

COL 341 Homework2

Saket Kandoi

TOTAL POINTS

22 / 22

QUESTION 1

1 22 / 22

✓ + **10 pts** Correct

✓ + **5 pts** Click here to replace this description.

✓ + **5 pts** Click here to replace this description.

✓ + **2 pts** Click here to replace this description.

+ **2 pts** Click here to replace this description.

+ **1 pts** Click here to replace this description.

+ **0.5 pts** Click here to replace this description.

+ **0.25 pts** Click here to replace this description.

Homework 2

Lecturer: Prof. Chetan Arora

Saket Kandoi 2021MT60265

Question 1

10marks

Primal problem of soft SVM is:

$$\min_{\mathbf{w}, b} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{n=1}^N \zeta_n \quad (1)$$

such that, $\forall n = 1, \dots, N$:

$$y_n(\mathbf{w}^T \mathbf{x}_n + b) \geq 1 - \zeta_n \quad (2)$$

$$\zeta_n \geq 0 \quad (3)$$

A general primal problem,

$$\begin{aligned} \min_{\mathbf{u}} \quad & \frac{1}{2} \mathbf{u}^T \theta \mathbf{u} + P^T \mathbf{u} \\ \text{s.t.} \quad & \mathbf{a}_n^T \mathbf{u} \geq \mathbf{c}_n \end{aligned}$$

Has its corresponding Lagrangian dual of the form,

$$\max_{\alpha} \min_{\mathbf{u}} L$$

$$\begin{aligned} \text{where } L = \quad & \frac{1}{2} \mathbf{u}^T \theta \mathbf{u} + P^T \mathbf{u} + \sum_{n=1}^N \alpha_n (\mathbf{c}_n - \mathbf{a}_n^T \mathbf{u}) \\ \text{s.t.} \quad & \alpha_n \geq 0 \end{aligned}$$

Hence, L for soft SVM is:

$$L = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{n=1}^N \zeta_n + \sum_{n=1}^N \alpha_n (1 - \zeta_n - y_n(\mathbf{w}^T \mathbf{x}_n + b)) - \sum_{n=1}^N \beta_n \zeta_n$$

And its lagrangian dual is,

$$\max_{\alpha} \min_{\mathbf{w}, b, \zeta} L$$

$$\text{s.t. } \alpha_n \geq 0 \text{ and } \beta_n \geq 0$$

Applying Stationarity conditions w.r.t each parameter,

$$\begin{aligned} \frac{\delta L}{\delta b} &= - \sum_n \alpha_n y_n \\ \frac{\delta L}{\delta b} &= 0 \\ \sum_n \alpha_n y_n &= 0 \end{aligned} \tag{4}$$

$$\begin{aligned} \frac{\delta L}{\delta \mathbf{w}} &= \mathbf{w} - \sum_n \alpha_n y_n \mathbf{x}_n \\ \frac{\delta L}{\delta \mathbf{w}} &= 0 \\ \mathbf{w} &= \sum_n \alpha_n y_n \mathbf{x}_n \end{aligned} \tag{5}$$

$$\frac{\delta L}{\delta \boldsymbol{\zeta}} = \frac{\delta L}{\delta \zeta_1} \hat{\mathbf{e}}_1 + \frac{\delta L}{\delta \zeta_2} \hat{\mathbf{e}}_2 + \dots + \frac{\delta L}{\delta \zeta_n} \hat{\mathbf{e}}_n$$

$$\begin{aligned} \frac{\delta L}{\delta \zeta_i} &= C - \alpha_n - \beta_n \\ \frac{\delta L}{\delta \boldsymbol{\zeta}} &= 0 \\ \frac{\delta L}{\delta \zeta_i} &= 0 \\ \alpha_n &= C - \beta_n \end{aligned} \tag{6}$$

Dual condition implies

$$\begin{aligned} \beta_n &\geq 0 \\ \alpha_n &\leq C \end{aligned}$$

$$\begin{aligned} L &= \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{n=1}^N \zeta_n + \sum_{n=1}^N \alpha_n - \sum_{n=1}^N \alpha_n \zeta_n - \sum_{n=1}^N \alpha_n y_n \mathbf{w}^T \mathbf{x}_n + \sum_{n=1}^N \alpha_n y_n b - \sum_{n=1}^N \beta_n \zeta_n \\ L &= \frac{1}{2} \mathbf{w}^T \mathbf{w} + \sum_{n=1}^N \alpha_n + \sum_{n=1}^N (C - \alpha_n - \beta_n) \zeta_n - \mathbf{w}^T \sum_{n=1}^N \alpha_n y_n \mathbf{x}_n + b \sum_{n=1}^N \alpha_n y_n \end{aligned}$$

Substituting equations (4), (5) and (6) in Lagrangian,

$$\begin{aligned}
L &= \frac{1}{2} \mathbf{w}^T \mathbf{w} + \sum_{n=1}^N \alpha_n - \mathbf{w}^T \mathbf{w} \\
L &= -\frac{1}{2} \mathbf{w}^T \mathbf{w} + \sum_{n=1}^N \alpha_n \\
L &= -\frac{1}{2} \sum_{m=1}^N \sum_{n=1}^N y_n y_m \alpha_n \alpha_m \mathbf{x}_n^T \mathbf{x}_m + \sum_{n=1}^N \alpha_n
\end{aligned}$$

Hence, dual problem reduces to

$$\max_{\alpha} -\frac{1}{2} \sum_{m=1}^N \sum_{n=1}^N y_n y_m \alpha_n \alpha_m \mathbf{x}_n^T \mathbf{x}_m + \sum_{n=1}^N \alpha_n$$

which is equivalent to

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

$$\begin{aligned}
&\text{s.t. } \sum_{n=1}^N y_n \alpha_n = 0 \\
&0 \leq \alpha_n \leq C \quad \forall n
\end{aligned}$$

Question 2 12marks

Given, N is even and $N/2$ labels of y_1, \dots, y_n are $+1$ and the remaining $N/2$ are -1 .

$$\begin{aligned}
\sum_{n=1}^N y_n &= 0 \\
\|\mathbf{x}\|^2 &= \|\mathbf{x} \cdot \mathbf{x}\|
\end{aligned}$$

1.

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

Since terms of the summation are scalars,

$$\left\| \sum_{n=1}^N y_n \mathbf{x}_n \right\|^2 = \sum_{n=1}^N \sum_{m=1}^N y_n y_m \mathbf{x}_n^T \mathbf{x}_m \tag{7}$$

2. When $n = m$,

$$y_n y_m = y_n^2$$

Label y_n can be either 1 or -1 .

$$\begin{aligned}
y_n y_m &= 1 \\
\mathbb{P}[y_n y_m = 1] &= 1
\end{aligned}$$

$$\mathbb{E}[y_n y_m] = 1$$

When $n \neq m$,

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

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

Hence,

$$\mathbb{E}[y_n y_m] = \begin{cases} 1 & m = n \\ -\frac{1}{N-1} & m \neq n \end{cases} \quad (8)$$

3. By Equation (7),

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

Since n , m and N are fixed, and \mathbf{x}_n and \mathbf{x}_m are not random variables,

$$\begin{aligned}
\mathbb{E} \left[\left\| \sum_{n=1}^N y_n \mathbf{x}_n \right\|^2 \right] &= \sum_{n=1}^N \sum_{m=1}^N \mathbb{E} [y_n y_m] \mathbf{x}_n^T \mathbf{x}_m \\
&= \sum_{n=1}^N \sum_{m=n}^N \mathbb{E} [y_n y_m] \mathbf{x}_n^T \mathbf{x}_m + \sum_{n=1}^N \sum_{\substack{m=1 \\ m \neq n}}^N \mathbb{E} [y_n y_m] \mathbf{x}_n^T \mathbf{x}_m \\
&= \sum_{n=1}^N \sum_{m=n}^N \mathbf{x}_n^T \mathbf{x}_m + \sum_{n=1}^N \sum_{\substack{m=1 \\ m \neq n}}^N \frac{-\mathbf{x}_n^T \mathbf{x}_m}{N-1} \\
&= \sum_{n=1}^N \|\mathbf{x}_n\|^2 + \sum_{n=1}^N \sum_{\substack{m=1 \\ m \neq n}}^N \frac{-\mathbf{x}_n^T \mathbf{x}_m}{N-1} \\
&= \sum_{n=1}^N \|\mathbf{x}_n\|^2 + \sum_{n=1}^N \sum_{m=1}^N \frac{-\mathbf{x}_n^T \mathbf{x}_m}{N-1} - \sum_{n=1}^N \sum_{m=n}^N \frac{-\mathbf{x}_n^T \mathbf{x}_m}{N-1} \\
&= \sum_{n=1}^N \|\mathbf{x}_n\|^2 - \frac{1}{N-1} \left\| \sum_{n=1}^N \mathbf{x}_n \right\|^2 + \frac{1}{N-1} \sum_{n=1}^N \|\mathbf{x}_n\|^2 \\
&= \frac{N}{N-1} \sum_{n=1}^N \|\mathbf{x}_n\|^2 - \frac{1}{N-1} \left\| \sum_{n=1}^N \mathbf{x}_n \right\|^2 \\
&= \frac{N}{N-1} \sum_{n=1}^N \|\mathbf{x}_n\|^2 - \frac{N^2}{N-1} \|\tilde{\mathbf{x}}\|^2 \\
&= \frac{N}{N-1} \left(\sum_{n=1}^N \|\mathbf{x}_n\|^2 - N \|\tilde{\mathbf{x}}\|^2 \right) \\
&= \frac{N}{N-1} \left(\sum_{n=1}^N \|\mathbf{x}_n\|^2 - 2N \|\tilde{\mathbf{x}}\|^2 + N \|\tilde{\mathbf{x}}\|^2 \right) \\
&= \frac{N}{N-1} \left(\sum_{n=1}^N \|\mathbf{x}_n\|^2 - 2N \|\tilde{\mathbf{x}}\|^2 + \sum_{n=1}^N \|\tilde{\mathbf{x}}\|^2 \right) \\
&= \frac{N}{N-1} \left(\sum_{n=1}^N \|\mathbf{x}_n\|^2 - 2N \tilde{\mathbf{x}}^T \tilde{\mathbf{x}} + \sum_{n=1}^N \|\tilde{\mathbf{x}}\|^2 \right) \\
&= \frac{N}{N-1} \left(\sum_{n=1}^N \|\mathbf{x}_n\|^2 - 2\tilde{\mathbf{x}}^T \cdot \sum_{n=1}^N \mathbf{x}_n + \sum_{n=1}^N \|\tilde{\mathbf{x}}\|^2 \right)
\end{aligned}$$

$$\begin{aligned}
&= \frac{N}{N-1} \left(\sum_{n=1}^N \|\mathbf{x}_n\|^2 - 2 \sum_{n=1}^N \tilde{\mathbf{x}}^T \cdot \mathbf{x}_n + \sum_{n=1}^N \|\tilde{\mathbf{x}}\|^2 \right) \\
&= \frac{N}{N-1} \sum_{n=1}^N (\|\mathbf{x}_n\|^2 - 2\tilde{\mathbf{x}}^T \cdot \mathbf{x}_n + \|\tilde{\mathbf{x}}\|^2) \\
&= \frac{N}{N-1} \sum_{n=1}^N (\|\mathbf{x}_n - \tilde{\mathbf{x}}\|^2) \tag{9}
\end{aligned}$$

4.

$$\|\mathbf{x}_n - \tilde{\mathbf{x}}\|^2 \leq \sum_{n=1}^N \|\mathbf{x}_n - \tilde{\mathbf{x}}\|^2$$

Since $\sum_{n=1}^N \|\mathbf{x}_n - \boldsymbol{\mu}\|^2$ is minimum at $\boldsymbol{\mu} = \tilde{\mathbf{x}}$,

$$\|\mathbf{x}_n - \tilde{\mathbf{x}}\|^2 \leq \sum_{n=1}^N \|\mathbf{x}_n - \boldsymbol{\mu}\|^2 \text{ for all } \boldsymbol{\mu} \in \mathbb{R}^d$$

Putting $\boldsymbol{\mu} = \mathbf{0}$,

$$\|\mathbf{x}_n - \tilde{\mathbf{x}}\|^2 \leq \sum_{n=1}^N \|\mathbf{x}_n\|^2$$

Since input space is a ball of radius R in \mathbb{R}^d ,

$$\begin{aligned}
\|\mathbf{x}_n\| &\leq R \\
\sum_{n=1}^N \|\mathbf{x}_n\|^2 &\leq \sum_{n=1}^N R^2 = NR^2
\end{aligned}$$

Hence,

$$\|\mathbf{x}_n - \tilde{\mathbf{x}}\|^2 \leq NR^2$$

(*Note:* This is the required proof in the question, however I shall prove the same bound for the summation, which has an application in the following part.)

Proof of Hint:

Triangle inequality on norms gives us

$$\|A_1\| + \|A_2\| + \dots + \|A_n\| \geq \|A_1 + A_2 + \dots + A_n\|$$

Hence,

$$\begin{aligned}\sum_{n=1}^N \|\mathbf{x}_n - \boldsymbol{\mu}\|^2 &\geq \left\| \sum_{n=1}^N (\mathbf{x}_n - \boldsymbol{\mu}) \right\|^2 \\ \sum_{n=1}^N \|\mathbf{x}_n - \boldsymbol{\mu}\|^2 &\geq \|N\tilde{\mathbf{x}} - N\boldsymbol{\mu}\|^2\end{aligned}$$

RHS is minimum at $\tilde{\mathbf{x}} = \boldsymbol{\mu}$, hence proved.

Since $\sum_{n=1}^N \|\mathbf{x}_n - \boldsymbol{\mu}\|^2$ is minimum at $\boldsymbol{\mu} = \tilde{\mathbf{x}}$,

$$\sum_{n=1}^N \|\mathbf{x}_n - \tilde{\mathbf{x}}\|^2 \leq \sum_{n=1}^N \|\mathbf{x}_n - \boldsymbol{\mu}\|^2 \text{ for all } \boldsymbol{\mu} \in \mathbb{R}^d$$

Putting $\boldsymbol{\mu} = \mathbf{0}$,

$$\sum_{n=1}^N \|\mathbf{x}_n - \tilde{\mathbf{x}}\|^2 \leq \sum_{n=1}^N \|\mathbf{x}_n\|^2$$

Since input space is a ball of radius R in \mathbb{R}^d ,

$$\begin{aligned}\|\mathbf{x}_n\| &\leq R \\ \sum_{n=1}^N \|\mathbf{x}_n\|^2 &\leq \sum_{n=1}^N R^2 = NR^2\end{aligned}$$

Hence,

$$\sum_{n=1}^N \|\mathbf{x}_n - \tilde{\mathbf{x}}\|^2 \leq NR^2 \tag{10}$$

5. By equations (9) and (10),

$$\begin{aligned}\mathbb{E} \left[\left\| \sum_{n=1}^N y_n \mathbf{x}_n \right\|^2 \right] &\leq \frac{N}{N-1} \cdot NR^2 \\ \mathbb{E} \left[\left\| \sum_{n=1}^N y_n \mathbf{x}_n \right\|^2 \right] &\leq \frac{N^2 R^2}{N-1}\end{aligned}$$

Since expectation of $\left\|\sum_{n=1}^N y_n \mathbf{x}_n\right\|^2$ is less than $\frac{N^2 R^2}{N-1}$, it can attain a value less than $\frac{N^2 R^2}{N-1}$ with non-zero probability. Hence,

$$\mathbb{P}\left[\left\|\sum_{n=1}^N y_n \mathbf{x}_n\right\|^2 \leq \frac{N^2 R^2}{N-1}\right] > 0$$

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

Thus, for some choice of a balanced dichotomy y_1, \dots, y_n ,

$$\sum_{n=1}^N y_n = 0 \tag{11}$$

$$\left\|\sum_{n=1}^N y_n \mathbf{x}_n\right\| \leq \frac{NR}{\sqrt{N-1}} \tag{12}$$

Since N points are being shattered, they can be separated by SVM with margin at least ρ . Hence, for some \mathbf{w} and b ,

$$\rho \leq d_n \forall n$$

where d_n is distance from the hyper-plane.

$$d_n = \frac{y_n(\mathbf{w}^T \mathbf{x}_n + b)}{\|\mathbf{w}\|}$$

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

Taking summation of inequalities over all n ,

$$\begin{aligned} \rho N &\leq \frac{1}{\|\mathbf{w}\|} \left(\sum_{n=1}^N y_n \mathbf{w}^T \mathbf{x}_n + \sum_{n=1}^N b y_n \right) \\ &\leq \frac{1}{\|\mathbf{w}\|} \left(\left\| \sum_{n=1}^N y_n \mathbf{w}^T \mathbf{x}_n \right\| + b \sum_{n=1}^N y_n \right) \\ &\leq \frac{1}{\|\mathbf{w}\|} \left(\left\| \mathbf{w}^T \sum_{n=1}^N y_n \mathbf{x}_n \right\| \right) && \text{(By Equation (11))} \\ &\leq \frac{1}{\|\mathbf{w}\|} \left(\|\mathbf{w}^T\| \left\| \sum_{n=1}^N y_n \mathbf{x}_n \right\| \right) && \text{(By Cauchy-Schwarz inequality } \|\mathbf{u}\mathbf{v}\| \leq \|\mathbf{u}\| \cdot \|\mathbf{v}\|) \\ &\leq \left\| \sum_{n=1}^N y_n \mathbf{x}_n \right\| \\ &\leq \frac{NR}{\sqrt{N-1}} && \text{(By Equation (12))} \end{aligned}$$

$$\begin{aligned}
\rho N &\leq \frac{NR}{\sqrt{N-1}} \\
\sqrt{N-1} &\leq \frac{R}{\rho} \\
N &\leq \frac{R^2}{\rho^2} + 1 \\
\therefore d_{VC}(\rho) &\leq \lceil R^2/\rho^2 \rceil + 1
\end{aligned}$$