# Gyujin Park

# CSE 6740 HOMEWORK #3

## 1. Questions

**Question 1.** (a)

$$
\begin{aligned}
\hat{\theta} &= (X^T X)^{-1} X^T Y \\
E(\hat{\theta}) &= E((X^T X)^{-1} X^T Y) \\
&= (X^T X)^{-1} X^T E(Y) \\
&= (X^T X)^{-1} X^T X \theta \\
&= \theta
\end{aligned}
$$

$\therefore \hat{\theta}$ is unbiased estimator for $\theta$.

(b)

$$
\begin{aligned}
Cov(\hat{\theta}) &= Cov((X^T X)^{-1} X^T Y) \\
&= (X^T X)^{-1} X^T Cov(Y) ((X^T X)^{-1} X^T)^T \\
&= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} \\
&= \sigma^2 (X^T X)^{-1}
\end{aligned}
$$

(c)
I agree. $\hat{\theta}$ follows Gaussian distribution because $\hat{\theta}$ is a linear transformation of $Y$.

**Question 2.** (a)

$$\hat{\theta} = (X^T X + \lambda I)^{-1} X^T Y$$
$$E(\hat{\theta}) = (X^T X + \lambda I)^{-1} X^T X \theta$$

$$\begin{aligned} Cov(\hat{\theta}) &= Cov((X^T X + \lambda I)^{-1} X^T Y) \\ &= (X^T X + \lambda I)^{-1} X^T Cov(y)((X^T X + \lambda I)^{-1} X^T)^T \\ &= \sigma^2 (X^T X + \lambda I)^{-1} X^T X (X^T X + \lambda I)^{-1} \end{aligned}$$

**Question 3.2.** (a)
Let's assume the class distribution is Bernoulli with parameter $\theta$.

$$p(x|y=1) = (2\pi)^{-\frac{d}{2}}|\Sigma_1|^{-\frac{1}{2}}e^{-\frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)}$$
$$p(x|y=-1) = (2\pi)^{-\frac{d}{2}}|\Sigma_2|^{-\frac{1}{2}}e^{-\frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)}$$

$$\begin{aligned} h(x) &= \log\left(\frac{p(x|y=1)\theta}{p(x|y=-1)(1-\theta)}\right) \\ &= \log\left(\sqrt{\frac{|\Sigma_2|}{|\Sigma_1|}}\frac{\theta}{1-\theta}e^{-\frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)+\frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)}\right) \\ &= \frac{1}{2}\log\frac{|\Sigma_2|}{|\Sigma_1|} + \log\frac{\theta}{1-\theta} - \frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) + \frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2) \end{aligned}$$

Decision boundary($h(x)=0$): A quadratic curve

(b)
If $\Sigma_1 = \Sigma_2$, Quadratic terms in $h(x)$ cancel out and $h(x)$ becomes linear.
$\therefore$ Decision boundary: A line

(c)
If $\Sigma_1 = \Sigma_2 = I_d$,

$$(x-\mu_1)^T(x-\mu_1) - (x-\mu_2)^T(x-\mu_2) = 2\log\frac{\theta}{1-\theta}$$
$$\|x-\mu_1\|^2 - \|x-\mu_2\|^2 = 2\log\frac{\theta}{1-\theta}$$

Decision boundary: A line perpendicular to the line connecting $\mu_1$ and $\mu_2$.

**Question 4.** (a)

$$
\begin{aligned}
L(z) &= \log(1+e^{-z}) \\
L'(z) &= \frac{-e^{-z}}{1+e^{-z}} \\
L''(z) &= \frac{e^{z}}{(1+e^{z})^2} > 0
\end{aligned}
$$

$\therefore L(z)$ is a convex function.

(b)

We can prove it using mathematical induction. By definition, it's true when $m = 2$.
Let's assume the inequality holds when $m = n$.
If $m = n+1$ and $\sum_{i=1}^{n+1} \alpha_i = 1$,

$$
\begin{aligned}
f\left(\sum_{i=1}^{n+1} \alpha_i x_i\right) &= f\left((\alpha_1+\alpha_2)\left(\frac{\alpha_1}{\alpha_1+\alpha_2}x_1 + \frac{\alpha_2}{\alpha_1+\alpha_2}x_2\right) + \sum_{i=3}^{n+1} \alpha_i x_i\right) \\
&\geq (\alpha_1+\alpha_2) f\left(\frac{\alpha_1}{\alpha_1+\alpha_2}x_1 + \frac{\alpha_2}{\alpha_1+\alpha_2}x_2\right) + \sum_{i=3}^{n+1} \alpha_i f(x_i) \\
&\geq (\alpha_1+\alpha_2)\left(\frac{\alpha_1}{\alpha_1+\alpha_2}f(x_1) + \frac{\alpha_2}{\alpha_1+\alpha_2}f(x_2)\right) + \sum_{i=3}^{n+1} \alpha_i f(x_i) \\
&= \sum_{i=1}^{n+1} \alpha_i f(x_i)
\end{aligned}
$$

$\therefore$ We are done.

**Question 5.** (a)

$$
\begin{aligned}
P(Y=1|X=x) &= \frac{e^{w_0+w^T x}}{1+e^{w_0+w^T x}} \\
P(Y=0|X=x) &= \frac{1}{1+e^{w_0+w^T x}} \\
\log\left(\frac{P(Y=1|X=x)}{P(Y=0|X=x)}\right) &= \log\left(e^{w_0+w^T x}\right) \\
&= w_0 + w^T x
\end{aligned}
$$

$\therefore$ log-odds of success is a linear function.

(b) [Extra Credit]

$$
\begin{aligned}
P(y^i|x^i) &= \frac{e^{\theta_{y^i}\cdot x^i}}{\sum_c e^{\theta_c \cdot x^i}} \\
L = \Pi_i P(y^i|x^i) &= \Pi_i \frac{e^{\theta_{y^i}\cdot x^i}}{\sum_c e^{\theta_c \cdot x^i}}
\end{aligned}
$$

$$
\begin{aligned}
l = \log L &= \sum_i \log \frac{e^{\theta_{y^i}\cdot x^i}}{\sum_c e^{\theta_c \cdot x^i}} \\
&= \sum_i \left(\theta_{y^i}\cdot x^i - \log \sum_c e^{\theta_c \cdot x^i}\right)
\end{aligned}
$$

$$
\begin{aligned}
\therefore \frac{dl}{d\theta_c} &= \sum_i \frac{d}{d\theta_c}(\theta_{y^i}\cdot x^i) - \frac{d}{d\theta_c}\left(\sum_i \log \sum_c e^{\theta_c \cdot x^i}\right) \\
&= \sum_i I(y_i = c)x^i - \sum_i \frac{e^{\theta_c \cdot x^i}}{\sum_c e^{\theta_c \cdot x^i}} x^i
\end{aligned}
$$

**Question 6.** (a)

$$E = \sum_{u,i} \left( M_{ui} - \sum_{k=1}^{r} U_{uk} V_{ik} \right)^2$$
$$\frac{\partial E}{\partial U_{uk}} = \sum_{i} 2 \left( M_{ui} - \sum_{k=1}^{r} U_{uk} V_{ik} \right) (-V_{ik})$$
$$\frac{\partial E}{\partial V_{ik}} = \sum_{u} 2 \left( M_{ui} - \sum_{k=1}^{r} U_{uk} V_{ik} \right) (-U_{uk})$$

(b)

$$E = \sum_{u,i} \left( M_{ui} - \sum_{k=1}^{r} U_{uk} V_{ik} \right)^2 + \lambda \sum_{u,k} U_{uk}^2 + \lambda \sum_{i,k} V_{ik}^2$$
$$\frac{\partial E}{\partial U_{uk}} = \sum_{i} 2 \left( M_{ui} - \sum_{k=1}^{r} U_{uk} V_{ik} \right) (-V_{ik}) + \lambda \cdot 2 \cdot U_{uk}$$
$$\frac{\partial E}{\partial V_{ik}} = \sum_{u} 2 \left( M_{ui} - \sum_{k=1}^{r} U_{uk} V_{ik} \right) (-U_{uk}) + \lambda \cdot 2 \cdot V_{ik}$$