$ $(y-1)(\phi-1) + \phi(y)$

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

if $y = 1$, $\frac{1}{\phi}$

if $y = 0$, $\frac{1}{(1-\phi)}$

$= \sum_{i=1}^{n} \frac{1}{\quad}$

Let x be how many $y = 1$

$\therefore$ if $y = 1$, $\frac{1}{\phi}$

$y = 0$ $\frac{1}{\phi - 1}$

$\therefore 0 = \frac{x}{\phi} + \frac{(n-x)}{(\phi - 1)}$

$\therefore \phi = \frac{x}{n}$

$= \frac{1}{n} \sum_{i=1}^{n} 1\{y = 1\}$

$\frac{\partial \ell}{\partial \mu} = 0$ $\quad x^{(i)} | y^{(i)}$ $\quad$ Let x be how many $y = 1$

(d) ii, iii

$$p(x^{(i)}, y^{(i)}) = \frac{1}{\sim} \exp\left(\frac{-1}{2}\overbrace{(x^{(i)} - ((1-y)\mu_0 + yu_1))}^{①}{}^{T} \Sigma^{-1} (x - ((1-y)u_0 + yu_1))\right)$$

$$\frac{\partial l}{\partial \mu_0} = \sum \log p(x+y)$$

$$= \sum \cancel{\log\left(\frac{1}{\sim}\right)} + \left(\frac{-1}{2}(\cdots)\right) - 0 \qquad \frac{\partial a^T A a}{\partial a} = a^T(A + A^T)$$

$$\therefore \frac{\partial l}{\partial ①} \cdot \frac{\partial ①}{\partial \mu_0} = \sum -\frac{1}{2} \underset{1 \times n}{①^T} \, 2\underset{n\times n}{\Sigma^{-1}} \cdot (-1 + y^{(i)})\underset{n \times n}{(-I)} = 0$$

$$\sum ①^T \Sigma^{-1} (-1 + y^{(i)})(I) = 0$$

if $y^{(i)} = 1$ $\quad = 0$

if $y^{(i)} = 0$, $-I \times (-I) = 0 \times (-I)$

$$\therefore \sum_{\substack{i=1\\ y=0}}^{n} x^{(i)} - (1 - \underset{=0}{y^{(i)}})\mu_0 + \underset{=0}{y^{(i)}}u_1))^T \underbrace{\Sigma^{-1}}(I) = 0$$

$$\Sigma^{-1} \times \Sigma = 0 \times \Sigma$$
$$= I$$

$= \Sigma$ row $\times I = \Sigma$ colum

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

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

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

(iv) $\frac{\partial l}{\partial \Sigma_1} = 0$

$$\frac{\partial\, a^T X^{-1} b}{\partial X} = -X^{-T} a b^T X^{-T}$$

$$-\frac{1}{2}\sum_{i=1}^{n} -\Sigma^{-T} ⓘⓘ^T \Sigma^{-T} = 0$$

$$\Sigma = \Sigma^T \quad \therefore (\Sigma^{-1})^T = (\Sigma^T)^{-1} = \Sigma^{-1}$$

$$\therefore \cancel{-\frac{1}{2}}\sum_{i=1}^{n} -\Sigma^{-1} ⓘⓘ^T \Sigma^{-1} = 0$$

$$\frac{\partial}{\partial}\left(\log \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\right)$$

$$= \frac{\partial}{\partial \Sigma}\left(\cancel{\log 1}^{\,0} - \cancel{\log (2\pi)^{d/2}}^{\,0} - \frac{1}{2}\log|\Sigma|\right)$$

$$\cancel{-\frac{1}{2}} \frac{1}{|\Sigma|}|\Sigma|\Sigma^{-1} +$$

$$\frac{\partial \det|A|}{\partial A} = |A|(A^{-1})^T$$

$$\therefore n\Sigma^{-1} = \sum_{i=1}^{n} \Sigma^{-1} ⓘⓘ^T \Sigma^{-1}$$

$$\times\Sigma \qquad \times\Sigma$$

$$I \qquad \Sigma\times$$

$$\Sigma \times I$$

$$n\Sigma = \sum_{i=1}^{n} ⓘⓘ^T$$

$$\rightarrow \cancel{\Sigma = I}$$

$$\boxed{\Sigma_1 = \frac{1}{n}\sum_{i=1}^{n} ⓘⓘ^T}$$

$\sum h(x^{(i)}) \mathbb{1}\{y=1\} \cdot p(y=1)$.

(f) log reg higher accuracy, boundary looks ~~be~~ nicer

(g) Data 2: similar, probably normal data so GDA is good
DATA 1: bad as not normal distribution

(h) $x_2' = \log(x_2+1)$, maybe z-score, maybe x_1^2, x_2^2 etc needed

(Q2) (c) $\Rightarrow \frac{p(y=1 \mid t=1, x)\, p(t=1 \mid x)}{p(y=1 \mid x)}$

$$= \frac{NR}{\underbrace{p(y=1 \mid t=0, x)}_{0}\, p(t=0 \mid x) + \dots t=1}$$

$=$

$\therefore NR = DR$

$\therefore = 1$

(d) $p(y=1 \mid x) = p(t=1 \mid x)\, p(y=1 \mid t=1, x)$
$= p(t=1 \mid x)\, \alpha$

(e) show $\alpha = \mathbb{E}[h(x^{(i)}) \mid y^{(i)}=1]$ $\quad h(x^{(i)}) = p(y^{(i)}=1 \mid x^{(i)})$

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

$\mathbb{E}[h \mid y=1]$

$= \frac{\mathbb{E}[h(x^{(i)}) \cdot \mathbb{1}\{y^{(i)}=1\}]}{p(y^{(i)}=1)}$ if $t=1$

$p_{y^{(i)}} = p$ if $t=1$

$\therefore \frac{\mathbb{E}[h(x^{(i)}) \mid t=1] \cdot p(y^{(i)}=1)}{p(y^{(i)}=1)} = \alpha$

Q3. (a) $p(y;\lambda) = \dfrac{e^{-\lambda}\lambda^{y}}{y!}$

$\dfrac{1}{y!}(e^{-\lambda} e^{y\log\lambda})$

$\dfrac{1}{y!} \cdot (e^{y\log\lambda - \lambda})$

$\dfrac{1}{y!} \exp(y \log\lambda - \lambda)$

$\boxed{\begin{aligned} \therefore b(y) &= 1/y! \qquad \eta = \log\lambda \\ a(\eta) &= \lambda = e^{\eta} \\ T(y) &= y \end{aligned}}$

(b) $g(\eta) = \mathbb{E}[T(y);\eta]$ $\qquad \because h_\theta(x) = g(\eta)$

$= \mathbb{E}[y;\eta]$

$= \boxed{\lambda = e^{\eta}}$

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

$\therefore \dfrac{\partial l^{i}(\theta)}{\partial\theta_j} = \dfrac{\partial \log\left(\frac{1}{y^{(i)}!}\exp((\theta^T x^{(i)})^T y^{(i)} - e^{\eta})\right)}{\partial\theta_j}$

$= \dfrac{\partial\left((\theta^T x^{(i)})^T y^{(i)} - e^{\theta^T x^{(i)}}\right)}{\partial\theta_j}$ $\qquad \underset{1\times n \;\; n\times 1}{\theta^T x^{(i)}}$ scalar

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

$\therefore \theta_j := \theta_j + \alpha \dfrac{\partial l(\theta)}{\partial\theta_j}$

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

PS1

Q4 (a) Show $E[Y;\eta] = \frac{\partial a(\eta)}{\partial \eta}$

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

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

① $= \int_{-\infty}^{\infty} b(y)\exp(\eta y - a(\eta)) \cdot (y - a'(\eta))\,dy$

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

$E[y;\eta] = \int_{-\infty}^{\infty} y \cdot b(y)\exp(\eta y - a(\eta))\,dy$

$\therefore$ ① $\frac{\partial \int p(y;\eta)\,dy}{\partial \eta} = 0$ $\quad \& = \frac{\partial 1}{\partial \eta}$

$\int_{-\infty}^{\infty}(b(y)\exp(\eta y - a(\eta))\,y\,dy - \int_{-\infty}^{\infty} p\,a'(\eta))\,dy$

$E[y;\eta] - a'(\eta)\int_{-\infty} p\,dy$

$\therefore E[y;\eta] - a'(\eta) = 0$

$\therefore \boxed{E[y;\eta] = a'(\eta)}$

(b) Show $Var(Y;\eta) = \frac{\partial^2}{\partial \eta^2} a(\eta)$

$$Var(y;\eta) = E[X^2] - E[X]^2$$

$$\underbrace{= \int y^2 b(y) \exp(\eta y - a(\eta))\, dy - a'(\eta)^2}_{\text{from a}} \rightarrow ②$$

$$\cancel{= b(y)\exp(\eta y - a(\eta)) - a'(\eta)}$$

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

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

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

$$= \int_{-\infty}^{\infty} y\, b(y) \exp(\eta y - a(\eta))\,(y - a'(\eta))\, dy$$

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

$$- \int_{-\infty}^{\infty} a'(\eta)\, y\, b(y) \exp(\eta y - a(\eta)\, dy$$

$$= - a'(\eta) \underbrace{\int_{-\infty}^{\infty} y\, b(y) \exp(\eta y - a(\eta)\, dy}_{a'(\eta)}$$

$$= E[y^2;\eta] - E[y;\eta]^2$$

$$= \boxed{Var(y;\eta) = a''(\eta)}$$

■

(c) $$\ell'(\theta) = -\log p(y;\eta)$$
$$= -\log b(y)\exp(\eta y - a(\eta))$$
$$= -\log b(y) + (-\log(\exp(\eta y - a(\eta)))$$
$$= a(\eta) - \eta y - \log b(y)$$
$$= a(\theta^T x) - \theta^T x\, y - \log b(y)$$

$$\nabla_\theta \ell'(\theta) = \frac{\partial a(\theta^T x)}{\partial \theta} - \frac{\partial \theta^T x y}{\partial \theta} - 0$$
$$= \frac{\partial a(\theta^T x)}{\partial \theta^T x} x - y\frac{\partial \theta^T x}{\partial \theta}$$
$$= \frac{\partial a(\eta)}{\partial \eta} x - y x$$

Hessian $$\nabla^2_\theta \ell'(\theta) = \frac{\partial}{\partial \theta}\left(\frac{\partial a(\eta)}{\partial \eta}\cdot x - yx\right)$$ $\to 0$
$$= \frac{\partial}{\partial \theta}\left(\frac{\partial a(\theta^T x)}{\partial \theta^T x}\cdot x\right)$$
$$= x\,\frac{\partial^2 a(\theta^T x)}{\partial (\theta^T x)^2}\cdot\frac{\partial \theta^T x}{\partial \theta}$$
$$= x x^T \mathrm{Var}(y)$$

Eg. $x \in \mathbb{R}^3$

$[x_1^i\ x_2^i\ x_3^i]\begin{bmatrix} x_1^i \\ x_2^i \\ x_3^i \end{bmatrix}\cdot \mathrm{Var}(y^i)$

$\therefore$ Σ gives hessian

$$\ell(\theta) = -\log \prod_{i=1}^{n} p(y;\eta)$$
$$= \sum_{i=1}^{n} \ell'(\theta)$$
$$\therefore \nabla_\theta \ell(\theta) = \sum_{i=1}^{n} \nabla_\theta \ell'(\theta) = \sum_{i=1}^{n} E[y;\eta]\, x^{(i)} - y^{(i)} x^{(i)}$$

$$\therefore \nabla^2_\theta \ell(\theta) = \sum \nabla^2_\theta \ell(\theta) = \sum_{i=1}^{n} \underbrace{x^{(i)} x^{(i)T}}_{\therefore\ \geq 0} \underbrace{\mathrm{Var}(y^{(i)})}_{\geq 0}$$

$\therefore$ PSD

$\therefore$ NLL convex ▨

Q5:-

(a) $J(\theta) = \frac{1}{2} \sum_{i=1}^{n} (\theta^T \hat{x}^{(i)} - y^{(i)})^2$

$$\nabla_\theta J(\theta) = \sum_{i=1}^{n} (\theta^T \hat{x}^{(i)} - y^{(i)})(\hat{x}^{(i)})$$

$$\therefore \theta := \theta - \alpha \nabla_\theta J(\theta)$$

(b) $(X^T X)^{-1} X^T Y$

(c) Higher degree tends to fit better. But numerical issues with inversion / Lin alg solving.

(d) Even $k = 0$ can fit with right basis of sin
Higher k can overfit

(e) Higher degree can fit all but seems like overfitting rather than having a predictive use