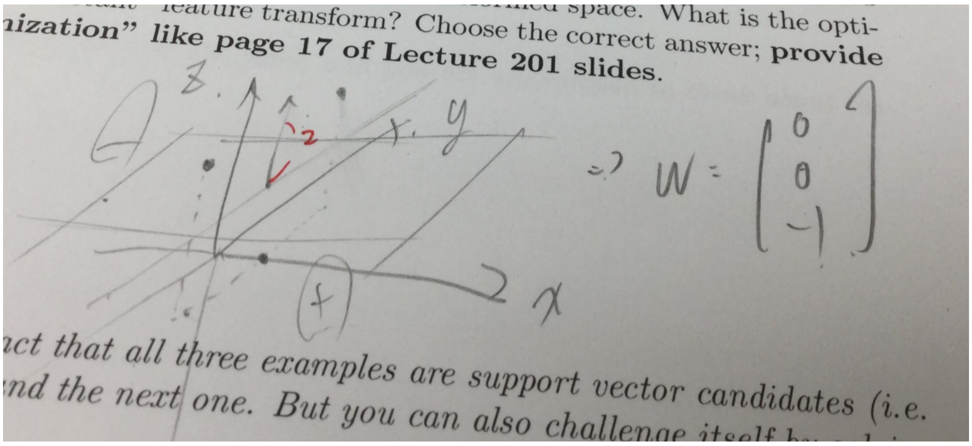
Machine Learning Foundations

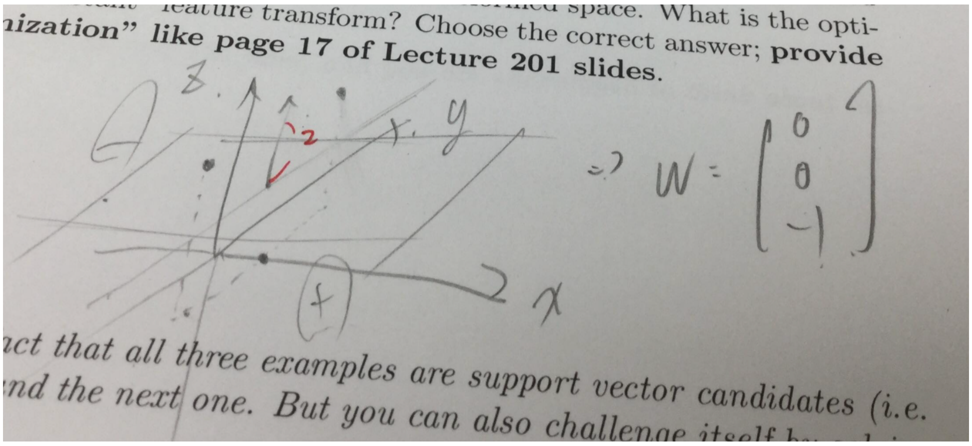
Homework 5

B05505009 謝惟遠

1. [e]



2. [b]



3. [e]

   The best separation line will be the center of segment $\overline{x_M x_{M+1}}$

4. [a]

$$P(|x_1 - x_2| \geq 2\rho) = P(x_1 - x_2 \geq 2\rho) + P(x_2 - x_1 \geq 2\rho)$$
$$= 2P(x_2 - x_1 \geq 2\rho) = 2\int_0^{1-2\rho}\int_{x_1+2\rho}^{1} dx_2\, dx_1$$
$$= 2\int_0^{1-2\rho}(1-2\rho-x_1)dx_1 = 2\left((1-2\rho)x_1 - \frac{1}{2}x_1^2\right)\Bigg|_0^{1-2\rho}$$
$$= 2(1-2\rho)x_1 - x_1^2\Bigg|_0^{1-2\rho} = (1-2\rho)^2$$
$$2P(|x_1 - x_2| < \rho) + 4P(|x_1 - x_2| \geq \rho) = 2 - 2(1-2\rho)^2 + 4(1-2\rho)^2$$
$$= 2 + 2(1-2\rho)^2$$

5. [c]

Karush-Kuhn-Tucker:

$$\mathcal{L}(b, \boldsymbol{w}, \boldsymbol{\mu}) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \sum_{i=1}^{n}\mu_i\big(\rho_+[\![y_i = 1]\!] + \rho_-[\![y_i = -1]\!] - y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)\big)$$
$$\nabla_{\boldsymbol{w}}\mathcal{L}(b, \boldsymbol{w}, \boldsymbol{\mu}) = \boldsymbol{w} - \sum_{i=1}^{n}\mu_i y_i \boldsymbol{x}_i$$
$$\nabla_{\boldsymbol{w}}\mathcal{L}(b^*, \boldsymbol{w}^*, \boldsymbol{\mu}) = \boldsymbol{w}^* - \sum_{i=1}^{n}\mu_i y_i \boldsymbol{x}_i = \boldsymbol{0}$$
$$\boldsymbol{w}^* = \sum_{i=1}^{n}\mu_i y_i \boldsymbol{x}_i$$
$$\mathcal{L}(b^*, \boldsymbol{w}^*, \boldsymbol{\mu}^*) = \frac{1}{2}\boldsymbol{w}^{*T}\boldsymbol{w}^*$$
$$+ \sum_{i=1}^{n}\mu_i^*\left(\rho_+[\![y_i = 1]\!] + \rho_-[\![y_i = -1]\!] - y_i\big(\boldsymbol{w}^{*T}\boldsymbol{x}_i + b\big)\right)$$
$$= \max_{\boldsymbol{\mu}}\left(\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mu_i y_i\, \mu_j y_j \boldsymbol{x}_j^T\boldsymbol{x}_i\right.$$
$$\left. + \sum_{i=1}^{n}\mu_i\left(\rho_+[\![y_i = 1]\!] + \rho_-[\![y_i = -1]\!] - y_i\big(\boldsymbol{w}^{*T}\boldsymbol{x}_i + b\big)\right)\right)$$
$$= \max_{\boldsymbol{\mu}}\left(\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mu_i\mu_j y_i\, y_j \boldsymbol{x}_j^T\boldsymbol{x}_i - \sum_{i=1}^{n} y_i\big(\boldsymbol{w}^{*T}\boldsymbol{x}_i + b\big)\right.$$
$$\left. + \sum_{i=1}^{n}\mu_i(\rho_+[\![y_i = 1]\!] + \rho_-[\![y_i = -1]\!])\right)$$

$$= \max_{\boldsymbol{\mu}} \left(-\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mu_i \mu_j y_i \, y_j \boldsymbol{x}_j^T \boldsymbol{x}_i \right.$$
$$\left. + \sum_{i=1}^{n} \mu_i (\rho_+ [\![y_i = 1]\!] + \rho_- [\![y_i = -1]\!]) \right)$$
$$= \min_{\boldsymbol{\mu}} \left(\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mu_i \mu_j y_i \, y_j \boldsymbol{x}_j^T \boldsymbol{x}_i - \sum_{i=1}^{n} \mu_i (\rho_+ [\![y_i = 1]\!] + \rho_- [\![y_i = -1]\!]) \right)$$

6. [e]

$$\nabla_{\boldsymbol{\mu}} \left(\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mu_i \mu_j y_i \, y_j \boldsymbol{x}_j^T \boldsymbol{x}_i - \sum_{i=1}^{n} f(i) \mu_i \right)$$
$$= \nabla_{\boldsymbol{\mu}} \left(\frac{1}{2} \boldsymbol{w}^{*T} \boldsymbol{w}^* - \sum_{i=1}^{n} f(i) \mu_i \right)$$
$$= \boldsymbol{w}^{*T} \nabla_{\boldsymbol{\mu}} \left(\sum_{i=1}^{n} \mu_i y_i \boldsymbol{x}_i \right) - \nabla_{\boldsymbol{\mu}} \left(\sum_{i=1}^{n} f(i) \mu_i \right)$$
$$= \boldsymbol{w}^{*T} \begin{bmatrix} | & | & & | \\ y_1 \boldsymbol{x}_1 & y_2 \boldsymbol{x}_2 & \dots & y_n \boldsymbol{x}_n \\ | & | & & | \end{bmatrix} - \begin{bmatrix} f(1) \\ f(2) \\ \vdots \\ f(n) \end{bmatrix}^T = \boldsymbol{0}$$

$$\sum_{i=1}^{n} \mu_i y_i \boldsymbol{x}_i^T \begin{bmatrix} | & | & & | \\ y_1 \boldsymbol{x}_1 & y_2 \boldsymbol{x}_2 & \dots & y_n \boldsymbol{x}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} f(1) \\ f(2) \\ \vdots \\ f(n) \end{bmatrix}^T$$

$$\sum_{i=1}^{n} \mu_i y_i \boldsymbol{x}_i^T \begin{bmatrix} | & | & & | \\ y_1 \boldsymbol{x}_1 & y_2 \boldsymbol{x}_2 & \dots & y_n \boldsymbol{x}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} f(1) \\ f(2) \\ \vdots \\ f(n) \end{bmatrix}^T$$

for original case and new case:

$$\begin{cases} \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i^T \begin{bmatrix} | & | & & | \\ y_1 \boldsymbol{x}_1 & y_2 \boldsymbol{x}_2 & \dots & y_n \boldsymbol{x}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}^T \\ \sum_{i=1}^{n} \mu_i y_i \boldsymbol{x}_i^T \begin{bmatrix} | & | & & | \\ y_1 \boldsymbol{x}_1 & y_2 \boldsymbol{x}_2 & \dots & y_n \boldsymbol{x}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \rho_+ [\![y_1 = 1]\!] + \rho_- [\![y_1 = -1]\!] \\ \rho_+ [\![y_2 = 1]\!] + \rho_- [\![y_2 = -1]\!] \\ \vdots \\ \rho_+ [\![y_n = 1]\!] + \rho_- [\![y_n = -1]\!] \end{bmatrix}^T \end{cases}$$

$$\begin{cases} \sum_{i=1}^{n}(\mu_i - \rho_+\alpha_i)y_i\boldsymbol{x}_i^T \begin{bmatrix} | & | & & | \\ y_1\boldsymbol{x}_1 & y_2\boldsymbol{x}_2 & \dots & y_n\boldsymbol{x}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} (\rho_- - \rho_+)[\![y_1 = -1]\!] \\ (\rho_- - \rho_+)[\![y_2 = -1]\!] \\ \vdots \\ (\rho_- - \rho_+)[\![y_n = -1]\!] \end{bmatrix}^T \\ \sum_{i=1}^{n}(\mu_i - \rho_-\alpha_i)y_i\boldsymbol{x}_i^T \begin{bmatrix} | & | & & | \\ y_1\boldsymbol{x}_1 & y_2\boldsymbol{x}_2 & \dots & y_n\boldsymbol{x}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} (\rho_+ - \rho_-)[\![y_1 = 1]\!] \\ (\rho_+ - \rho_-)[\![y_2 = 1]\!] \\ \vdots \\ (\rho_+ - \rho_-)[\![y_n = 1]\!] \end{bmatrix}^T \end{cases}$$

$$\begin{cases} \sum_{i=1}^{n}(\mu_i - \rho_+\alpha_i)y_i\boldsymbol{x}_i^T \sum_{i=1}^{n}\alpha_i y_i \boldsymbol{x}_1 = \begin{bmatrix} (\rho_- - \rho_+)[\![y_1 = -1]\!] \\ (\rho_- - \rho_+)[\![y_2 = -1]\!] \\ \vdots \\ (\rho_- - \rho_+)[\![y_n = -1]\!] \end{bmatrix}^T \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \\ \sum_{i=1}^{n}(\mu_i - \rho_-\alpha_i)y_i\boldsymbol{x}_i^T \sum_{i=1}^{n}\alpha_i y_i \boldsymbol{x}_1 = \begin{bmatrix} (\rho_+ - \rho_-)[\![y_1 = 1]\!] \\ (\rho_+ - \rho_-)[\![y_2 = 1]\!] \\ \vdots \\ (\rho_+ - \rho_-)[\![y_n = 1]\!] \end{bmatrix}^T \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \end{cases}$$

$$\sum_{i=1}^{n}(2\mu_i - \rho_+\alpha_i - \rho_-\alpha_i)y_i\boldsymbol{x}_i^T \sum_{i=1}^{n}\alpha_i y_i \boldsymbol{x}_1 = (\rho_+ - \rho_-)\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_2 \end{bmatrix}^T \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$

given $\boldsymbol{\alpha}$ is optimal:

$$\sum_{j=1}^{n}\alpha_j y_j = 0$$

$$\sum_{i=1}^{n}(2\mu_i - \rho_+\alpha_i - \rho_-\alpha_i)y_i\boldsymbol{x}_i^T \sum_{i=1}^{n}\alpha_i y_i \boldsymbol{x}_1 = 0$$

$$2\mu_i = (\rho_+ + \rho_-)\alpha_i$$

7. [d]
   $log\ 0$ is not defined

8. [b]

$$exp(x) \in (0, 1]$$

9. [d]

Consider worst case, where every other point has the opposite label and is at a distance of ϵ:

$$1-\sum_{i\neq n} exp(-\gamma\epsilon^2) > 0$$
$$1 > (N-1)exp(-\gamma\epsilon^2)$$
$$exp(\gamma\epsilon^2) > (N-1)$$
$$\gamma\epsilon^2 > log(N-1)$$
$$\gamma > \frac{log(N-1)}{\epsilon^2}$$

10. [c]

$$\boldsymbol{w}_{t+1} = \boldsymbol{w}_t + y_{n(t)}\phi(\boldsymbol{x}_n)$$
$$a_{t+1,n} = y_{n(t)}$$

11. [a]

$$\boldsymbol{w}_t = \sum_{n=1}^{N} a_{t,n}\phi(\boldsymbol{x}_n)$$
$$\boldsymbol{w}_t^T\phi(\boldsymbol{x}) = \sum_{n=1}^{N} a_{t,n}\phi(\boldsymbol{x}_n)^T\phi(\boldsymbol{x}) = \sum_{n=1}^{N} a_{t,n}K(\boldsymbol{x}_n, \boldsymbol{x})$$

12. [b]

$$b = y_n - y_n\xi_n - w^T z_n$$

If $y_n = 1$,

$$b = 1 - \xi_n - w^T z_n$$
$$b \leq 1 - w^T z_n$$

If $y_n = -1$,

$$b = -1 + \xi_n - w^T z_n$$
$$b \geq -1 - w^T z_n$$

Upper bound at $y_n = 1$, thus

$$\min_{y_n>0}(1 - w^T z_n) \geq b$$

13. [e]

$$\mathcal{L}(b,\boldsymbol{w},\boldsymbol{\mu},\boldsymbol{\xi})=\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+C\sum_{i=1}^{N}\xi_i^2+\sum_{i=1}^{N}\mu_i\big(1-\xi_i-y_i(\boldsymbol{w}^T\boldsymbol{z}_i+b)\big)$$

$$\nabla_{\boldsymbol{w}}\mathcal{L}(b,\boldsymbol{w},\boldsymbol{\mu},\boldsymbol{\xi})=\boldsymbol{w}-\sum_{n=1}^{N}y_n\mu_n\nabla_{\boldsymbol{w}}(\boldsymbol{w}^T\boldsymbol{z}_n+b)=\boldsymbol{w}-\sum_{n=1}^{N}y_n\mu_n\boldsymbol{z}_n$$

$$\boldsymbol{w}^*=\sum_{n=1}^{N}y_n\mu_n\boldsymbol{z}_n$$

$$\nabla_{\boldsymbol{\mu}}\mathcal{L}(b,\boldsymbol{w},\boldsymbol{\mu},\boldsymbol{\xi})=\begin{bmatrix}1-\xi_1-y_1(\boldsymbol{w}^T\boldsymbol{z}_1+b)\\1-\xi_2-y_2(\boldsymbol{w}^T\boldsymbol{z}_2+b)\\\vdots\\1-\xi_n-y_n(\boldsymbol{w}^T\boldsymbol{z}_n+b)\end{bmatrix}$$

$$\nabla_{\boldsymbol{\xi}}\mathcal{L}(b,\boldsymbol{w},\boldsymbol{\mu},\boldsymbol{\xi})=2C\boldsymbol{\xi}-\boldsymbol{\mu}$$

$$\boldsymbol{\xi}^*=\frac{\boldsymbol{\mu}}{2C}$$

$$\mathcal{L}(\boldsymbol{\mu})=\frac{1}{2}\boldsymbol{w}^{*T}\boldsymbol{w}^*+C\sum_{i=1}^{N}\xi_i^{*2}+\sum_{i=1}^{N}y_n\mu_n\big(1-\xi_i^*-y_i(\boldsymbol{w}^{*T}\boldsymbol{z}_i+b)\big)$$

$$=\frac{1}{2}\boldsymbol{w}^{*T}\boldsymbol{w}^*+\frac{\boldsymbol{\mu}^T\boldsymbol{\mu}}{4C}+\sum_{i=1}^{N}\mu_i\left(1-\frac{\mu_i}{2C}-y_i\left(\sum_{n=1}^{N}y_n\mu_n\boldsymbol{z}_n^T\,\boldsymbol{z}_i\right)\right)$$

$$=\frac{1}{2}\boldsymbol{w}^{*T}\boldsymbol{w}^*+\frac{\boldsymbol{\mu}^T\boldsymbol{\mu}}{4C}+\sum_{i=1}^{N}\mu_i\left(1-\frac{\mu_i}{2C}-y_i\left(\sum_{n=1}^{N}y_n\mu_n\boldsymbol{z}_n^T\,\boldsymbol{z}_i\right)\right)$$

$$=\frac{1}{2}\boldsymbol{w}^{*T}\boldsymbol{w}^*+\frac{\boldsymbol{\mu}^T\boldsymbol{\mu}}{4C}+\sum_{i=1}^{N}\mu_i-\frac{\boldsymbol{\mu}^T\boldsymbol{\mu}}{2C}-\sum_{i=1}^{N}\mu_iy_i\left(\sum_{n=1}^{N}y_n\mu_n\boldsymbol{z}_n^T\,\boldsymbol{z}_i\right)$$

$$=\frac{1}{2}\boldsymbol{w}^{*T}\boldsymbol{w}^*-\frac{\boldsymbol{\mu}^T\boldsymbol{\mu}}{4C}+\sum_{i=1}^{N}\mu_i-\sum_{i=1}^{N}\sum_{n=1}^{N}\mu_i\mu_ny_iy_nK(\boldsymbol{x}_n,\boldsymbol{x}_i)$$

$$=\sum_{i=1}^{N}\mu_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{n=1}^{N}\left(\mu_i\mu_ny_iy_nK(\boldsymbol{x}_n,\boldsymbol{x}_i)+\frac{\mu_i\mu_n}{2C}[\![i=n]\!]\right)$$

14. [e]

$$\mathcal{L}(b,\boldsymbol{w},\boldsymbol{\mu},\boldsymbol{\xi})=\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+C\sum_{i=1}^{N}\xi_i^2+\sum_{i=1}^{N}\mu_i\big(1-\xi_i-y_i(\boldsymbol{w}^T\boldsymbol{z}_i+b)\big)$$

$$\nabla_{\boldsymbol{\xi}}\mathcal{L}(b,\boldsymbol{w},\boldsymbol{\mu},\boldsymbol{\xi})=2C\boldsymbol{\xi}-\boldsymbol{\mu}$$

$$\boldsymbol{\xi}^*=\frac{\boldsymbol{\mu}}{2C}$$