COL341 - Homework-2

Rajat Bhardwaj
2020CS50436

March 2023

1 Question - 1

Soft SVM

$$min\frac{1}{2}w^Tw + C\sum_{n=1}^{N}\xi_n$$

S.T.

$$y_n(w^Tx_n + b) > 1 - \xi_n$$

and

$$\xi_n \geq 0$$

The Lagrange function will be

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

We need to minimize L w.r.t w,b and we need to maximize L w.r.t. $\alpha \geq 0$ and $\beta \geq 0$ We will first show that this lagrangian is equivalent to the soft SVM
Consider the third term $\sum_{n=1}^{N}\alpha_n(1 - \xi_n - y_n(w^Tx_n + b))$
This term can be either > 0 or ≤ 0
Case 1
$\sum_{n=1}^{N}\alpha_n(1 - \xi_n - y_n(w^Tx_n + b)) > 0$
we know that $\alpha_n \geq 0$ therefore to maximize L we take $\alpha = \inf$. Since we need to minimize L w.r.t w and b, we can make $(1 - \xi_n - y_n(w^Tx_n + b)) = 0$

Thus This case is not possible since we can always choose w,b such that $(1 - \xi_n - y_n(w^Tx_n + b)) \leq 0$ to minimize L. Now to maximize L w.r.t. α we keep $\alpha = 0$ Therefore,

$$\sum_{n=1}^{N}\alpha_n(1 - \xi_n - y_n(w^Tx_n + b)) = 0$$

Similarly for the last term we can argue that $-\xi \leq 0$ therefore $\xi \geq 0$ also $\sum_{n=1}^{N}\beta_n\xi_n = 0$ (because we can choose ξ to minimize $-\sum_{n=1}^{N}\beta_n\xi_n$ thus $\xi \geq 0$ also we need to maximize $-\sum_{n=1}^{N}\beta_n\xi_n$ w.r.t β. we can chose $\beta = 0$ when $\xi > 0$ This will always lead to $\sum_{n=1}^{N}\beta_n\xi_n = 0$

We end up with the same equation and constraints as soft SVM. i.e.

Minimize

$$\frac{1}{2}w^Tw + C\sum_{n=1}^{N}\xi_n$$

s.t.

$$(1 - \xi_n - y_n(w^Tx_n + b)) \leq 0$$

Which implies

$$y_n(w^T x_n + b) > 1 - \xi_n$$

and

$$\xi_n \geq 0$$

Now we have shown that the Lagrange equation is the same as the soft SVM
We will now derive the final result
According to the KKT conditions, $\frac{\partial L}{\partial \xi} = 0$ Therefore

$$\sum_{n=1}^{N} C - \sum_{n=1}^{N} \alpha_n - \sum_{n=1}^{N} \beta_n = 0$$

$$\sum_{n=1}^{N} (C - \alpha_n - \beta_n) = 0$$

This implies

$$\beta_n = C - \alpha_n$$

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

Which gives us

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

We know that $\frac{\partial L}{\partial w} = 0$ and $\frac{\partial L}{\partial b} = 0$ Therefore

$$\frac{\partial L}{\partial w} = 0$$

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

therefore

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

also

$$\frac{\partial L}{\partial b} = 0$$

implies

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

Now we plug back these values into L. We get

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

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

The term with β disappears because $\sum_{n=1}^{N} \alpha_n y_n = 0$
We notice that the first and their term have same variables therefore we get

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

We need to maximize this equation w.r.t α
or we can minimize

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

w.r.t α

Thus we have derived the following equations and conditions
Minimize

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

w.r.t α (where α is a n-dimensional vector) s.t.

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

and

$$\beta_n = C - \alpha_n$$

or

$$C \geq \alpha_n \geq 0$$

(because $\alpha_n \geq 0$ according to our initial conditions)

2 Question 2

2.1 Show $\| \sum_{n=1}^{N} y_n x_n \|^2 = \sum_{n=1}^{N}\sum_{m=1}^{N} y_n y_m x_n^T x_m$

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

$$= \sum_{n=1}^{N} y_n x_n^T \sum_{m=1}^{N} y_m x_m$$

Now we need to multiply each term of the left term of the product with the right term of the product since y_n is a real number and x_n is a row. Therefore the term $y_n x_n^T$ show be multiplied with each $y_m x_m$ term
Therefore we get

$$= \sum_{n=1}^{N}\sum_{m=1}^{N} y_n x_n^T y_m x_m$$

$$= \sum_{n=1}^{N}\sum_{m=1}^{N} y_n y_m x_n^T x_m$$

2.2 Show that

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

We know that $y_i \in \{-1, 1\}$ therefore,
for m = n, $y_n y_m = 1 \times 1 = 1$ or $y_n y_m = (-1) \times (-1) = 1$,
hence $y_n y_m = 1$
$P[y_n y_m = 1] =$? such that n $\neq$ m
Now $y_n y_m = 1$ when $y_n = y_m$ This is only possible if both are 1 or both are -1 the number of possible ways to select y_n and y_m such that both are 1 is ${}^{\frac{N}{2}}C_2$ because we have $\frac{N}{2}$ points as $+1$ and the same

number of points as -1. Similarly for both points to be -1 the number of possible ways to select y_n and y_m are ${}^{\frac{N}{2}}C_2$ due to the same reason

Therefore the number of possibilities for $y_n y_m = 1$ are ${}^{\frac{N}{2}}C_2 + {}^{\frac{N}{2}}C_2 = 2\,{}^{\frac{N}{2}}C_2$ Now the total number of ways to select 2 points from N point is ${}^{N}C_2$ Thus

$$P[y_n y_m = 1] = \frac{2\,{}^{\frac{N}{2}}C_2}{{}^{N}C_2}$$

This becomes

$$\frac{\frac{N}{2} - 1}{N - 1}$$

proving

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

Now for the first case i.e. n = m, we already know that the expected value will be 1. for the second case we have :

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

Now

$$P[y_n y_m = -1] = ?$$

Now number of ways to select y_n and y_m such that $y_n \neq y_m$ is $\frac{N}{2} \times \frac{N}{2}$ Now

$$P[y_n y_m = -1] = \frac{\frac{N}{2} \times \frac{N}{2}}{{}^{N}C2}$$

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

Therefore, for $m \neq n$

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

$$E[y_n y_m] = \frac{-1}{N - 1}$$

Hence we have

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

2.3 show that

$$E[\| \sum_{n=1}^{N} y_n x_n \|^2] = \frac{N}{N - 1} \sum_{n=1}^{N} \| x_n - \bar{x} \|^2$$

Proof :

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

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

$$E[\| \sum_{n=1}^{N} y_n x_n \|^2] = \sum_{n=1}^{N} (E[y_n y_n] x_n^T x_n + \sum_{m \neq n}^{N} E[y_n y_n] x_n^T x_m)$$

$$E[\| \sum_{n=1}^{N} y_n x_n \|^2] = \sum_{n=1}^{N} (x_n^T x_n - \sum_{m \neq n}^{N} \frac{1}{N - 1} x_n^T x_m)$$

Divide and multiply by N

$$E[\|\sum_{n=1}^{N} y_n x_n\|^2] = N\sum_{n=1}^{N}(\frac{1}{N}x_n^T x_n - \frac{1}{N}\sum_{m\neq n}^{N}\frac{1}{N-1}x_n^T x_m)$$

Taking $\frac{1}{N-1}$ common

$$E[\|\sum_{n=1}^{N} y_n x_n\|^2] = \frac{N}{N-1}\sum_{n=1}^{N}(\frac{N-1}{N}x_n^T x_n - \frac{1}{N}\sum_{m\neq n}^{N} x_n^T x_m)$$

Now we adjust $\frac{1}{N}\sum_{m\neq n}^{N} x_n^T x_m$ to $\frac{1}{N}\sum_{m=1}^{N} x_n^T x_m$ we need to add a term $\frac{1}{N}x_n x_n^T$ to $\frac{N-1}{N}x_n^T x_n$ Thus we get

$$E[\|\sum_{n=1}^{N} y_n x_n\|^2] = \frac{N}{N-1}\sum_{n=1}^{N}(x_n^T x_n - x_n^T\frac{1}{N}\sum_{m=1}^{N} x_m)$$

Now according to the question $\frac{1}{N}\sum_{m=1}^{N} x_m = \bar{x}$. Thus we get

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

We will prove that $\sum_{n}^{N}(x_n^T x_n - x_n^T\bar{x}) = \sum_{n}^{N}\|x_n - \bar{x}\|^2$
R.H.S.

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

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

$$= \sum_{n}^{N}(x_n^T x_n - x_n^T\bar{x}) = L.H.S.$$

This is because $\sum_{n}^{N}\bar{x}^T x_n = \sum_{n}^{N}\bar{x}^T\bar{x}$ Thus we get

$$E[\|\sum_{n=1}^{N} y_n x_n\|^2] = \frac{N}{N-1}\sum_{n=1}^{N}\|x_n - \bar{x}\|^2$$

2.4 Show that $\sum_{n=1}^{N}\|x_n - \bar{x}\|^2 \leq \sum_{n=1}^{N}\|x\|^2 \leq NR$

We know that $\|x_n - \mu\|^2$ attains minimum value at $\mu = \frac{1}{N}\sum_{n=1}^{N} x_n$ Therefore we get $\|x_n - 0\|$ will be greater than $\|x_n - \bar{x}\|^2$ or

$$\sum_{n=1}^{N}\|x_n - \bar{x}\|^2 \leq \sum_{n=1}^{N}\|x_n\|^2$$

It is given that $\|x\| \leq R$ therefore $\sum_{n=1}^{N}\|x_n\|^2 \leq \sum_{n=1}^{N} R^2 = NR^2$

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

2.5 Conclude that $E[\| \sum_{n=1}^{N} y_n x_n \|^2] \leq \frac{N^2R^2}{N-1}$

We have already proved that

$$E[\| \sum_{n=1}^{N} y_n x_n \|^2] = \frac{N}{N-1} \sum_{n=1}^{N} \|x_n - \bar{x}\|^2$$

Therefore from previous inequality

$$E[\| \sum_{n=1}^{N} y_n x_n \|^2] \leq \frac{N}{N-1}(NR^2)$$

$$E[\| \sum_{n=1}^{N} y_n x_n \|^2] \leq \frac{N^2R^2}{N-1}$$

Now we will conclude that $P[\| \sum_{n=1}^{N} y_n x_n \| \leq \frac{NR}{\sqrt{N-1}}] > 0$ Let us use contradiction, Assume $P[\| \sum_{n=1}^{N} y_n x_n \| \leq \frac{NR}{\sqrt{N-1}}] = 0$.(since LHS and RHS both are non-negative) therefore

$$P[\| \sum_{n=1}^{N} y_n x_n \| > \frac{NR}{\sqrt{N-1}}] = 1$$

We get

$$E[\| \sum_{n=1}^{N} y_n x_n \|] = \int_0^{\infty} MP[\| \sum_{n=1}^{N} y_n x_n \| = M]$$

$$E[\| \sum_{n=1}^{N} y_n x_n \|] > \int_0^{\infty} \frac{NR}{\sqrt{N-1}} P[\| \sum_{n=1}^{N} y_n x_n \|]$$

because $M > \frac{NR}{\sqrt{N-1}}$

$$E[\| \sum_{n=1}^{N} y_n x_n \|] > \frac{NR}{\sqrt{N-1}} \int_0^{\infty} P[\| \sum_{n=1}^{N} y_n x_n \|]$$

$$E[\| \sum_{n=1}^{N} y_n x_n \|] > \frac{NR}{\sqrt{N-1}}$$

which is a contradiction, therefore, $P[\| \sum_{n=1}^{N} y_n x_n \| \leq \frac{NR}{\sqrt{N-1}}] > 0$

Thus we also conclude that there exists a balanced dichotomy $y_1, y_2, ...y_n$ such that $y_1 + y_2 + ...y_n = 0$(since half of the points were +1 and half were -1 and $\| \sum_{n=1}^{N} y_n x_n \| \leq \frac{NR}{\sqrt{N-1}}$

2.6 VC dimension upper bound

Now for any data point (x_n, y_n) , the distance of this point from the separating hyperplane ($w^T x_n + b = 0$) is greater than the margin (ρ) and we will use y_n to balance the sign Therefore we have

$$\frac{y_n(w^T x_n + b)}{\|w\|} \geq \rho$$

Therefore

$$\rho\|w\| \leq y_n(w^T x_n + b)$$

Taking sum on both sides we get

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

$$N\rho\|w\| \leq \sum_{n=1}^{N} y_n w^T x_n + \sum_{n=1}^{N} y_n b$$

The second term in the above equation = 0 because $\sum y_n = 0$ since half of the y_n are -1 and others +1

We get

$$N\rho\|w\| \leq \sum_{n=1}^{N} y_n w^T x_n$$

$$N\rho\|w\| \leq w^T \sum_{n=1}^{N} y_n x_n$$

using Cauchy- Schwartz inequality

$$N\rho\|w\| \leq \|w\|\|\sum_{n=1}^{N} y_n x_n\|$$

$$N\rho \leq \|\sum_{n=1}^{N} y_n x_n\|$$

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

$$\rho \leq \frac{R}{\sqrt{N-1}}$$

This implies that

$$N \leq \frac{R^2}{N^2} + 1$$