# COL341 Spring 2023
# Homework 2
# (To be done Individually)

Due Date: 19th March 2023, Friday, 11:55 PM (No extensions)

## Instructions

Type the solutions in LaTeX(you may use Overleaf for ease of use). Submit the `.tex` source and the compiled `pdf` in a single `.zip` file in Moodle. Name the file as `<your-entry-number>.zip`, e.g. `2019CSZ8406.zip`. The homework is to be done individually. Plagiarism and academic dishonesty will be penalized as per the course policy. No deadline extension will be provided.

## Question 1 $[1 \times 10 = 10$ **marks**$]$

The SVM hard margin formulation assumes that the data is linearly separable and tries to draw the decision boundary with maximum margin so that the generalization error is less. In class, we have seen the primal and corresponding dual problem in this scenario.

The primal problem corresponding to the separable case:

$$\min_{\boldsymbol{w},b} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} \qquad \text{s.t} \qquad y_n(\boldsymbol{w}^T\boldsymbol{x}_n + b) \geq 1 \qquad\qquad (n = 1, \ldots N)$$

The corresponding dual problem is

$$\min_{\alpha \in \mathcal{R}^N} \frac{1}{2}\sum_{m=1}^{N}\sum_{n=1}^{N} y_n y_m \alpha_n \alpha_m \boldsymbol{x}_n^T \boldsymbol{x}_m - \sum_{n=1}^{N}\alpha_n$$

$$\text{s.t} \qquad \sum_{n=1}^{N} y_n\alpha_n = 0 \qquad \text{and} \quad \alpha_n \geq 0 \qquad (n = 1, \ldots N)$$

However, in practice, the training data is not linearly separable, and we need to introduce slack variables to handle the noise. Again, as discussed in the class, the primal problem corresponding to the soft SVM is:

$$\min_{\boldsymbol{w},b} \; \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{n=1}^{N}\zeta_n$$

$$\text{s.t} \qquad y_n(\boldsymbol{w}^T\boldsymbol{x}_n + b) \geq 1 - \zeta_n \quad \text{and} \quad \zeta_n \geq 0 \qquad (n = 1, \dots N)$$

**The task in this homework is to derive the dual of the soft SVM**, which is:

$$\min_{\alpha\in\mathcal{R}^N} \; \frac{1}{2}\sum_{m=1}^{N}\sum_{n=1}^{N} y_n y_m \alpha_n \alpha_m \boldsymbol{x}_n^T\boldsymbol{x}_m - \sum_{n=1}^{N}\alpha_n$$

$$\text{s.t} \qquad \sum_{n=1}^{N} y_n\alpha_n = 0 \qquad \text{and} \quad C \geq \alpha_n \geq 0 \qquad (n = 1, \dots N)$$

[Hint: Find the Lagrangian of the primal problem. Use the KKT conditions to get certain relationships among the variables and substitute these into the Lagrangian to get the dual.]

# Question 2 [1+2+3+2+1=10 marks]

In this question, we will find an upper bound of the VC-dimension of SVM. Suppose the input space is the ball of radius $R$ in $\mathcal{R}^d$, so $||x|| \leq R$. Then,

$$d_{VC}(\rho) \leq \left\lceil R^2/\rho^2 \right\rceil + 1 \tag{1}$$

where $\left\lceil R^2/\rho^2 \right\rceil$ is the smallest integer greater than or equal to $R^2/\rho^2$.

To start the proof, fix $x_1, \dots, x_n$ that are shattered by hyperplanes with margin $\rho$. We will show when $N$ is even, $N \leq R^2/\rho^2 + 1$. When $N$ is odd, a similar analysis can show that $N \leq R^2/\rho^2 + 1 + \frac{1}{N}$. In both cases, $N \leq \left\lceil R^2/\rho^2 \right\rceil + 1$.

## Proof sketch (N is even)

Suppose you randomly select $N/2$ of the labels $y_1, \dots, y_n$ to be +1, the others being -1. By construction, $\sum_{n=1}^{N} y_n = 0$.

1. Show $\left\|\sum_{n=1}^{N} y_n\boldsymbol{x}_n\right\|^2 = \sum_{n=1}^{N}\sum_{m=1}^{N} y_n y_m \boldsymbol{x}_n^T\boldsymbol{x}_m$

2. When $n = m$, what is $y_n y_m$? Show that $\mathbb{P}[y_n y_m = 1] = (\frac{N}{2} - 1)/(N-1)$ when $n \neq m$. Hence show that

$$\mathbb{E}[y_n y_m] = \begin{cases} 1 & m = n; \\ -\frac{1}{N-1} & m \neq n. \end{cases}$$

3. Show that

$$\mathbb{E}\left[\left\|\sum_{n=1}^{N} y_n \boldsymbol{x_n}\right\|^2\right] = \frac{N}{N-1}\sum_{n=1}^{N} \|\boldsymbol{x}_n - \bar{\boldsymbol{x}}\|^2,$$

where the average vector $\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n$. [Hint: Use linearity of expectation in (1), and consider the case $m = n$ and $m \neq n$ separately.]

4. Show that $\sum_{n=1}^{N} \|\boldsymbol{x}_n - \bar{\boldsymbol{x}}\|^2 \leq \sum_{n=1}^{N} \|\boldsymbol{x}_n\|^2 \leq NR^2$

$$\text{[Hint: } \sum_{n=1}^{N} \|\boldsymbol{x}_n - \boldsymbol{\mu}\|^2 \text{ is minimized at } \boldsymbol{\mu} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n\text{]}$$

5. Conclude that

$$\mathbb{E}\left[\left\|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right\|^2\right] \leq \frac{N^2R^2}{N-1},$$

and hence that

$$\mathbb{P}\left[\left\|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right\| \leq \frac{NR}{\sqrt{N-1}}\right] > 0$$

This means for some choice of $y_n$, $\left\|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right\| \leq NR/\sqrt{N-1}$

This proof is called a *probabilistic existence proof: if some random process can generate an object with positive probability, then that object must exist. Note that you prove the existence of the required dichotomy without actually constructing it. In this case, the easiest way to construct a desired dichotomy is to randomly generate the balanced dichotomies until you have one that works*

Use the above **proof of sketch** to show that there exist a balanced dichotomy $y_1, \ldots, y_n$ s.t.

$$\sum_{n=1}^{N} y_n = 0, \text{and} \left\|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right\| \leq \frac{NR}{\sqrt{N-1}}$$

## VC dimension upper bound (2 marks)

As these $N$ points are being shattered, they can be separated by the SVM with margin at least $\rho$. So, for some $W$ and $b$, we have

$$\rho\|\boldsymbol{w}\| \leq y_n(\boldsymbol{w}^T\boldsymbol{x}_n + b)\forall n$$

Then, use Cauchy- Schwartz inequality to prove the given upper bound for VC dimension.

1) $\min_{b,w,\xi} \frac{1}{2} w^T w + C \sum_{n=1}^{N} \xi_n$ ①

s.t $y_n(w^T x_n + b) \geq 1 - \xi_n$ & $\xi_n \geq 0$

$$\mathcal{L}(w, b, \xi, \alpha, \beta) = \min_{b,w,\xi} \frac{1}{2} w^T w + C \sum_{n=1}^{N} \xi_n + \sum_{n=1}^{N} \alpha_n (1 - \xi_n - y_n(w^T x_n + b)) - \sum_{n=1}^{N} \beta_n \xi_n, \quad \alpha_n \geq 0, \ \beta_n \geq 0$$

Find the dual :-

$$\frac{\partial \mathcal{L}}{\partial w} = w - \sum_{n=1}^{N} \alpha_n y_n x_n = 0$$

$$\Rightarrow w = \sum_{n=1}^{N} \alpha_n y_n x_n$$

$$\frac{\partial \mathcal{L}}{\partial b} = \sum_{n=1}^{N} \alpha_n y_n = 0$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i} = C + \alpha_i - \beta_i = 0$$

Note $C - \beta_i = \alpha_i$

since $\beta_i \geq 0$

$\alpha_i \leq C. \Rightarrow 0 \leq \alpha_i \leq C.$

Let's use these constraints to find the dual.

$$\mathcal{L}(\alpha, \beta) = \frac{1}{2} w^T w + \sum_{n=1}^{N} \cancelto{0}{(C + \alpha_n - \beta_n)} \xi_n + b \cancelto{0}{\sum \alpha_n y_n} - w^T \sum \alpha_n y_n x_n + \sum \alpha_n$$

$$= \frac{1}{2} \left(\sum_{n=1}^{N} \alpha_n y_n x_n\right)^T \left(\sum \alpha_n y_n x_n\right) - \left(\sum \alpha_n y_n x_n\right)^T \left(\sum \alpha_n y_n x_n\right) + \sum \alpha_n$$

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

So, the dual problem is

$$\mathcal{L}(\alpha, \beta) = \max_{\alpha \in \mathbb{R}} \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{m,n=1}^{N} \alpha_n \alpha_m y_n y_m x_n^T x_m$$

$$\text{s.t} \sum_{n=1}^{N} y_n \alpha_n = 0 \quad \& \quad C \geq \alpha_n \geq 0$$

## Answer to Question2

1.

$$\begin{aligned}\left\|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right\|^2 &= \left(\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right)^T \left(\sum_{n=1}^{M} y_m \boldsymbol{x}_m\right) \\ &= \left(\sum_{n=1}^{N} y_n \boldsymbol{x}_n^T\right) \left(\sum_{n=1}^{M} y_m \boldsymbol{x}_m\right) \\ &= \sum_{n=1}^{N} \sum_{m=1}^{M} y_n y_m \boldsymbol{x}_n^T \boldsymbol{x}_m\end{aligned}$$

2. When,

$$n = m \Rightarrow y_n y_m = 1 \Rightarrow P(y_n y_m = 1) = 1 \Rightarrow \mathbb{E}(y_n y_m = 1) = 1 \tag{1}$$

When, $n \neq m$

$$\begin{aligned}\mathbb{P}(y_n y_m = 1) &= \mathbb{P}(y_n = +1, y_m = +1) + \mathbb{P}(y_n = -1, y_n = -1) \\ &= 2 \times \frac{\binom{N/2}{2}}{\binom{N}{2}} \\ &= \frac{(N/2) - 1}{N - 1}\end{aligned}$$

Similarly, $\mathbb{P}(y_n y_m = -1) = 1 - \frac{(N/2)-1}{N-1} = \frac{(N/2)}{N-1}$

Now,

$$\begin{aligned}\mathbb{E}[y_n y_m] &= 1 \times \mathbb{P}(y_n y_m = 1) + (-1) \times \mathbb{P}(y_n y_m = -1) \\ &= \frac{(N/2) - 1}{N - 1} - \frac{(N/2)}{N - 1} \\ &= \frac{N - 2}{2(N - 1)} - \frac{N}{2(N - 1)} \\ &= -\frac{1}{N - 1}\end{aligned} \tag{2}$$

Combining both equation 1 and equation 2, we have,

$$\mathbb{E}[y_n y_m] = \begin{cases} 1 & m = n; \\ -\frac{1}{N-1} & m \neq n. \end{cases}$$

3.

$$\mathrm{E}\left[\left\|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right\|^2\right] = \mathbb{E}\left[\sum_{n=1}^{N}\sum_{m=1}^{N} y_n y_m \boldsymbol{x}_n^T \boldsymbol{x}_m\right]$$
$$= \sum_{n=1}^{N}\sum_{m=1}^{N} \mathbb{E}\left[y_n y_m \boldsymbol{x}_n^T \boldsymbol{x}_m\right]$$
$$= \sum_{n=1}^{N}\sum_{m=1}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_m \mathbb{E}\left[y_n y_m\right]$$

$$=\sum_{\substack{m,n=1\\ m=n}}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_m \mathbb{E}\left[y_n y_m\right] + \sum_{\substack{m,n=1\\ m\neq n}}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_m \mathbb{E}\left[y_n y_m\right]$$
$$= \sum_{\substack{m,n=1\\ m=n}}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_m - \frac{1}{N-1}\sum_{\substack{m,n=1\\ m\neq n}}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_m$$
$$= \sum_{\substack{m,n=1\\ m=n}}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_m - \left(\frac{-1}{N-1}\right)\sum_{\substack{m,n=1\\ m=n}}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_m + \left(\frac{-1}{N-1}\right)\sum_{\substack{m,n=1\\ m=n}}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_m - \frac{1}{N-1}\sum_{\substack{m,n=1\\ m\neq n}}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_m$$
$$= \frac{N}{N-1}\sum_{\substack{m,n=1\\ m=n}}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_m - \frac{1}{N-1}\sum_{m,n=1}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_m$$
$$= \frac{N}{N-1}\left[\sum_{n=1}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_n - \frac{1}{N}\sum_{m,n=1}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_m\right]$$
$$= \frac{N}{N-1}\left[\sum_{n=1}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_n - \sum_{n=1}^{N} \boldsymbol{x}_n^T \bar{\boldsymbol{x}}\right]$$
$$= \frac{N}{N-1}\left[\sum_{n=1}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_n - N\bar{\boldsymbol{x}}^T \bar{\boldsymbol{x}}\right]$$
$$= \frac{N}{N-1}\left[\sum_{n=1}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_n - 2N\bar{\boldsymbol{x}}^T \bar{\boldsymbol{x}} + N\bar{\boldsymbol{x}}^T \bar{\boldsymbol{x}}\right]$$
$$= \frac{N}{N-1}\left[\sum_{n=1}^{N} \boldsymbol{x}_n^T \boldsymbol{x}_n - 2\bar{\boldsymbol{x}}^T \sum_{n=1}^{N} \boldsymbol{x}_n + N\bar{\boldsymbol{x}}^T \bar{\boldsymbol{x}}\right]$$
$$= \frac{N}{N-1}\sum_{n=1}^{N}\left[\boldsymbol{x}_n^T \boldsymbol{x}_n - 2\bar{\boldsymbol{x}}^T \boldsymbol{x}_n + \bar{\boldsymbol{x}}^T \bar{\boldsymbol{x}}\right]$$
$$= \frac{N}{N-1}\sum_{n=1}^{N}\left[\boldsymbol{x}_n^T \boldsymbol{x}_n - \boldsymbol{x}_n^T \bar{\boldsymbol{x}} - \bar{\boldsymbol{x}}^T \boldsymbol{x}_n + \bar{\boldsymbol{x}}^T \bar{\boldsymbol{x}}\right]$$
$$= \sum_{n=1}^{N}(\boldsymbol{x}_n^T - \bar{\boldsymbol{x}}^T)(\boldsymbol{x}_n - \bar{\boldsymbol{x}})$$
$$= \sum_{n=1}^{N}\left\|\boldsymbol{x}_n - \bar{\boldsymbol{x}}\right\|^2$$

Part 3 and 4 are is easy to show.

4. $\sum_{n=1}^{N} \|x_n - \mu\|^2$ is minimized at $\mu = \bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$

$$\Rightarrow \sum_{n=1}^{N} \|x_n - \bar{x}\|^2 \le \sum_{n=1}^{N} \|x_n - 0\|^2$$
$$\le NR^2$$

( $\because$ since the inputs are present inside a ball of radius R)

5. $$E\left[\left\|\sum_{n=1}^{N} y_n x_n\right\|^2\right] \le \frac{N^2 R^2}{N-1}$$

consider, $Z = \left\|\sum_{n=1}^{N} y_n x_n\right\|^2$, $a = \frac{N^2R^2}{N-1}$

& use Markov inequality

For any non-negative random variable Z, we have
$$P(Z \ge a) \le \frac{E(Z)}{a} \qquad a > 0$$

$$P(Z \ge a) \le \frac{E(Z)}{a}$$
$$\Rightarrow 1 - P(Z \le a) \le \frac{E(Z)}{a}$$
$$\Rightarrow P(Z \le a) \ge 1 - \frac{E(Z)}{a}.$$

So, $$P\left(\left\|\sum_{n=1}^{N} y_n x_n\right\|^2 \le \frac{N^2R^2}{N-1}\right) \ge \underbrace{1 - \frac{E\left[\left\|\sum_{n=1}^{N} y_n x_n\right\|^2\right]}{\left(N^2R^2/N-1\right)}}_{\ge 0}$$

$$\Rightarrow P\left(\left\|\sum_{n=1}^{N} y_n x_n\right\| \le \frac{NR}{\sqrt{N-1}}\right) > 0.$$

As these N points are being shattered, they can be separated by the SVM with margin at least $\rho$.

So, for some $W$ & $b$.

$$\rho \|w\| \le y_n(w^T x_n + b) \quad \forall n$$
$$\Rightarrow \sum_{n=1}^{N} \rho \|w\| \le \sum_{n=1}^{N} y_n (w^T x_n + b)$$

$$\Rightarrow N\rho \|w\| \le w^T \sum y_n x_n + b \sum y_n$$

we know by construction $\sum y_n = 0$ ,

$$\Rightarrow N\rho \|w\| \le w^T \sum y_n x_n$$

using cauchy-schwartz inequality,

$$(w^T \sum y_n x_n)^2 \le \|w\|^2 \|\sum y_n x_n\|^2$$

$$\Rightarrow (N\rho \|w\|)^2 \le \|w\|^2 \|\sum y_n x_n\|^2$$

$$\Rightarrow N^2 \rho^2 \le \|\sum y_n x_n\|^2 \le \frac{N^2 R^2}{N-1}$$

$$\Rightarrow \rho^2 \le \frac{R^2}{N-1}$$

$$\Rightarrow N-1 \le \frac{R^2}{\rho^2} \quad \Rightarrow N \le 1 + \frac{R^2}{\rho^2}$$

So, when N is ~~odd~~ even

$$N \le \frac{R^2}{\rho^2} + 1$$

& when N is odd

$$N \le \frac{R^2}{\rho^2} + 1 + \frac{1}{N}$$

$$\Rightarrow d_{vc}(\rho) \le \left\lceil \frac{R^2}{\rho^2} \right\rceil + 1 .$$

# Guidelines:
(common instances where marks have been cut)

Question1:
1) The lagrangian is not written properly. There should be extra two terms corresponding to the two constraints in the primal.
2) The constarint of alpha I the dual is not explained properly.

Question2-partC:
1)some steps are missing. Some results which have been used are not proved.

Question2-partD:
1)No explanations are given.

Question2-partE:
1)One can use Markov inequality or any other forms of it.