COL341: Fundamentals of Machine Learning Spring 2023

Homework 2

Lecturer: Prof. Chetan Arora Saket Kandoi 2021MT60265

Question 1

Primal problem of soft SVM is:

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

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

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

$$\zeta_n \geq 0 \tag{3}$$

A general primal problem,

$$\min_{\boldsymbol{u}} \frac{1}{2}\boldsymbol{u}^T\theta\boldsymbol{u} + P^T\boldsymbol{u}$$
$$\text{s.t. } \boldsymbol{a}_n^T\boldsymbol{u} \geq \boldsymbol{c}_n$$

Has its corresponding Lagrangian dual of the form,

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

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

Hence, L for soft SVM is:

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

And its lagrangian dual is,

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

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

Applying Stationarity conditions w.r.t each parameter,

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

$$\begin{aligned}
\frac{\delta L}{\delta \boldsymbol{w}} &= \boldsymbol{w} - \sum_n \alpha_n y_n \boldsymbol{x}_n \\
\frac{\delta L}{\delta \boldsymbol{w}} &= 0 \\
\boldsymbol{w} &= \sum_n \alpha_n y_n \boldsymbol{x}_n \qquad (5)
\end{aligned}$$

$$\frac{\delta L}{\delta \boldsymbol{\zeta}} = \frac{\delta L}{\delta \zeta_1}\hat{\boldsymbol{e_1}} + \frac{\delta L}{\delta \zeta_2}\hat{\boldsymbol{e_2}} + ... + \frac{\delta L}{\delta \zeta_n}\hat{\boldsymbol{e_n}}$$

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

Dual condition implies

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

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

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

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

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

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

Hence, dual problem reduces to

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

which is equivalent to

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

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

$$0 \le \alpha_n \le C \ \forall n$$

Question 2

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

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

$$||\boldsymbol{x}||^2 = ||\boldsymbol{x} \cdot \boldsymbol{x}||$$

1.

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

Since terms of the summation are scalars,

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

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 $n \neq m$,

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

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

Hence,

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

3. By Equation (7),

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

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

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

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

4.

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

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

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

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

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

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

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

Hence,

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

(*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

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

Hence,

$$\sum_{n=1}^{N} ||\boldsymbol{x}_n - \boldsymbol{\mu}||^2 \geq || \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{\mu})||^2$$
$$\sum_{n=1}^{N} ||\boldsymbol{x}_n - \boldsymbol{\mu}||^2 \geq ||N\tilde{\boldsymbol{x}} - N\boldsymbol{\mu}||^2$$

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

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

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

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

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

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

$$||\boldsymbol{x}_n|| \leq R$$
$$\sum_{n=1}^{N} ||\boldsymbol{x}_n||^2 \leq \sum_{n=1}^{N} R^2 = NR^2$$

Hence,

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

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

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

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

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

Thus, for some choice of of a balanced dichotomy $y_1, ..., y_n$,

$$\sum_{n=1}^{N} y_n = 0 \tag{11}$$
$$\left|\left|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right|\right| \leq \frac{NR}{\sqrt{N-1}} \tag{12}$$

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

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

where d_n is distance from the hyper-plane.

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

Taking summation of inequalities over all n,

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

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