

# ML HW5

B06303126 Lo Yun Chien

December 2020

## Question 1

Correct answer: (d)

After transformation, we yield three inequality,

$$\begin{aligned} 2w_1 - 4w_2 - b &\geq 1 \\ b &\geq 1 \\ -2w_1 - 4w_2 - b &\geq 1 \end{aligned}$$

We want to minimize  $\mathbf{w}^T \mathbf{w}$ , we should let  $2w_1 + 4w_2$  as small as possible, then we want  $b = 1$ . For the last two inequality, after plotting it on 2D plane, the smallest circle centered at 0 intersecting with the area would be  $(0, \frac{1}{2})$ . Therefore  $w_1 = 0$ .

## Question 2

Correct answer: (c)

By definition, margin is  $\frac{1}{\|\mathbf{w}\|}$ , with  $\mathbf{w} = (0, \frac{1}{2})$ , we should have margin is 4.

## Question 3

Correct answer: (e)

For this kind of data, only points in  $(x_M, x_{M+1})$  can separate the data. The margin is  $\min\{x^* - x_M, x_{M+1} - x^*\}$ , then to max this value, the biggest margin is  $\frac{1}{2}(x_{M+1} - x_M)$

## Question 4

Correct answer: (a)

For  $\mathbf{y} = (1, 1)$  or  $\mathbf{y} = (-1, -1)$ , 1D perceptron can always produce these dichotomy by setting cutting point at -0.5. So the expected dichotomies should be

$$E[Dichotomy] = 2 + 2P(|x_1 - x_2| \geq 2\rho)$$

To calculate  $P(|x_1 - x_2| \geq 2\rho)$ , note that the graph of  $|x_1 - x_2| \geq 2\rho$  intersect with  $[0, 1] \times [0, 1]$  on 2D plane yield two triangles, each with area  $\frac{1}{2}(1 - 2\rho)^2$ . Since it's uniform distribution, hence  $P(|x_1 - x_2| \geq 2\rho) = (1 - 2\rho)^2$ .

## Question 5

Correct answer:(c)

Using Lagrange method, define a new function as

$$\frac{1}{2} \mathbf{w}^T \mathbf{w} + \sum_{i=1}^N \alpha_i (\rho^+ - y_i (\mathbf{w}^T \mathbf{x}_i + b)) [y_i = 1] + \sum_{i=1}^N \alpha_i (\rho^- - y_i (\mathbf{w}^T \mathbf{x}_i + b)) [y_i = -1]$$

with  $\alpha_i \geq 0$  for all  $i$ . Solve the dual problem with these equation, take partial derivative about  $\mathbf{w}, b$ . we can find  $\sum_{i=1}^N \alpha_i y_i = b$ , and  $\sum_{i=1}^N \alpha_i y_i x_i = \mathbf{w}$  when maxing it. Therefore the remaining term is

$$-\frac{1}{2} \left\| \sum_{i=1}^N \alpha_i y_i x_i \right\|^2 + \sum_{i=1}^N \alpha_i \rho^+ [y_i = 1] + \sum_{i=1}^N \alpha_i \rho^- [y_i = -1]$$

When taking opposite, the result is (c)

## Question 6

Correct answer:(e)

By question 5, we are solving

$$\frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \alpha_n \alpha_m y_n y_m x_n^T x_m - \sum_{n=1}^N \rho_+ [y_n = 1] \alpha_n - \sum_{n=1}^N \rho_- [y_n = -1] \alpha_n$$

By using the constraint  $\sum_{n=1}^N y_n \alpha_n = 0$ , we have  $\sum_{n=1}^N [y_n = 1] \alpha_n = \sum_{n=1}^N [y_n = -1] \alpha_n = \frac{1}{2} \sum_{n=1}^N \alpha_n$ . Using this fact to rewrite the function,

$$\frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \alpha_n \alpha_m y_n y_m x_n^T x_m - \frac{1}{2} \sum_{n=1}^N \alpha_n (\rho_+ + \rho_-)$$

Define  $\alpha'_n = \alpha_n \frac{\rho_+ + \rho_-}{2}$  and put it into the function

$$\frac{1}{2} \left( \frac{\rho_+ + \rho_-}{2} \right)^2 \sum_{n=1}^N \sum_{m=1}^N \alpha'_n \alpha'_m y_n y_m x_n^T x_m - \left( \frac{\rho_+ + \rho_-}{2} \right)^2 \sum_{n=1}^N \alpha'_n$$

Since all the constraints won't change by simply adding a factor, this new problem is the original hard margin SVM, Which solution is  $\alpha^*$ . Therefore, by  $\alpha_n = \alpha'_n \frac{\rho_+ + \rho_-}{2}$ , the uneven SVM solution is (e)

## Question 7

Correct answer: (d)

While setting  $K = 0$ ,  $\log_2 0$  is undefined.

## Question 8

Correct answer: (c)

$$\begin{aligned}\|\phi(x) - \phi(x')\| &= \phi(x)^T \phi(x) - 2\phi(x')^T \phi(x) + \phi(x')^T \phi(x') \\ &= K(x, x) - 2K(x', x) + K(x', x') \\ &= 2 - 2e^{-\gamma \|x-x'\|}\end{aligned}$$

We want to maximize  $\|\phi(x) - \phi(x')\|$ , so takes  $\|x' - x\|$  to infinity, then the answer is 2.

## Question 9

Correct answer: (d)

We have

$$h_{1,0}(x) = \text{sign}(\sum_{i=1}^N y_i e^{-\gamma \|x_i - x\|})$$

Since we want  $E_{in} = 0$ , consider an example  $(x_0, y_0)$  with  $y_0 = 1$ . We need  $h_{1,0}(x_0) = 1$  i.e.  $\sum_{i=1}^N y_i e^{-\gamma \|x_i - x_0\|} > 0$ . Worst case is that all the other  $y_n = -1$ , adding that when  $n = 0$ , the term is 1, we yield

$$\sum_{n \neq 0} -e^{-\gamma \|x_n - x_0\|} > -1$$

With all  $\|x_n - x_0\| \geq \epsilon^2$ ,  $\sum_{n \neq 0} -e^{-\gamma \|x_n - x_0\|} \geq \sum_{n \neq 0} -e^{-\gamma \epsilon^2}$ ,  $(N-1)e^{-\gamma \epsilon^2} \geq -1$ ,  $(N-1)e^{-\gamma \epsilon^2} \leq 1$ . Solve this inequality, we have

$$\gamma \geq \frac{\ln(N-1)}{\epsilon^2}$$

## Question 10

Correct answer: (c)

$w_t = \sum_{n=1}^N \alpha_{t,n} \phi(x_n)$ , using this equation, we have

$$w_{t+1} = \sum_{n=1}^N \alpha_{t,n} \phi(x_n) + y_{n(t)} \phi(x_n)$$

As we can see,  $\alpha_{t+1} = \alpha_t$ . except for index n will add  $y_n$ .

## Question 11

Correct answer: (a)

$w_t^T \phi(x) = \sum_{n=1}^N \alpha_{t,n} \phi(x_n)^T \phi(x)$ , by using  $w_t = \sum_{n=1}^N \alpha_{t,n} \phi(x_n)$ . Therefore, kernel trick imply  $\sum_{n=1}^N \alpha_{t,n} \phi(x_n)^T \phi(x) = \sum_{n=1}^N \alpha_{t,n} K(x_n, x)$

## Question 12

Correct answer: (b)

Using KKT condition, we yield for  $\alpha_n > 0, b = y_n - y_n \xi_n - w^T z_n$ . represent it as  $\xi_n = \frac{b - y_n w^T z_n}{-y_n}$ , since  $\xi_n \geq 0$ , consider  $y_n = 1$ , we yield  $b \leq 1 - w^T z_n$ , for  $y_n = -1$ , we yield  $b \geq -1 - w^T z_s$ . To maximize b, we choose b satisfying all n constraint and is the largest, which is the lowest upper bound when  $y_n = 1$ .

## Question 13

Correct answer: (e)

Using Lagrange method, we want to minimize the below value

$$L(w, b, \xi) = \frac{1}{2} w^T w + C \sum_{n=1}^N \xi_n^2 + \sum_{n=1}^N \alpha_n (1 - \xi_n - y_n (w^T z_n + b))$$

Where  $\frac{\partial L}{\partial \xi_i} = 2C\xi_i - \alpha_i = 0$ , substitute  $\xi$  for  $\alpha$

$$\frac{1}{2} w^T w + C \sum_{n=1}^N \frac{\alpha_n^2}{(2C)^2} + \sum_{n=1}^N \alpha_n \left(1 - \frac{\alpha_n}{2C} - y_n (w^T z_n + b)\right)$$

Using the same method calculate  $\frac{\partial L}{\partial w} = 0$  and  $\frac{\partial L}{\partial b} = 0$  should have  $\sum_{n=1}^N \alpha_n y_n = 0$  and  $w = \sum \alpha_n y_n z_n$  separately, finally we have

$$-\frac{1}{2} w^T w - \sum_{n=1}^N \frac{\alpha_n^2}{4C} + \sum_{n=1}^N \alpha_n$$

Notice that  $\sum_{n=1}^N \frac{\alpha_n^2}{4C}$  only affect diagonal entry, and the coefficient is  $\frac{1}{2C}$ .

## Question 14

Correct answer: (e)

Notice that the last maximization problem is solved under constraint  $2C\xi_i - \alpha_i = 0$ , then the answer is (e).