COL341 Homework 2

- Q1

Solution.

The primal problem corresponding to the soft SVM is

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

$$\text{s.t} \quad y_n(\boldsymbol{\omega}^T\boldsymbol{x}_n + b) \geq 1 - \zeta_n \ \text{ and } \ \zeta_n \geq 0 \quad (n = 1, ...N)$$

The constraints in the standard form are

$$y_n(\boldsymbol{\omega}^T\boldsymbol{x}_n + b) - 1 + \zeta_n \geq 0 \quad (n = 1, ...N)$$

$$\zeta_n \geq 0 \quad (n = 1, ...N)$$

Using the KKT conditions, taking the lagrange multipliers with constraints as $\alpha_n \geq 0$ and $\beta_n \geq 0$ (n = 1,...N) we obtain the optimisation function under the above and lagrange multiplier constraints as follows

$$\max_{\alpha,\beta} \ (\min_{\boldsymbol{\omega},b,\zeta_n} \frac{1}{2}\boldsymbol{\omega}^T\boldsymbol{\omega} + C\sum_{n=1}^{N}\zeta_n - \sum_{n=1}^{N}\alpha_n(y_n(\boldsymbol{\omega}^T\boldsymbol{x}_n + b) - 1 + \zeta_n) - \sum_{n=1}^{N}\beta_n\zeta_n)$$

We define

$$\mathrm{L}(\boldsymbol{\omega}, b, \alpha, \zeta, \beta) = \frac{1}{2}\boldsymbol{\omega}^T\boldsymbol{\omega} + C\sum_{n=1}^{N}\zeta_n - \sum_{n=1}^{N}\alpha_n(y_n(\boldsymbol{\omega}^T\boldsymbol{x}_n + b) - 1 + \zeta_n) - \sum_{n=1}^{N}\beta_n\zeta_n$$

L can be written as

$$\mathrm{L}(\boldsymbol{\omega}, b, \alpha, \zeta, \beta) = \frac{1}{2}\boldsymbol{\omega}^T\boldsymbol{\omega} + \sum_{n=1}^{N}(C - \alpha_n - \beta_n)\zeta_n - \sum_{n=1}^{N}\alpha_n(y_n(\boldsymbol{\omega}^T\boldsymbol{x}_n) - 1) - b\sum_{n=1}^{N}\alpha_n y_n$$

Now we need to apply the following conditions on L

$$\nabla_w L = 0, \ \frac{\partial L}{\partial b} = 0 \ \text{ and } \ \frac{\partial L}{\partial \zeta_n} = 0 \ \ (n = 1, ...N)$$

Applying the first condition

$$\nabla_{\boldsymbol{\omega}} L = \boldsymbol{\omega} - \sum_{n=1}^{N}\alpha_n y_n \boldsymbol{x}_n = 0$$

This gives us

$$\boldsymbol{\omega} = \sum_{n=1}^{N} \alpha_n y_n \boldsymbol{x}_n$$

Applying the second condition

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

This gives us

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

Applying the third condition

$$\frac{\partial L}{\partial \zeta_n} = C - \alpha_n - \beta_n = 0 \quad (n = 1, ...N)$$

This gives us

$$C - \alpha_n - \beta_n = 0 \quad (n = 1, ...N)$$

Substituting these three values into L we get

$$\mathrm{L} = \frac{1}{2}(\sum_{n=1}^{N} \alpha_n y_n \boldsymbol{x}_n)^T(\sum_{n=1}^{N} \alpha_n y_n \boldsymbol{x}_n) + \sum_{n=1}^{N}(0)\zeta_n - \sum_{n=1}^{N} \alpha_n(y_n(\boldsymbol{(}\sum_{n=1}^{N} \alpha_n y_n \boldsymbol{x}_n)^T \boldsymbol{x}_n) - 1) - b.0$$

$$\mathrm{L}(\boldsymbol{\omega}, b, \alpha, \zeta, \beta) = \frac{1}{2}(\sum_{n=1}^{N} \alpha_n y_n \boldsymbol{x}_n^T)(\sum_{n=1}^{N} \alpha_n y_n \boldsymbol{x}_n) - \sum_{n=1}^{N} \alpha_n(y_n(\boldsymbol{(}\sum_{n=1}^{N} \alpha_n y_n \boldsymbol{x}_n^T\boldsymbol{)} \boldsymbol{x}_n) - 1)$$

$$\mathrm{L}(\boldsymbol{\omega}, b, \alpha, \zeta, \beta) = \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n y_n \boldsymbol{x}_n^T \alpha_m y_m \boldsymbol{x}_m - \sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n y_n \boldsymbol{x}_n^T \alpha_m y_m \boldsymbol{x}_m + \sum_{n=1}^{N} \alpha_n$$

$$\mathrm{L}(\boldsymbol{\omega}, b, \alpha, \zeta, \beta) = -\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n y_n \boldsymbol{x}_n^T \alpha_m y_m \boldsymbol{x}_m + \sum_{n=1}^{N} \alpha_n$$

$$\mathrm{L}(\boldsymbol{\omega}, b, \alpha, \zeta, \beta) = -\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} y_n y_m \alpha_n \alpha_m \boldsymbol{x}_n^T \boldsymbol{x}_m + \sum_{n=1}^{N} \alpha_n$$

As we can see, L no longer depends on $\boldsymbol{\omega}, b, \zeta$ or β

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

The optimisation problem now is

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

which is equivalent to

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

under the initial constraints $\alpha_n \geq 0$ and $\beta_n \geq 0$ and the conditions obtained from the above calculation:

$$\sum_{n=1}^{N} \alpha_n y_n = 0 \text{ and } C - \alpha_n - \beta_n = 0 \ (n = 1, ...N)$$

From the condition $(C - \alpha_n - \beta_n = 0,\ (n = 1, ...N))$ we get $(\alpha_n = C - \beta_n,\ (n = 1, ...N))$ and since $(\beta_n \geq 0,\ (n = 1, ...N))$ which means $(C - \beta_n \leq C,\ (n = 1, ...N))$, we get the constraint

$$\alpha_n \leq C,\ (n = 1, ...N)$$

Combining this with the constraint ($\alpha_n \geq 0$, (n=1,...N))we get the new constraint

$$0 \leq \alpha_n \leq C,\ (n = 1, ...N)$$

Since the expression to be minimised does not contain β, it does depend on the constraint $(\beta_n \geq 0 \ (n = 1, ...N))$ Hence, the final optimisation problem is obtained as follows

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

under the constraints

$$\sum_{n=1}^{N} \alpha_n y_n = 0 \text{ and } 0 \leq \alpha_n \leq C \ (n = 1, ...N)$$

- Q2. 1

Proof. By definition $||A||^2 = A^T A$

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

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

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

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

□

- Q2. 2

Proof. If n = m then

$$y_n y_m = y_n y_n = y_n^2$$

y_n can take values +1 and -1. In either case $y_n^2 = 1$. Hence, when n = m then $y_n y_m = 1$.

$$P(y_n y_m = 1 |\ n = m) = 1$$

When n $\neq m$ then we need n and m with y_n and y_m of opposite signs to get $y_n y_m = -1$.

$$P(y_n y_m = -1 |\ n \neq m) = \frac{(\frac{N}{2})(\frac{N}{2}) + (\frac{N}{2})(\frac{N}{2})}{2.\binom{N}{2}}$$

$$P(y_n y_m = -1 |\ n \neq m) = \frac{\frac{N^2}{2}}{2.\frac{N(N-1)}{2}}$$

$$P(y_n y_m = -1 |\ n \neq m) = \frac{N}{2(N-1)}$$

Similarly we can calculate

$$P(y_n y_m = 1 |\ n \neq m) = \frac{2.\binom{\frac{N}{2}}{2} + 2.\binom{\frac{N}{2}}{2}}{2.\binom{N}{2}}$$

$$P(y_n y_m = 1 |\ n \neq m) = \frac{4.\binom{\frac{N}{2}}{2}}{2.\binom{N}{2}}$$

$$P(y_n y_m = 1|\ n \neq m) = \frac{2.\frac{(\frac{N}{2})(\frac{N}{2}-1)}{2}}{\frac{N(N-1)}{2}}$$

$$P(y_n y_m = 1|\ n \neq m) = \frac{2.(\frac{N}{2})(\frac{N}{2}-1)}{N(N-1)}$$

$$P(y_n y_m = 1|\ n \neq m) = \frac{(\frac{N}{2}-1)}{N-1}$$

Using these calculations we can say

$$\text{when n = m: } E[y_n y_m] = 1.P(y_n y_m = 1|\ n = m) + (-1).P(y_n y_m = -1|\ n = m)$$

$$\text{when n = m: } E[y_n y_m] = 1.1 + (-1).0$$

We get, when n = m: $E[y_n y_m] = 1$

Similarly, we can say

$$\text{when n} \neq \text{m: } E[y_n y_m] = 1.P(y_n y_m = 1|\ n \neq m) + (-1).P(y_n y_m = -1|\ n \neq m)$$

$$\text{when n} \neq \text{m: } E[y_n y_m] = 1.\frac{(\frac{N}{2}-1)}{N-1} + (-1).\frac{N}{2(N-1)}$$

$$\text{when n} \neq \text{m: } E[y_n y_m] = 1.\frac{(N-2)}{2(N-1)} + \frac{(-N)}{2(N-1)}$$

$$\text{when n} \neq \text{m: } E[y_n y_m] = \frac{(-2)}{2(N-1)}$$

$$\text{when n} \neq \text{m: } E[y_n y_m] = \frac{-1}{N-1}$$

From this, we conclude

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

□

- Q2. 3

Proof.

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

Using linearity of expectations

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

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

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

Using the expectation values obtained in part 2

$$E\left[\left|\left|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right|\right|^2\right] = \sum_{n=1}^{N}\left(\sum_{m=1,m=n}^{N} 1.\boldsymbol{x}_n^T \boldsymbol{x}_m + \sum_{m=1,m\neq n}^{N} \left(\frac{-1}{N-1}\right) \boldsymbol{x}_n^T \boldsymbol{x}_m\right)$$

$$E\left[\left|\left|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right|\right|^2\right] = \sum_{n=1}^{N}\left(\boldsymbol{x}_n^T \boldsymbol{x}_n + \left(\frac{-1}{N-1}\right) \boldsymbol{x}_n^T \sum_{m=1,m\neq n}^{N} \boldsymbol{x}_m\right)$$

$$E\left[\left|\left|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right|\right|^2\right] = \sum_{n=1}^{N}\left(\boldsymbol{x}_n^T \boldsymbol{x}_n + \left(\frac{-1}{N-1}\right) \boldsymbol{x}_n^T (\sum_{m=1}^{N} \boldsymbol{x}_m - \boldsymbol{x}_n)\right)$$

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

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

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

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

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

Modifying to the form required (add subtract $\bar{\boldsymbol{x}}$ to $\boldsymbol{x}_n^T$)

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

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

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

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

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

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

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

which gives us the required expression

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

$\square$

- Q2. 4

Proof. We first prove the hint that

$$\sum_{n=1}^{N} ||\boldsymbol{x}_n - \boldsymbol{\mu}||^2 \text{ is minimised at } \boldsymbol{\mu} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n$$

Expanding the expression we get

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

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

We minimise the RHS expression by taking its derivative and equating it to 0

$$\nabla_{\boldsymbol{\mu}}(\sum_{n=1}^{N} ||\boldsymbol{x}_n - \boldsymbol{\mu}||^2) = \nabla_{\boldsymbol{\mu}}(\sum_{n=1}^{N} (\boldsymbol{x}_n^T\boldsymbol{x}_n - \boldsymbol{\mu}^T\boldsymbol{x}_n - \boldsymbol{x}_n^T\boldsymbol{\mu} + \boldsymbol{\mu}^T\boldsymbol{\mu})) = 0$$

$$\sum_{n=1}^{N} (\nabla_{\boldsymbol{\mu}}(\boldsymbol{x}_n^T\boldsymbol{x}_n) - \nabla_{\boldsymbol{\mu}}(\boldsymbol{\mu}^T\boldsymbol{x}_n) - \nabla_{\boldsymbol{\mu}}(\boldsymbol{x}_n^T\boldsymbol{\mu}) + \nabla_{\boldsymbol{\mu}}(\boldsymbol{\mu}^T\boldsymbol{\mu})) = 0$$

$$\sum_{n=1}^{N} (0 - (\boldsymbol{x}_n) - (\boldsymbol{x}_n) + 2(\boldsymbol{\mu})) = 0$$

$$-2\sum_{n=1}^{N} \boldsymbol{x}_n + 2\sum_{n=1}^{N} \boldsymbol{\mu} = 0$$

$$-2\sum_{n=1}^{N} \boldsymbol{x}_n + 2N\boldsymbol{\mu} = 0$$

$$2N\boldsymbol{\mu} = 2\sum_{n=1}^{N} \boldsymbol{x}_n$$

$$\boldsymbol{\mu} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n$$

We can confirm that this is the minimum by taking the double derivative of the original expression

$$\nabla_{\boldsymbol{\mu}}^2(\sum_{n=1}^{N} ||\boldsymbol{x}_n - \boldsymbol{\mu}||^2) = \nabla_{\boldsymbol{\mu}}^2(\sum_{n=1}^{N} (\boldsymbol{x}_n^T\boldsymbol{x}_n - \boldsymbol{\mu}^T\boldsymbol{x}_n - \boldsymbol{x}_n^T\boldsymbol{\mu} + \boldsymbol{\mu}^T\boldsymbol{\mu}))$$

$$\nabla^2_{\boldsymbol{\mu}}(\sum_{n=1}^{N} ||\boldsymbol{x}_n - \boldsymbol{\mu}||^2) = \nabla_{\boldsymbol{\mu}}(-2\sum_{n=1}^{N} \boldsymbol{x}_n + 2N\boldsymbol{\mu})$$

$$\nabla^2_{\boldsymbol{\mu}}(\sum_{n=1}^{N} ||\boldsymbol{x}_n - \boldsymbol{\mu}||^2) = -2\sum_{n=1}^{N} 0 + 2N.1$$

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

$$2N > 0$$

As the double derivative is positive, we can confirm that we have obtained the minimum. So, we have proved the hint.

Now using the result obtained we can say that

$$\sum_{n=1}^{N} ||\boldsymbol{x}_n - \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n||^2 \leq \sum_{n=1}^{N} ||\boldsymbol{x}_n - 0||^2 \text{ as the minimum is obtained at } \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n$$

Using $\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n$ we write

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

And from the information given in the question that $||\boldsymbol{x}|| \leq R$, we obtain

$$||\boldsymbol{x}||^2 \leq R^2$$

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

Using this result in the previously obtained inequality we can write

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

which is the result we were trying to prove. $\square$

- Q2. 5

Proof. Using the result obtained in part 4:

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

we can say that

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

Now using this inequality in the result obtained in part 3

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

we can conclude that

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

Using this result we now have to show that

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

Proof (by contradiction):

Let us assume that

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

which is equivalent to saying

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

This means that

$$\left|\left|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right|\right|^2 > \frac{N^2R^2}{N-1}$$

We also know by definition of expectation that

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

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

$$E\left[\left|\left|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right|\right|^2\right] = \left|\left|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right|\right|^2 .0 + \left|\left|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right|\right|^2 .1$$

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

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

which contradicts the fact that

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

Hence our assumption is wrong.

Therefore, we can conclude that

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

This means for some choice of y_n, $||\sum_{n=1}^{N} y_n \boldsymbol{x_n}|| \leq NR/\sqrt{N-1}$ $\square$

- VC Dimension Upper Bound

Proof. The above 5 parts are enough to show that there exist a balanced dichotomy $y_1, ..., y_n$ s.t.

$$\sum_{n=1}^{N} y_n = 0, \text{ and } \left|\left|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right|\right| \leq \frac{NR}{\sqrt{N-1}}$$

As these N points are being shattered, they can be separated by the SVM with margin at least ρ. So, for some $\boldsymbol{\omega}$ and b, we have

$$\rho||\boldsymbol{\omega}|| \leq y_n(\boldsymbol{\omega}^T \boldsymbol{x}_n + b) \ \forall \ n$$

Let us call our VC dimension N, i.e. N is the maximum number of points that can be shattered by the SVM with margin ρ.

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

$$\left|\left|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right|\right|^2 \leq \frac{N^2R^2}{N-1}$$

$$\frac{R^2}{N-1} \geq \frac{\left|\left|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right|\right|^2}{N^2}$$

We keep this expression on hold for now.

As we had seen

$$\rho||\boldsymbol{\omega}|| \leq y_n(\boldsymbol{\omega}^T \boldsymbol{x}_n + b) \ \forall \ n$$

Since this is true for every n, summing this over all n we get

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

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

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

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

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

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

$$N\rho||\boldsymbol{\omega}|| \leq \ < \boldsymbol{\omega}, \sum_{n=1}^{N} y_n \boldsymbol{x}_n >$$

Squaring both sides

$$N^2\rho^2||\boldsymbol{\omega}||^2 \leq \ \ (< \boldsymbol{\omega}, \sum_{n=1}^{N} y_n \boldsymbol{x}_n >)^2$$

Using the Cauchy Schwartz inequality

$$N^2\rho^2||\boldsymbol{\omega}||^2 \leq \ \ ||\boldsymbol{\omega}||^2 \left|\left|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right|\right|^2$$

$$N^2\rho^2 \leq \left|\left|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right|\right|^2$$

$$\rho^2 \leq \frac{\left|\left|\sum_{n=1}^{N} y_n \boldsymbol{x}_n\right|\right|^2}{N^2}$$

Substituting this into the expression we had put on hold, we get

$$\frac{R^2}{N-1} \geq \rho^2$$

$$\frac{R^2}{\rho^2} \geq N-1$$

$$N-1 \leq \frac{R^2}{\rho^2}$$

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

Also

$$\frac{R^2}{\rho^2} \leq \left\lceil \frac{R^2}{\rho^2} \right\rceil$$

We do this because N is an integer. Hence, we get the final result:

$$N \leq \left\lceil \frac{R^2}{\rho^2} \right\rceil + 1$$

which is the required bound on the VC dimension of the SVM with margin ρ

$$d_{vc(\rho)} \leq \left\lceil \frac{R^2}{\rho^2} \right\rceil + 1$$

$\square$