Chinmay Mittal
(2020CS10336)

Homework 2 | COL341
Machine Learning

March 18, 2023

Question 1

The primal optimization problem for the soft-margin SVM is as follows

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

$$\text{subject to } y_n(\boldsymbol{w}^T\boldsymbol{x}_n + b) \geq 1 - \xi_n \quad \text{for } n = 1, 2, 3 \dots N$$

$$\xi_n \geq 0 \text{ for } n = 1, 2, 3 \dots N$$

The Lagrangian of the primal optimization problem is given by:

$$\mathcal{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{w}\boldsymbol{w}^T + C\sum_{n=1}^{N}\xi_n + \sum_{n=1}^{N}\alpha_n(1 - \xi_n - y_n(\boldsymbol{w}^T\boldsymbol{x}_n + b)) - \sum_{n=1}^{N}\beta_n\xi_n$$

Here $\alpha_n \geq 0$ are the Lagrange multipliers for the constraints $y_n(\boldsymbol{w}^T\boldsymbol{x}_n + b) \geq 1 - \xi_n$ and $\beta_n \geq 0$ are the Lagrange multipliers for the constraints $\xi_n \geq 0$.

From the KKT optimality conditions we have, $\frac{\partial \mathcal{L}}{\partial \xi_n} = 0$ (stationarity condition) which gives us:

$$\frac{\partial \mathcal{L}}{\partial \xi_n} = C + \alpha_n - \beta_n = 0 \Rightarrow \alpha_n + \beta_n = C \text{ for } n = 1, 2 \dots N \tag{1}$$

Substituting $\beta_n = C - \alpha_n$ in the Lagrangian we have:

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

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

$$\text{subject to } \alpha_n \geq 0, \beta_n \geq 0 \text{ for } n = 1, 2, \dots N$$

the constraints simplify to

$$\alpha_n \geq 0, C - \alpha_n \geq 0 \Rightarrow 0 \leq \alpha_n \leq C \text{ for } n = 1, 2, \dots N \tag{2}$$

From the KKT optimality conditions we have $\frac{\partial \mathcal{L}}{\partial b} = 0$, $\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = 0$ (stationarity condition)

$$\frac{\partial \mathcal{L}}{\partial b} = -\sum_{n=1}^{N}\alpha_n y_n = 0 \tag{3}$$

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

Substituting in the Lagrangian we get:

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

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

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

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

The dual optimization problem thus becomes:

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

The constraints come from equation (2) and (3)

$$\text{subject to } \sum_{n=1}^{N} \alpha_n y_n = 0, 0 \leq \alpha_n \leq C$$

which is same as

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

$$\text{subject to } \sum_{n=1}^{N} \alpha_n y_n = 0, 0 \leq \alpha_n \leq C$$

Question 2

N even

Fix $x_1, x_2 \dots x_N$ be N points that are shattered by hyperplanes with margin ρ.
We randomly assign $\frac{N}{2}$ of the labels from $y_1, y_2, \dots y_N$ to be $+1$ and the others to be -1, thus by construction $\sum_{n=1}^{N} y_n = 0$

---

1)

$$|| \sum_{n=1}^{N} y_n \boldsymbol{x}_n ||^2 = (\sum_{n=1}^{N} y_n \boldsymbol{x_n})^T (\sum_{m=1}^{N} y_m \boldsymbol{x_m}) = (\sum_{n=1}^{N} y_n \boldsymbol{x_n^T})(\sum_{m=1}^{N} y_m \boldsymbol{x_m}) = \sum_{n=1}^{N} \sum_{m=1}^{N} y_n y_m \boldsymbol{x_n^T x_m}$$

---

2)

$$\text{When } n = m \Rightarrow y_n = y_m \Rightarrow y_n y_m = y_n^2 = |y_n|^2 = 1$$

$$\text{Thus we get } \mathbb{P}[y_n y_m = 1] = 1 \text{when } n = m$$

$$\text{Next consider the case when } n \neq m, y_n y_m = 1 \text{ when both } y_n, y_m \text{ are 1 or } -1.$$

$$\mathbb{P}[y_n y_m = 1] = \frac{N/2}{N} \frac{N/2 - 1}{N - 1} + \frac{N/2}{N} \frac{N/2 - 1}{N - 1} = \frac{N/2 - 1}{N - 1}$$

We have $N/2$ labels +1 and $N/2$ labels which are -1. The first part is when both are +1 and the second part is when both are -1.

$$\text{Thus } \mathbb{E}[y_n y_m] = 1 \text{ when } n = m \text{ (constant).}$$

When $n \neq m$

$$\mathbb{E}[y_n y_m] = (+1(\frac{N/2 - 1}{N - 1}) - 1(1 - \frac{N/2 - 1}{N - 1}))$$

$$\mathbb{E}[y_n y_m] = 2(\frac{N/2 - 1}{N - 1}) - 1 = \frac{N - 2}{N - 1} - 1 = -\frac{1}{N - 1}$$

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

---

3) From Linearity of Expectation

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

$$= \sum_{n=1}^{N} \boldsymbol{x_n^T x_n} + \sum_{n=1}^{N} \sum_{m=1, m\neq n}^{N} -\frac{1}{N-1} \boldsymbol{x_n^T x_m}$$

$$= \sum_{n=1}^{N} \boldsymbol{x_n^T x_n} - \frac{1}{N-1} \sum_{n=1}^{N} \boldsymbol{x_n^T} (\sum_{m=1}^{N} \boldsymbol{x}_m - \boldsymbol{x}_n) = \sum_{n=1}^{N} \boldsymbol{x_n^T x_n} + \frac{1}{N-1} \sum_{n=1}^{N} \boldsymbol{x_n^T x_n} - \frac{1}{N-1} \sum_{n=1}^{N} \boldsymbol{x_n^T} (N \bar{\boldsymbol{x}})$$

$$= \frac{N}{N-1} \sum_{n=1}^{N} \boldsymbol{x_n^T x_n} - \frac{N}{N-1} \sum_{n=1}^{N} \boldsymbol{x_n^T} \bar{\boldsymbol{x}}$$

$$= \frac{N}{N-1} (\sum_{n=1}^{N} \boldsymbol{x_n^T x_n} + \sum_{n=1}^{N} \bar{\boldsymbol{x}}^T \boldsymbol{x_n} - 2 \sum_{n=1}^{N} \bar{\boldsymbol{x}}^T \boldsymbol{x_n})$$

We also have $\sum_{n=1}^{N} \boldsymbol{x}_n = \sum_{n=1}^{N} \bar{\boldsymbol{x}}$
Thus we get

$$\mathbb{E}[|| \sum_{n=1}^{N} y_n \boldsymbol{x}_n ||^2] = \frac{N}{N-1} (\sum_{n=1}^{N} \boldsymbol{x_n^T x_n} + \sum_{n=1}^{N} \bar{\boldsymbol{x}}^T \bar{\boldsymbol{x}} - 2 \sum_{n=1}^{N} \bar{\boldsymbol{x}}^T \boldsymbol{x_n})$$

$$= \frac{N}{N-1} \sum_{n=1}^{N} (\bar{\boldsymbol{x}}^T - \boldsymbol{x_n^T})(\bar{\boldsymbol{x}} - \boldsymbol{x_n})$$

$$= \frac{N}{N-1} \sum_{n=1}^{N} (||\bar{\boldsymbol{x}} - \boldsymbol{x_n}||^2)$$

---

4)

$$\sum_{n=1}^{N} ||\bar{\boldsymbol{x}} - \boldsymbol{x}_n||^2 = \sum_{n=1}^{N} (\bar{\boldsymbol{x}}^T - \boldsymbol{x_n^T})(\bar{\boldsymbol{x}} - \boldsymbol{x_n})$$

$$= \sum_{n=1}^{N} \boldsymbol{x_n^T x_n} + \sum_{n=1}^{N} \bar{\boldsymbol{x}}^T \bar{\boldsymbol{x}} - 2 \sum_{n=1}^{N} \bar{\boldsymbol{x}}^T \boldsymbol{x_n} = \sum_{n=1}^{N} \boldsymbol{x_n^T x_n} + \sum_{n=1}^{N} \bar{\boldsymbol{x}}^T \bar{\boldsymbol{x}} - 2 \bar{\boldsymbol{x}}^T \sum_{n=1}^{N} \boldsymbol{x}_n$$

Since $\sum_{n=1}^{N} \boldsymbol{x}_n = \sum_{n=1}^{N} \bar{\boldsymbol{x}}$

$$= \sum_{n=1}^{N} \boldsymbol{x_n^T x_n} + \sum_{n=1}^{N} \bar{\boldsymbol{x}}^T \bar{\boldsymbol{x}} - 2 \bar{\boldsymbol{x}}^T \sum_{n=1}^{N} \bar{\boldsymbol{x}}$$

$$= \sum_{n=1}^{N} \boldsymbol{x_n^T x_n} + \sum_{n=1}^{N} \bar{\boldsymbol{x}}^T \bar{\boldsymbol{x}} - 2 \sum_{n=1}^{N} \bar{\boldsymbol{x}}^T \boldsymbol{x_n} = \sum_{n=1}^{N} \boldsymbol{x_n^T x_n} - \sum_{n=1}^{N} \bar{\boldsymbol{x}}^T \bar{\boldsymbol{x}}$$

$$\leq \sum_{n=1}^{N} \boldsymbol{x_n^T x_n} \text{ since } \bar{\boldsymbol{x}}^T \bar{\boldsymbol{x}} \geq 0$$

Hence

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

---

5)

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

Consider a positive Random Variable X

$$\mathbb{P}[X \leq c] = 0 \Rightarrow \mathbb{E}[X^2] > c^2$$

Thus

$$\mathbb{E}[X^2] \leq c^2 => \mathbb{P}[X \leq c] > 0$$

Thus we can conclude that

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

Thus there exists some assignment $y_1, y_2, \ldots y_N$, $\sum_{n=1}^{N} y_n = 0$ such that $||\sum_{n=1}^{N} y_n \boldsymbol{x}_n|| \leq \frac{NR}{\sqrt{N-1}}$ and $\boldsymbol{x_1}, \ldots$ $\boldsymbol{x_N}$ are shattered by hyperplanes with margin ρ.

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

$$\Rightarrow N\rho||\boldsymbol{w}|| \leq \sum_{n=1}^{N} y_n(\boldsymbol{w}^T \boldsymbol{x}_n + b) = \boldsymbol{w}^T \sum_{n=1}^{N}(y_n \boldsymbol{x}_n) + b\sum_{n=1}^{N} y_n = \boldsymbol{w}^T \sum_{n=1}^{N}(y_n \boldsymbol{x}_n) + 0$$

Using Cauchy-Schwartz Inequality ($||x^T y|| \leq \ ||x|| \ ||y||$):

$$N\rho||\boldsymbol{w}|| \leq \boldsymbol{w}^T \sum_{n=1}^{N}(y_n \boldsymbol{x}_n) \leq ||\boldsymbol{w}^T|| \ ||\sum_{n=1}^{N}(y_n \boldsymbol{x}_n)||$$

$$N\rho||\boldsymbol{w}|| \leq ||\boldsymbol{w}||\frac{NR}{\sqrt{N-1}}$$

$$\rho \leq \frac{R}{\sqrt{N-1}} => \Rightarrow N \leq \frac{R^2}{\rho^2} + 1$$
