EhsanMerrikhi HW02_400101967

1-

1)

$$W = \begin{pmatrix} -5 \\ 0.1 \\ 0.25 \end{pmatrix} \qquad x = \begin{pmatrix} 1 \\ 80 \\ 18 \end{pmatrix} \qquad w^T x = -5 + 8 + 4.5 = 7.5$$

$$Ber(y \mid \sigma(7.5)) = P(y \mid x; \theta) \rightarrow P(y=1 \mid x; \theta) = \frac{1}{1+e^{-7.5}} = 0.99944$$

2)

$$P(y \mid x; \theta) = 0.9 \rightarrow \frac{1}{1+e^{-a}} = 0.9 \rightarrow a = \ln(9.9) \simeq 2.2$$

$$\left.\begin{array}{l} x_1 = ? \\ x_2 = 16 \end{array}\right\} \rightarrow w^T x = 2.2 \rightarrow -5 + x_2/10 + 4 = 2.2 \rightarrow x_1 = 32$$

2-

1) $$\sum_{k=1}^{K} P(Y=y_k|X)=1$$

$$P(Y=y_k|X)+\sum_{k=1}^{K-1} P(Y=y_k|X)=1 \longrightarrow P(Y=y_k|X)=1-\sum_{k=1}^{K-1} P(Y=y_k|X)$$

$$=1-\sum_{k=1}^{K-1} C\exp\left(w_{k0}+\sum_{i=1}^{d} w_{k0} X_i\right)=C\exp\left(w_{k0}+\sum_{k=1}^{d} w_{ki} X_i\right)$$

$$\rightarrow C=\frac{1}{\sum_{k=1}^{K}\exp\left(w_{k0}+\sum_{k=1}^{d} w_{ki} X_i\right)} \longrightarrow P(Y=y_k|X)=\frac{\exp\left(w_{k0}+\sum_{i=1}^{d} w_{ki} X_i\right)}{\sum_{k=1}^{K}\exp\left(w_{k0}+\sum_{k=1}^{d} w_{ki} X_i\right)}$$

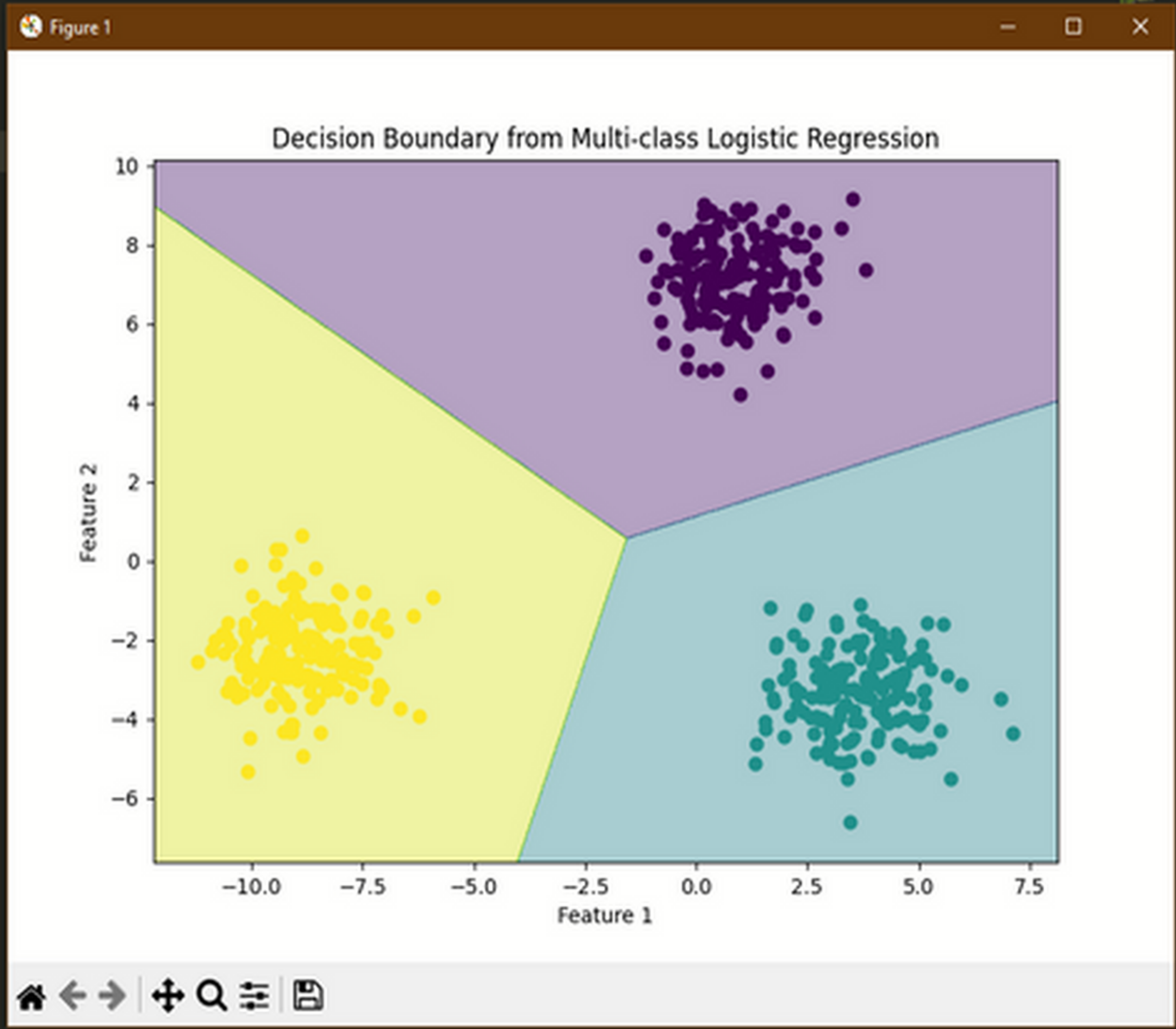
2) برای این که نتیجه بگیریم داده در کلاس y_k باشد، باید $P(Y=y_k|X)$ ماکسیمم شود.

$$\hat{y}=\underset{k}{\arg\max}\ \frac{\exp\left(w_{k0}+\sum_{i=1}^{d} w_{ki} X_i\right)}{\sum_{k=1}^{K}\exp\left(w_{k0}+\sum_{k=1}^{d} w_{ki} X_i\right)}$$

3)

D: > Ehsan > studies > uni > sem 6 > ML > HW > HW2 > 2_3.py > ...

```python
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import make_blobs
from sklearn.linear_model import LogisticRegression

X, y = make_blobs(n_samples=500, n_features=2, random_state=67, centers=3)

model = LogisticRegression(multi_class='multinomial')
model.fit(X, y)

y_pred = model.predict(X)

h = 0.02
x_min, x_max = X[:, 0].min() - 1, X[:, 0].max() + 1
y_min, y_max = X[:, 1].min() - 1, X[:, 1].max() + 1
xx, yy = np.meshgrid(np.arange(x_min, x_max, h), np.arange(y_min, y_max, h))

Z = model.predict(np.c_[xx.ravel(), yy.ravel()])
Z = Z.reshape(xx.shape)

plt.figure(figsize=(8, 6))
plt.contourf(xx, yy, Z, alpha=0.4)
plt.scatter(X[:, 0], X[:, 1], c=y, cmap='viridis')
plt.title('Decision Boundary from Multi-class Logistic Regression')
plt.xlabel('Feature 1')
plt.ylabel('Feature 2')
plt.show()
```



4) $\mathcal{L}(\theta) = \prod_{k=1}^{n} P(y = y_k \mid X = x_k, \theta) \rightarrow \sum_{k=1}^{n} \log P(y = y_k \mid X = x_k, \theta)$

$\rightarrow \mathcal{L}(\theta) = \sum_{k=1}^{n} \left(y_k \log\left(\frac{1}{1+\exp(-\theta^T x_k)}\right) + (1-y_k) \log\left(1 - \frac{1}{1+\exp(\theta^T x_k)}\right)\right)$

5)

$$\mathcal{L}(x) = \sum_{k=1}^{N} \left(y_k \log\left(\frac{1}{1+\exp(-\theta^T x_k)} \right) + (1-y_k) \log\left(1 - \frac{1}{1+\exp(\theta^T x_k)} \right) \right) - \lambda \sum_{k=1}^{K} \|w_k\|^2$$

$$\frac{\partial}{\partial \theta_i} \mathcal{L}(x) = \sum_{k=1}^{N} \left(\left(y_k - \frac{1}{1+\exp(-\theta^T x_k)} \right) x_i^k \right) - 2\lambda\theta_i$$

$$\rightarrow \nabla_\theta \mathcal{L}(x) = \sum_{k=1}^{N} \left(y_k - P(y_k \mid X_k, \theta) \right) x_k - 2\lambda\theta$$

6)

$$\theta^{t+1} = \theta^t + \alpha \nabla_\theta \mathcal{L}(x) = \theta^t + \alpha \sum_{k=1}^{N} \left(y_k - P(y_k \mid X_k, \theta^t) \right) x_k - 2\alpha\lambda\theta^t$$

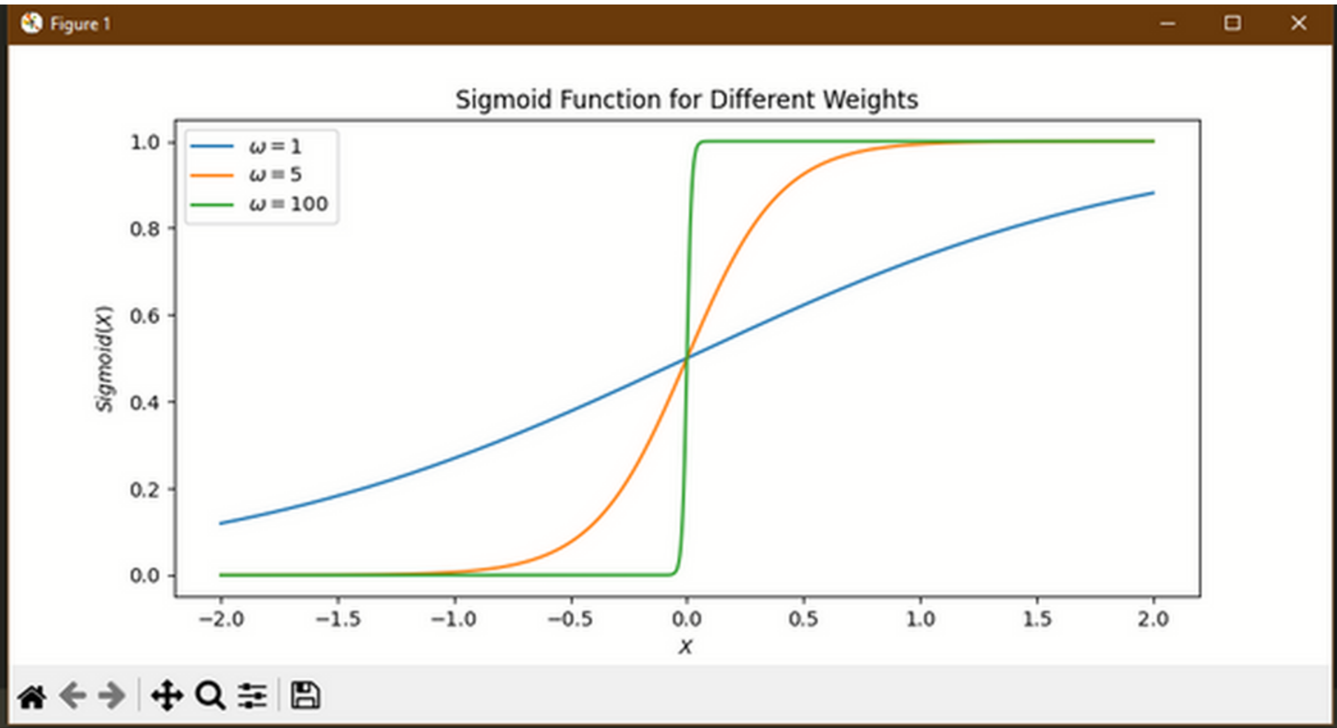
3-

1)

افزایش w باعث می‌شود تابع سیگما به تابع u(t) شبیه‌تر شود

این موضوع باعث می‌شود تابع سیگما به داده‌ها خیلی نزدیک شود و overfit رخ دهد

```python
import numpy as np
import matplotlib.pyplot as plt


def sigmoid(Z):
    return 1 / (1 + np.exp(Z))


X = np.linspace(-2, 2, 1000)
plt.figure(figsize=(10, 6))

for w in [1, 5, 100]:
    y = sigmoid(-w * X)
    plt.plot(X, y, label=f"$\omega = {w}$")

plt.title("Sigmoid Function for Different Weights")
plt.xlabel("$X$")
plt.ylabel("$Sigmoid(X)$")
plt.legend()
plt.show()
```



2)

$$\hat{\omega}_{MAP} = \max_{w} \prod_{i=1}^{\sim} P(y_i|X_i,w)\,P(w) = \max_{w} \sum_{i=1}^{\sim} \log P(y_i|X_i,w) + \log P(w)$$

$$= \max_{w}\left(\sum_{i=1}^{\sim} \log P(y_i,X_i,w) + \log(2\pi)^{-\frac{d+1}{2}} - \frac{1}{2}w^Tw\right) = \min_{w} \overbrace{\left(-\sum_{i=1}^{\sim} \log P(y_i|X_i;w) + \frac{1}{2}w^Tw\right)}^{b}$$

$$\nabla_w\left(-\sum_{i=1}^{\sim} \log P(y_i|X_i;w) + \frac{1}{2}w^Tw\right) = \sum_{i=1}^{\sim}\left(\frac{1}{1+\exp(w^TX_i)} - y_i\right)X_i + w \rightarrow w^{t+1} = w^t + \alpha\nabla_w b$$

5-

1)

$$P(Y \mid x_1=0, x_2=0) = \frac{P(x_1=0, x_2=0 \mid y)\, P(y)}{P(x_1=0, x_2=0)} = \frac{P(x_1=0 \mid y)\, P(x_2=0 \mid y)\, P(y)}{\sum_{c'} P(x_1=0 \mid y=c')\, P(x_2=0 \mid y=c')\, P(y=c')}$$

$$P(x_1 \mid y=c) = Ber(x_1; \theta_c) \longrightarrow P(x_1 \mid y=c) = \theta_c^{x_1}(1-\theta_c)^{1-x_1} \longrightarrow P(x_1 \mid y=c) = \theta_c^{0}(1-\theta_c)^{1}$$
$$= 1-\theta_c$$

$$P(x_2 \mid y=c) = \mathcal{N}(x_2; \mu_c, \sigma_c^2) \longrightarrow P(x_2 \mid y=c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(-\frac{(x_2-\mu_c)^2}{2\sigma_c^2}\right) = \frac{\exp\left(-\frac{\mu_c^2}{2\sigma_c^2}\right)}{\sqrt{2\pi\sigma_c^2}}$$

$$c=0 \longrightarrow \theta_c = 0,\ \mu_c = -1,\ \sigma_c^2 = 1 \longrightarrow P(x_1=0, x_2=0 \mid y=0)\, P(y=0) = P(x_1=0 \mid y)\, P(x_2=0 \mid y)\, P(y=0)$$
$$= (0.5)\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\right)(0.5) = \frac{e^{-1/2}}{4\sqrt{2\pi}}$$

$$c=1 \longrightarrow \theta_c = 0,\ \mu_c = 0,\ \sigma_c^2 = 1 \longrightarrow P(x_1=0, x_2=0 \mid y=1)\, P(y=1) = P(x_1=0 \mid y)\, P(x_2=0 \mid y)\, P(y=1)$$
$$= (0.5)\frac{1}{\sqrt{2\pi}} \exp(-1/2)(0.25) = \frac{e^{-1/2}}{8\sqrt{2\pi}}$$

$$\sum_{c'=0}^{2} P(x_1=0|y=c')\,P(x_2=0|y=c')\,P(y=c') = \frac{1}{8\sqrt{2\pi}}\left(3\exp(-1/2)+1\right)$$

$$P(y=0|x_1=0,x_2=0) = \frac{\frac{1}{4\sqrt{2\pi}}\exp(-1/2)}{\frac{1}{8\sqrt{2\pi}}\left(3\exp(-1/2)+1\right)} = \frac{2e^{-1/2}}{3e^{-1/2}+1}$$

$$P(y=1|x_1=0,x_2=0) = \frac{\frac{1}{8\sqrt{2\pi}}}{\frac{1}{8\sqrt{2\pi}}\left(3\exp(-1/2)+1\right)} = \frac{1}{3e^{-1/2}+1}$$

$$P(y=2|x_1=0,x_2=0) = \frac{\frac{1}{8\sqrt{2\pi}}\exp(-1/2)}{\frac{1}{8\sqrt{2\pi}}\left(3\exp(-1/2)+1\right)} = \frac{e^{-1/2}}{3e^{-1/2}+1}$$

$$P(Y|x_1=0,x_2=0) = \frac{1}{3e^{-1/2}+1}\begin{pmatrix} 2e^{-1/2} \\ 1 \\ e^{-1/2} \end{pmatrix}$$

2)

$$P_{Y|X_1}(y|x_1=0) = \frac{P(y=c|\pi)P(x_1=0|y=c,\theta_c)}{\sum_{c'} P(y=c'|\pi)P(x_1=0|y=c',\theta_{c'})}$$

$$\sum_{c'} P(y=c'|\pi)P(x_1=0|y=c',\theta_{c'}) = \frac{1}{2}\times\frac{1}{2}+\frac{1}{4}\times\frac{1}{2}+\frac{1}{4}\times\frac{1}{2}=\frac{1}{2}$$

$$P(y=c|x_1=0;\theta) = \begin{cases} {}^{1/4}/_{1/2} = 1/2 & c=0 \\ {}^{1/8}/_{1/2} = 1/4 & c=1 \\ {}^{1/8}/_{1/2} = 1/4 & c=2 \end{cases}$$

3)

$$P_{Y|X_2}(y|x_2=0) = \frac{P(y=c|\pi)P(x_2=0|y=c,\theta_c)}{\sum_{c'} P(y=c'|\pi)P(x_2=0|y=c',\theta_{c'})}$$

$$\sum_{c'} P(y=c'|\pi)P(x_2=0|y=c',\theta_{c'}) = \frac{1}{2}\frac{e^{-1/2}}{\sqrt{2\pi}}+\frac{1}{4}\frac{1}{\sqrt{2\pi}}+\frac{1}{4}\frac{e^{-1/2}}{\sqrt{2\pi}} = \frac{1}{4\sqrt{2\pi}}\left(3e^{-1/2}+1\right)$$

$$P(y=c|X_2=0): \quad \frac{2e^{-1/2}}{3e^{-1/2}+1},\ c=0 \qquad \frac{1}{3e^{-1/2}+1},\ c=1 \qquad \frac{e^{-1/2}}{3e^{-1/2}+1},\ c=2$$

7-

1) $P(x|y=1;\theta) = \mathcal{N}(x|\hat{\mu}_1, \hat{\sigma}_1^2)$ $\quad P(x|y=2,\theta) = \mathcal{N}(x|\hat{\mu}_2, \hat{\sigma}_2^2)$

$y=1$ if $\quad P(y=1|x;\theta) > P(y=2|x;\theta)$

$$P(y=1|x;\theta) \sim P(x|y=1;\theta)P(y=1;\theta) = \frac{n_1}{N}\frac{1}{\hat{\sigma}_1\sqrt{2\pi}} e^{-\frac{(x-\hat{\mu}_1)^2}{2\sigma_1^2}}$$

$$P(y=2|x;\theta) \sim P(x|y=2;\theta)P(y=2;\theta) = \frac{n_2}{N}\frac{1}{\hat{\sigma}_2\sqrt{2\pi}} e^{-\frac{(x-\hat{\mu}_2)^2}{2\sigma_2^2}}$$

$$P(y=1|x;\theta) > P(y=2|x;\theta)$$

$$\rightarrow \frac{n_1}{N}\frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{(x-\hat{\mu}_1)^2}{2\sigma_1^2}} > \frac{n_2}{N}\frac{1}{\hat{\sigma}_2\sqrt{2\pi}} e^{-\frac{(x-\hat{\mu}_2)^2}{2\sigma_2^2}}$$

$$\rightarrow \ln n_1 - \frac{1}{2}\left(\frac{x-\hat{\mu}_1}{2\hat{\sigma}_1^2}\right)^2 > \ln n_2 - \frac{1}{2}\left(\frac{x-\hat{\mu}_2}{2\hat{\sigma}_2^2}\right)^2 \rightarrow -\ln\left(\frac{n_2}{n_1}\right) > \frac{1}{2}\left(\left(\frac{x-\hat{\mu}_1}{\hat{\sigma}_1^2}\right)^2 - \left(\frac{x-\hat{\mu}_2}{\hat{\sigma}_2^2}\right)^2\right)$$

$$\frac{1}{2\sigma^2}\left(2x(\hat{\mu}_2-\hat{\mu}_1)+(\hat{\mu}_1^2-\hat{\mu}_2^2)\right) < -\ln\left(\frac{n_2}{n_1}\right)$$

in group 1 if $\frac{\hat{\mu}_2-\hat{\mu}_1}{\sigma^2} X < \frac{\hat{\mu}_2^2-\hat{\mu}_1^2}{2\sigma^2} - \ln\left(\frac{n_2}{n_1}\right)$

in group 2 if $\frac{\hat{\mu}_2-\hat{\mu}_1}{\sigma^2} X > \frac{\hat{\mu}_2^2-\hat{\mu}_1^2}{2\sigma^2} - \ln\left(\frac{n_2}{n_1}\right)$

2)

$P(y|x;\theta) = \mathcal{N}(y|w_0+w^T x, \sigma^2)$ $\quad \hat{\beta}_1 = (x^T x)^{-1} x^T y = \frac{\sum x_i y_i}{\sum |x_i|^2}$

$$y_i = \begin{cases} -n/n_1 & \text{class } 1 \\ n/n_2 & \text{class } 2 \end{cases} \qquad \sum x_i y_i = n\left(\frac{1}{n_2}\sum_{i=(y_i\in 2)} x_i - \frac{1}{n_1}\sum_{i=(y_i\in 1)} x_i\right) = n(\mu_2-\mu_1)$$

$$\sigma^2 = \frac{1}{n}\sum(x_i-\mu_i)^2 = \frac{1}{n}\left(\sum_{i=y_i\in 1}(x_i-\mu_1)^2 + \sum_{i=y_i\in 2}(x_i-\mu_2)^2\right)$$

$$= \frac{1}{n}\left(\sum x_i^2 - n_1\mu_1^2 - n_2\mu_2^2\right)$$

$$\hat{\beta}_1 = \frac{\sum_{i=y_i\in 1} -n/n_1\, x_i + \sum_{i=y_i\in 2} n/n_2\, x_i}{\sum_{i=1}^{n} x_i^2} = n\,\frac{\sum_{i=y_i\in 1} -1/n_1\, x_i + \sum_{i=y_i\in 2} 1/n_2\, x_i}{\sum_{i=1}^{n} x_i^2}$$

$$= \frac{\hat{\mu}_2 - \hat{\mu}_1}{\frac{1}{n}\sum x_i^2}$$

3)

$$\frac{\hat{\mu}_1 - \hat{\mu}_2}{\sigma^2} X > \frac{\hat{\mu}_2 - \hat{\mu}_1}{2\sigma^2} - \log\frac{n_2}{n_1} \to \frac{n-1}{n}\hat{\beta}_1 X > \frac{\hat{\mu}_2 - \hat{\mu}_1}{2\sigma^2} - \log\frac{n_2}{n_1}$$

$$\hat{\beta}_1 X > \frac{\hat{\mu}_2 - \hat{\mu}_1}{2\sigma^2} - \log\frac{n_2}{n_1}\left(\frac{n}{n-1}\right) \to \beta_0 + \hat{\beta}_1 X > C \to \beta_0 = C - \frac{n}{n-1}\left(\frac{\hat{\mu}_2 - \hat{\mu}_1}{2\sigma^2} - \log\frac{n_2}{n_1}\right)$$

8)

$$P(y=c|X,\theta) \propto \pi_c \mathcal{N}(x|\mu_c, \Sigma_c) \qquad \hat{\pi}_c = \frac{N_c}{N} \rightarrow \begin{cases} \hat{\pi}_1 = 1/2 \\ \hat{\pi}_2 = 1/2 \end{cases}$$

$$\hat{\mu}_c = \frac{1}{N_c} \sum_{n: y_n = c} x_n \rightarrow \begin{cases} \hat{\mu}_1 = \frac{1}{16} \sum \begin{pmatrix} 3\cos\theta_i \\ 3\sin\theta_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \hat{\mu}_2 = \frac{1}{16} \sum \begin{pmatrix} 3\cos\theta_i \\ 3\sin\theta_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{cases}$$

$$\hat{\Sigma}_c = \frac{1}{N_c} \sum_{n: y_n = c} (x_n - \hat{\mu}_c)(x_n - \hat{\mu}_c)^T = \frac{1}{16} \sum_{i=1}^{16} \begin{pmatrix} r\cos\theta_i \\ r\sin\theta_i \end{pmatrix} \begin{pmatrix} r\cos\theta_i & r\sin\theta_i \end{pmatrix}$$

$$= \begin{pmatrix} r^2/2 & 0 \\ 0 & r^2/2 \end{pmatrix} \longrightarrow \hat{\Sigma}_1 = \begin{pmatrix} 4.5 & 0 \\ 0 & 4.5 \end{pmatrix}, \hat{\Sigma}_2 = \begin{pmatrix} 12.5 & 0 \\ 0 & 12.5 \end{pmatrix}$$

QDA: $\log P(y=1|x,\theta) = -\log 2\pi - \log 4.5 + \log 1/2 + 1/9\, x_1^T x_1$

$$\log P(y=2|x,\theta) = -\log 2\pi - \log 12.5 + \log 1/2 + 1/25\, x_2^T x_2$$

$y=1$ if $\log P(y=1|x,\theta) > \log P(y=2|x,\theta) \rightarrow 25 x_1^T x_1 - 9 x_2^T x_2 > 25 \times 9 (2\log 9/25)$

9)

$$\frac{1}{N}|x_j^T y| = \lambda \qquad X\beta \overset{?}{=} y \to \hat{\beta} = (X^TX)^{-1}X^Ty \to \hat{y} = X(X^TX)^{-1}X^Ty$$

$$u(\alpha) = \alpha X\hat{\beta} = \alpha X(X^TX)^{-1}X^Ty \longrightarrow \frac{1}{N}|\langle x_j, y-u(\alpha)\rangle| = \frac{1}{N}|(x_j^T(y-\alpha X(X^TX)^{-1}X^Ty)|$$

$$\frac{1}{N}|\langle x_j, y-u(\alpha)\rangle| = \frac{1}{N}|(x_j^T(1-\alpha)y + \alpha x_j^T(y-X(X^TX)^{-1}X^Ty)|$$

$$X^T(y-X(X^TX)^{-1}X^Ty) = X^Ty - X^TX(X^TX)^{-1}X^Ty = X^Ty - X^Ty = 0$$

$$\to x_i^T(y-X(X^TX)^{-1}X^Ty) = 0 \longrightarrow \frac{1}{N}|\langle x_j, y-u(\alpha)\rangle| = \frac{1-\alpha}{N}|x_j^Ty| = \lambda(1-\alpha)$$

$$\langle y-u(\alpha), y-u(\alpha)\rangle = (y-\alpha X(X^TX)^{-1}X^Ty)^T(y-\alpha X(X^TX)^{-1}X^Ty)$$

$$= ((1-\alpha)y + \alpha(y-X(X^TX)^{-1}X^Ty))^T((1-\alpha)y+\alpha(y-X(X^TX)^{-1}X^Ty))$$

$$= (1-\alpha)^2y^Ty + 2\alpha(1-\alpha)(y^Ty - y^TX(X^TX)^{-1}X^Ty) + \alpha^2(y^Ty - 2y^TX(X^TX)^{-1}X^Ty +$$

$$y^TX(X^TX)^{-1}X^TX(X^TX)^{-1}Xy) = (1-\alpha)^2y^Ty + 2\alpha(1-\alpha)y^Ty + \alpha(\alpha-2)y^TX(X^TX)^{-1}X^Ty$$

$$\frac{1}{N}\langle y, y\rangle = 1 \to y^Ty = N$$

$$\frac{1}{N}\langle y-u(\alpha), y-u(\alpha)\rangle = (1-\alpha^2) + \alpha\,(2-\alpha) + \frac{\alpha(\alpha-2)\, y^T X(X^TX)^{-1}X^T y}{N}$$

$$(1-\alpha)^2 + \frac{\alpha(2-\alpha)}{N}\ RSS$$

$$RSS = (y - X(X^TX)^{-1}X^Ty)^T(y - X(X^TX)^{-1}X^Ty)$$

$$= y^Ty - 2y^TX(X^TX)^{-1}X^Ty + y^TX(X^TX)^{-2}X^TX(X^TX)^{-1}X^Ty = N - y^TX(X^TX)^{-2}X^Ty$$

$$\rightarrow (1-\alpha)^2 + \frac{\alpha(2-\alpha)}{N}\left(N - y^TX(X^TX)^{-2}X^Ty\right) = (1-\alpha)^2 + \alpha(2-\alpha) + \frac{\alpha(\alpha-2)\, y^TX(X^TX)X^Ty}{N}$$

$$\rightarrow \frac{1}{N}\langle y-u(\alpha), y-u(\alpha)\rangle = (1-\alpha)^2 + \frac{\alpha(2-\alpha)}{N}\ RSS$$

$$\lambda(\alpha) = \frac{\frac{1}{N}\left|\langle x_j, y-u(\alpha)\rangle\right|}{\sqrt{1/N\ \left|\langle y-u(\alpha), y-u(\alpha)\rangle\right|}} = \frac{1-\alpha}{\sqrt{(1-\alpha)^2 + \frac{\alpha(2-\alpha)}{N}\cdot RSS}}\cdot\lambda$$

10)

$$P(\beta|Y) \propto P(Y|\beta)P(\beta)$$

$$\rightarrow P(\beta|Y) \propto \frac{1}{|2\pi\tau^2 I|^{1/2}} \exp\left(-\frac{1}{2}\left(\beta^T(\tau^2 I)^{-1}\beta\right)\right) \frac{\exp\left(-\frac{1}{2}(y-X\beta)^T(\sigma^2 I)^{-1}(y-X\beta)\right)}{|2\pi\sigma^2 I|^{1/2}}$$

$$\underset{\beta}{\text{Argmax}}\ P(\beta|Y) = \underset{\beta}{\text{Argmin}} -\log P(\beta|Y) \propto \frac{1}{2}\frac{\beta^T\beta}{\tau^2} + \frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta)$$

$$\frac{\partial}{\partial\beta}\frac{1}{2}\frac{\beta^T\beta}{\tau^2} + \frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta) = \frac{\beta}{\tau^2} - \frac{X^T}{\sigma^2}(y-X\beta) = 0$$

$$\rightarrow \sigma^2\beta - \tau^2X^Ty + \tau^2X^TX\beta = 0 \rightarrow (\sigma^2I + \tau^2X^TX)\beta = \tau^2X^Ty$$

$$\rightarrow \hat{\beta}_{map} = (\sigma^2I + \tau^2X^TX)^{-1}\tau^2X^Ty = \left(X^TX + \frac{\sigma^2}{\tau^2}I\right)^{-1}X^Ty$$

$$\hat{W}_{map} = (X^TX + \lambda I)^{-1}X^Ty \rightarrow \lambda = \frac{\sigma^2}{\tau^2}$$

$$\tilde{X} = \begin{pmatrix} X \\ \sqrt{\lambda}I \end{pmatrix} \quad \tilde{y} = \begin{pmatrix} y \\ 0 \end{pmatrix}$$

برابر $\hat{\beta}_{map}$ است

$$\hat{\beta} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\tilde{y} = \left[(X^T\ \sqrt{\lambda}I)\begin{pmatrix} X \\ \sqrt{\lambda}I \end{pmatrix}\right]^{-1}(X^T\ \sqrt{\lambda}I)\begin{pmatrix} y \\ 0 \end{pmatrix} = \boxed{(X^TX + \lambda I)^{-1}(X^Ty)}$$

11-

1)

$$L(\mu|x) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \xrightarrow{\log} \sum_{i=1}^{N} -\frac{(x_i-\mu)^2}{2\sigma^2} \underbrace{-\frac{1}{2}\log(2\pi\sigma^2)}_{\text{constant}}$$

$$\rightarrow \frac{\partial \log L}{\partial \mu} = \sum_{i=1}^{N} \frac{+1}{2\sigma^2}(x_i-\mu) \rightarrow \hat{\mu}_{MLE} = \text{mean}(x)$$

2)

$$P(\mu|x) \propto P(x) P(\mu) \xrightarrow{\log} \frac{-1}{2\sigma^2} \sum_{i=1}^{N} -(x_i-\mu)^2 + \frac{1}{2\beta^2}(\mu+\mu_0)^2 + \text{Constant}$$

$$\rightarrow \hat{\mu}_{MAP} = \frac{\partial}{\partial \mu} \log(\mu|x) = \frac{-1}{\sigma^2} \sum_{i=1}^{N} (\mu - x_i) - \frac{1}{\beta^2}(\mu-\mu_0) = 0$$

$$\beta^2 \sum_{i=1}^{N} (\hat{\mu} - x_i) + \sigma^2(\hat{\mu} - \mu_0) = 0 \rightarrow \hat{\mu} = \frac{\beta^2}{\beta^2 N + \sigma^2} \sum_{i=1}^{N} x_i + \frac{\sigma^2}{\beta^2 N + \sigma^2} \mu_0$$

3)

1

$$\hat{\mu}_{MLE} = mean(x) = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$B\{\hat{\mu}_{MLE}\} = E\left\{\frac{1}{N}\sum_{i=1}^{N} x_i\right\} - \mu = \frac{1}{N}\sum_{i=1}^{N} \underbrace{E\{x\}}_{\mu} - \mu = \mu - \mu = 0 \rightarrow \text{converges to } \mu$$

2

$$\hat{\mu} = \frac{\beta^2}{\beta^2 N + \sigma^2}\sum_{i=1}^{N} x_i + \frac{\sigma^2}{\beta^2 N + \sigma^2}\mu_0$$

$$\lim_{N\to\infty} \frac{\beta^2}{\beta^2 N + \sigma^2}\sum_{i=1}^{N} x_i + \frac{\sigma^2}{\beta^2 N + \sigma^2}\mu_0 = \lim_{N\to\infty} \frac{\cancel{\beta^2}}{\cancel{\beta^2} N}\sum_{i=1}^{N} x_i = \mu$$

converges to μ as N increases Bias is zero.