COL341: Homework -1

Simran Mahawar
2020CS10387

March 19, 2023

Question 1

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:

$$\underset{w,b}{min}\ \frac{1}{2}w^Tw \quad s.t \qquad y_n(w^Tx_n+b) \geq 1 \qquad (n=1,2,...N)$$

The corresponding dual problem is:

$$\underset{\alpha}{min}\ \frac{1}{2}\sum_{m=1}^{N}\sum_{n=1}^{N} y_n y_m \alpha_n \alpha_m x_n^T x_m - \sum_{n=1}^{N}\alpha_n$$
$$s.t \qquad \sum_{n=1}^{N} y_n\alpha_n = 0 \quad \text{and} \quad \alpha_n \geq 0 \quad (n=1,2,...N)$$
$$We have to derive it.$$

Ans. Given primal problem corresponding to the soft SVM is:

$$\underset{w,b}{min}\ \frac{1}{2}w^Tw + C\sum_{n=1}^{N}\xi_n$$
$$s.t \qquad y_n(w^Tx_n+b) \geq 1-\xi_n \quad \text{and} \quad \xi_n \geq 0 \quad (n=1,2,...N)$$

The Lagrangian function that we will use is

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

$$\frac{\partial\mathcal{L}}{\partial\xi_n} = 0 \implies C - \alpha_n - \beta_n = 0$$

Thus we ca replace β_n by $C - \alpha_n$ then Lagrange function simplifies to

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

Now we have to minimize $\mathcal{L}$ with respect to (b,w) and then maximize to $\alpha \geq 0$. To minimize w.r.t (b,w) we have to find derivative of $\mathcal{L}$.

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial b} &= 0 \implies \sum_{n=1}^{N}\alpha_n y_n = 0 \\
\frac{\partial \mathcal{L}}{\partial w} &= 0 \implies w = \sum_{n=1}^{N}\alpha_n y_n x_n \implies w^T = \sum_{n=1}^{N}\alpha_n y_n x_n^T
\end{aligned}$$

Now by using these results we can say that,

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

Then to maximize $\mathcal{L}$ we can equivalently minimize $\mathcal{L}$ and this is subject to constrain $\alpha \geq 0$. Now the derived the dual of the soft SVM is:

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

$$s.t \qquad \sum_{n=1}^{N} y_n \alpha_n = 0 \quad \text{and} \quad C \geq \alpha \geq 0 \quad (n = 1, 2, ...N)$$

Question 2

Proof sketch (N is even)

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

(a) Show $E\left[\| \sum_{n=1}^{N} y_n x_n \|^2\right] = \sum_{n=1}^{N} \sum_{m=1}^{N} y_n y_m x_n^T x_m$.

$$
\begin{aligned}
|\sum_{n=1}^{N} y_n x_n|^2 &= |Xy|^2 \\
&= (Xy)^T (Xy) \\
&= y^T X^T X y \\
&= y^T M y \\
&= y^T \begin{bmatrix} \sum_{m=1}^{N} M_{1m} y_m \\ \dots \\ \sum_{m=1}^{N} M_{Nm} y_m \end{bmatrix} \\
&= \sum_{n=1}^{N} y_n \sum_{m=1}^{N} M_{nm} y_m \\
&= \sum_{n=1}^{N} \sum_{m=1}^{N} y_n y_m M_{nm} \\
&= \sum_{n=1}^{N} \sum_{m=1}^{N} y_n y_m x_n^T x_m
\end{aligned}
$$

Where $M = X^T X$ and $M_{nm} = x_n^T x_m$.

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

$$
\begin{aligned}
E[y_n y_m] = & \quad 1 \qquad m = n \\
= & \frac{1}{N-1} \quad m \neq n
\end{aligned}
$$

Ans When $n = m$, we have $y_n y_m = y_n^2 = 1$. When $n \neq m$, we have $P(y_n = 1) = \frac{1}{2}$ because there are $\frac{N}{2}$ positive points and $\frac{N}{2}$ negative points. We also have $P(y_n = 1 | y_m = 1) = \frac{\frac{N}{2} - 1}{N-1}$ because if we have $y_m = 1$, then we need choose another positive point (there are $\frac{N}{2} - 1$ of them) from the remaining $N - 1$ points.

$$\begin{aligned}
P(y_n y_m = 1) &= P(y_n y_m = 1|y_m = 1)P(y_m = 1) + P(y_n y_m = 1|y_m = -1)P(y_m = -1) \\
&= P(y_n = 1|y_m = 1)P(y_m = 1) + P(y_n = -1|y_m = -1)P(y_m = -1) \\
&= P(y_n = 1|y_m = 1)\frac{1}{2} + P(y_n = -1|y_m = -1)\frac{1}{2} \\
&= \frac{\frac{N}{2} - 1}{N-1}\frac{1}{2} + \frac{\frac{N}{2} - 1}{N-1}\frac{1}{2} \\
&= \frac{\frac{N}{2} - 1}{N-1}
\end{aligned}$$

We thus have $E[y_n y_m] = 1$ when $m = n$ because the when $n = m$, the probability of $P[y_n y_m = 1] = 1$. When $n \neq m$,

$$\begin{aligned}
E[y_n y_m] &= 1P(y_n y_m = 1 \cap n \neq m) - 1P(y_n y_m = -1 \cap n \neq m) \\
&= \frac{\frac{N}{2} - 1}{N-1} - \frac{\frac{N}{2}}{N-1} \\
&= -\frac{1}{N-1}
\end{aligned}$$

(c) Show that

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

Ans

$$\begin{aligned}
E\left[\|\sum_{n=1}^{N} y_n x_n\|^2\right] &= E\left[\sum_{n=1}^{N}\sum_{m=1}^{N} y_n y_m x_n^T x_m\right] \\
&= \sum_{n=1}^{N}\sum_{m=1}^{N} E[y_n y_m] x_n^T x_m \\
&= \sum_{n=1}^{N}\left[\sum_{m\neq n}^{N} E[y_n y_m] x_n^T x_m + E[y_n y_n] x_n^T x_n\right] \\
&= \sum_{n=1}^{N}\left[x_n^T x_n - \sum_{m\neq n}^{N} \frac{1}{N-1} x_n^T x_m\right] \\
&= \sum_{n=1}^{N}\left[x_n^T x_n - \frac{1}{N-1}\sum_{m\neq n}^{N} x_n^T x_m\right] \\
&= \frac{N}{N-1}\sum_{n=1}^{N}\left[\frac{N-1}{N} x_n^T x_n - \frac{1}{N}\sum_{m\neq n}^{N} x_n^T x_m\right] \\
&= \frac{N}{N-1}\sum_{n=1}^{N}\left[x_n^T x_n - \frac{1}{N}\sum_{m=1}^{N} x_n^T x_m\right] \\
&= \frac{N}{N-1}\sum_{n=1}^{N}\left[x_n^T x_n - x_n^T \frac{1}{N}\sum_{m=1}^{N} x_m\right] \\
&= \frac{N}{N-1}\sum_{n=1}^{N}\left[x_n^T x_n - x_n^T \bar{x}\right] \\
&= \frac{N}{N-1}\sum_{n=1}^{N}\left[x_n^T x_n - x_n^T \bar{x} - (\bar{x}^T x_n - \bar{x}^T\bar{x})\right] \\
&= \frac{N}{N-1}\sum_{n=1}^{N}(x_n - \bar{x})^T (x_n - \bar{x}) \\
&= \frac{N}{N-1}\sum_{n=1}^{N}\|x_n - \bar{x}\|^2
\end{aligned}$$

Where in the third to the last equation we have used the fact that $\sum_{n=1}^{N} \bar{x}^T x_n = \sum_{n=1}^{N} \bar{x}^T \bar{x}$

(d) Show that

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

Ans We see that $\sum_{n=1}^{N} \|x_n - \mu\|^2$ achieves minimal value when $\mu = \frac{1}{N}\sum_{n=1}^{N} x_n$, since $\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$, let $\mu = 0$, we have
$\sum_{n=1}^{N} \|x_n - \bar{x}\|^2 \leq \sum_{n=1}^{N} \|x_n\|^2 \leq NR^2$ by the assumption.

(e) Conclude that

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

and hence that

$$P[\|\sum_{n=1}^{N} y_n x_n\| \leq \frac{NR}{\sqrt{N-1}}] > 0$$

Ans. Take the result of problem (d) into the equation of (c), we have

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

Let $Z = \|\sum_{n=1}^{N} y_n x_n\|^2$, and $t = \frac{N^2R^2}{N-1}$, we claim that $P[\sqrt{Z} \leq \sqrt{t}] = P[\|\sum_{n=1}^{N} y_n x_n\| \leq \frac{N^2R^2}{\sqrt{N-1}}] > 0$, otherwise, if $P[\sqrt{Z} \leq \sqrt{t}] = 0$, then $P[\sqrt{Z} > \sqrt{t}] = 1$, we have (since $Z \leq 0$, and $t > 0$)

$$\begin{aligned} E[Z] &= \int_0^{\infty} ZP(dZ) \\ &= \int_t^{\infty} ZP(dZ) \\ &> t\int_t^{\infty} P(dZ) \\ &= t \end{aligned}$$

This contradicts with our conclusion that $E[Z] \leq t$
So $P[\|\sum_{n=1}^{N} y_n x_n\| \leq \frac{NR}{\sqrt{N-1}}] > 0$

VC dimension upper bound

Assume N is even. We need the following geometric fact There exists a (balanced) dichotomy $y_1, ...y_n$ such that

$$\sum_{n=1}^{N} y_n = 0, and \quad \|\sum_{n=1}^{N} y_n x_n\| \leq \frac{NR}{\sqrt{N-1}}$$

The dichotomy satisfying this equation is separated with margin at least p (since $x_1, ..., x_N$ is shattered). So, for some (w, b),
$p\|w\| \leq y_n(w^t x_n + b), forn = 1, ..., N.$

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

By the bound , the RHS is at most $\|w\|\frac{NR}{\sqrt{N-1}}$

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