CS229 Problem Set #1

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

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

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

$$H\,J(\theta) = \frac{\partial}{\partial\theta_j\partial\theta_k}J(\theta)$$

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

thus $$z^T H z = \frac{1}{m}\sum_{i=1}^{m}\sum_{j,k=1}^{m} g(\theta^T(x^{(i)}))(1-g(\theta^T x^{(i)}))x_j^{(i)}x_k^{(i)} z_j z_k$$

$$= \frac{1}{m}\sum_{i=1}^{m}\sum_{j,k=1}^{m} g(\theta^T(x^{(i)}))(1-g(\theta^T x^{(i)}))(x^{(i)}\cdot z)^2 \geq 0$$

1\. (c) see in notes

1\. (d) see in notes

2\. (a) since that each $y^{(i)}$ and $x^{(i)}$ are conditionally independent given $t^{(i)}$

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

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

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

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

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

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

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

$$= \alpha$$

3.(c) $P(y;\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}$ $\lambda = e^{\eta} = e^{\theta^T x}$

loss function : $L(\theta) = \prod_{i=1}^{m} \frac{e^{-\lambda^{(i)}} \lambda^{(i)y^{(i)}}}{y^{(i)}!}$

$$= \prod_{i=1}^{m} \frac{e^{-e^{\theta^T x^{(i)}}} (e^{\theta^T x^{(i)}})^{y^{(i)}}}{y^{(i)}!}$$

$$\log L(\theta) = \sum_{i=1}^{m} (-e^{\theta^T x^{(i)}} + \theta^T x^{(i)} y^{(i)} - \log(y^{(i)}!))$$

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

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

thus the step (update rule should be:

$$\theta_j := \theta_j + \alpha (y^{(i)} - \underbrace{e^{\theta^T x^{(i)}}}_{\text{this is } H_\theta(x)}) x^{(i)}$$

4.(a) $E(Y|X,\theta)$

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

$$= \int [\exp(\eta y - a(\eta))\, b(y) \cdot (y - a'(\eta))]\,dy$$

$$= \int P(y;\eta) \cdot y\,dy - \int P(y;\eta) \cdot a'(\eta)\,dy$$

$$= 0$$

$$E(y|\eta) = a'(\eta)$$

4.(b) $VAR(Y|X,\theta)$

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

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

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

$$= \int y \cdot b(y) \exp(\eta y - a(\eta))(y - a'(\eta))\,dy - a''(\eta)$$

$$= \int y^2 \cdot P(y;\eta)\, dy - \int y \cdot P(y;\eta) \cdot a'(\eta)\, dy - a''(\eta)$$

$$= E(y^2|\eta) - [E(y|\eta)]^2 - a''(\eta)$$

$$\Rightarrow a''(\eta) = VAR(Y|X,\eta)$$

4.(c) $$l(\theta) = -\prod_{i=1}^{m} \log P(y^{(i)}|X^{(i)},\theta)$$

$$= -\sum_{i=1}^{m} \log \left(b(y^{(i)}) \exp(\eta y^{(i)} - a(\eta)) \right)$$

$$= -\sum_{i=1}^{m} [\log b(y^{(i)}) + \eta y^{(i)} - a(\eta)] \quad \text{where } \eta = \theta^T X$$

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

$$\nabla l(\theta) = \frac{d}{d\theta} l(\theta)$$

$$= -\sum_{i=1}^{m} (X_k^{(i)} y^{(i)} - a'(\theta^T X_k^{(i)}) X_k^{(i)})$$

$$H\, l(\theta) = (\nabla l(\theta))'$$

$$= \frac{d^2}{d\theta^2} l(\theta)$$

$$= \sum_{i=1}^{m} a''(\theta^T X_k^{(i)}) X_k^{(i)} \cdot X_j^{(i)}$$

To show $H(l(\theta))$ is PSD, calculate $Z^T H Z$

$$Z^T H(l(\theta)) Z = \sum_{k,j=1}^{m} \sum_{i=1}^{N} a''(\theta^T X_k^{(i)}) X_k^{(i)} \cdot X_j^{(i)} \cdot Z_k \cdot Z_j$$

$$= \sum_{k,j=1}^{m} \sum_{i=1}^{N} a''(\theta^T X_k^{(i)}) (X_k^{(i)} \cdot Z_k)^2$$

Since $a''(\theta^T X_k^{(i)})$ is equals to $VAR(Y|X;\eta) \geq 0$, $(X_k^{(i)} \cdot Z_k)^2 \geq 0$ always.

We have $l(\theta)$ is <u>convex</u>. ▨

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

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

if $W = \begin{bmatrix} \frac{w^{(1)}}{2} & & & \\ & \frac{w^{(2)}}{2} & & \\ & & \ddots & \\ & & & \frac{w^{(m)}}{2} \end{bmatrix}$ which is a diagonal matrix

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

5.(a) ii if $W = \frac{1}{2}I$

$$J(\theta) = (X\theta - y)^T(\tfrac{1}{2}X\theta - \tfrac{1}{2}y)$$

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

$$\nabla J(\theta) = X^T X\theta - X^T y = 0$$

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

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

5.(a) iii $P(y^{(i)} | x^{(i)}; \theta) = \frac{1}{\sqrt{2\pi}\sigma^{(i)}} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right)$

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

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

$$\hat{\theta} = \arg\max l(\theta)$$

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

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

thus $w^{(i)} = 1/(\sigma^{(i)})^2$