CMPE 544: Pattern Recognition (Fall 2020)

Homework3 — Yaşar Alim Türkmen - 2019700123

1. (a) The hard-margin SVM can be written as

$$\min_{b,\mathrm{w}} \quad \frac{1}{2}\mathrm{w}^T\mathrm{w}$$

$$\text{subject to:} \quad y_n(\mathrm{w}^T X_n + b) \geq 1 \;\; n = 1, \ldots, N$$

After constructing the Lagrangian and rewriting w and b in terms of α we get

$$\mathcal{L}(\alpha) = -\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} y_n y_m \alpha_n \alpha_m X_n^T X_m + \sum_{n=1}^{N}\alpha_n$$

If we substitute inner product with kernel function ($\mathcal{K}(x_1, x_2) = (x_1^T x_2 + 1)^2$), we obtain the dual problem as

$$\max_{\alpha\in\mathbf{R}^4} \quad \sum_{n=1}^{4}\alpha_n - \frac{1}{2}\sum_{n=1}^{4}\sum_{m=1}^{4} y_n y_m \alpha_n \alpha_m (X_n^T X_m + 1)^2$$

$$\text{subject to:} \quad \sum_{n=1}^{4} y_n \alpha_n = 0$$

$$\alpha_n \geq 0 \;\; n = 1, \ldots, 4$$

(b) If we apply kernel function for each vector pairs we obtain

$$K = \begin{bmatrix} (X_1^T X_1 + 1)^2 & (X_1^T X_2 + 1)^2 & (X_1^T X_3 + 1)^2 & (X_1^T X_4 + 1)^2 \\ (X_2^T X_1 + 1)^2 & (X_2^T X_2 + 1)^2 & (X_2^T X_3 + 1)^2 & (X_2^T X_4 + 1)^2 \\ (X_3^T X_1 + 1)^2 & (X_3^T X_2 + 1)^2 & (X_3^T X_3 + 1)^2 & (X_3^T X_4 + 1)^2 \\ (X_4^T X_1 + 1)^2 & (X_4^T X_2 + 1)^2 & (X_4^T X_3 + 1)^2 & (X_4^T X_4 + 1)^2 \end{bmatrix}$$

$$K = \begin{bmatrix} (1+1+1)^2 & (1-1+1)^2 & (-1-1+1)^2 & (-1+1+1)^2 \\ (1-1+1)^2 & (1+1+1)^2 & (-1+1+1)^2 & (-1-1+1)^2 \\ (-1-1+1)^2 & (-1+1+1)^2 & (1+1+1)^2 & (1-1+1)^2 \\ (-1+1+1)^2 & (-1-1+1)^2 & (1-1+1)^2 & (1+1+1)^2 \end{bmatrix} = \begin{bmatrix} 9 & 1 & 1 & 1 \\ 1 & 9 & 1 & 1 \\ 1 & 1 & 9 & 1 \\ 1 & 1 & 1 & 9 \end{bmatrix}$$

A matrix is positive semi-definite if it is symmetric and all of its eigenvalues are greater than or equal to 0. K is symmetric and we can find K's eigenvalues by solving $\det(K - \lambda I) = 0$.

$$\begin{vmatrix} 9-\lambda & 1 & 1 & 1 \\ 1 & 9-\lambda & 1 & 1 \\ 1 & 1 & 9-\lambda & 1 \\ 1 & 1 & 1 & 9-\lambda \end{vmatrix} = 0.$$

After calculating this using numpy, eigenvalues are found as 8, 12, 8 and 8. Since all of the eigeenvalues of K is positive, we can say that K is positive semi-definite. Therefore, it satisfies Mercer's condition.

(c) From part(a), $\sum_{n=1}^{N} y_n\alpha_n = 0$. Therefore, $y_1\alpha_1 + y_2\alpha_2 + y_3\alpha_3 + y_4\alpha_4 = 0$. If we place y_n values, $\alpha_1 - \alpha_2 + \alpha_3 - \alpha_4 = 0$. Hence,

$$\alpha_1 + \alpha_3 = \alpha_2 + \alpha_4 \tag{1}$$

We try to maximize $\mathcal{L}(\alpha)$ with respect to α.

$$\mathcal{L}(\alpha) = \sum_{n=1}^{4}\alpha_n - \frac{1}{2}\sum_{n=1}^{4}\sum_{m=1}^{4} y_n y_m \alpha_n \alpha_m (X_n^T X_m + 1)^2$$

$$= (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) - \frac{1}{2}\sum_{n=1}^{4}\sum_{m=1}^{4} y_n y_m \alpha_n \alpha_m K_{n,m}$$

By using the kernel matrix we found in part (b),

$$\sum_{n=1}^{4}\sum_{m=1}^{4} y_n y_m \alpha_n \alpha_m K_{n,m} = 9\alpha_1^2 - 2\alpha_1\alpha_2 + 2\alpha_1\alpha_3 - 2\alpha_1\alpha_4 + 9\alpha_2^2 - 2\alpha_2\alpha_3 + 2\alpha_2\alpha_4 + 9\alpha_3^2 - 2\alpha_3\alpha_4 + 9\alpha_4^2$$

$$= 8(\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2) + (\alpha_1 - \alpha_2 + \alpha_3 - \alpha_4)^2$$

If we combine these, the problem becomes maximizing $\mathcal{L}(\alpha)$ where

$$\mathcal{L}(\alpha) = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) - \frac{1}{2}(8(\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2) + (\alpha_1 - \alpha_2 + \alpha_3 - \alpha_4)^2)$$

. In order to find, maximizing α values, we can use partial derivatives.

$$\frac{\partial \mathcal{L}}{\partial \alpha_1} = 1 - 8\alpha_1 - (\alpha_1 - \alpha_2 + \alpha_3 - \alpha_4) = 0 \tag{2}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_2} = 1 - 8\alpha_2 - (\alpha_1 - \alpha_2 + \alpha_3 - \alpha_4) = 0 \tag{3}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_3} = 1 - 8\alpha_3 - (\alpha_1 - \alpha_2 + \alpha_3 - \alpha_4) = 0 \tag{4}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_4} = 1 - 8\alpha_4 - (\alpha_1 - \alpha_2 + \alpha_3 - \alpha_4) = 0 \tag{5}$$

After combining equation (2) and (4) and equation (3) and (5), we can conclude that for optimum α values $\alpha_1 = \alpha_3$ and $\alpha_2 = \alpha_4$. Also, from equation (1) we know that $\alpha_1 + \alpha_3 = \alpha_2 + \alpha_4$. Hence, in order to maximize $\mathcal{L}(\alpha)$, all α values should be equal to each other. $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha^*$. We can use equation (2) to find a^* by rewriting it as

$$\frac{\partial \mathcal{L}}{\partial \alpha_1} = 1 - 8\alpha^* - (\alpha^* - \alpha^* + \alpha^* - \alpha^*) = 0$$

$$\alpha^* = \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \frac{1}{8} \tag{6}$$

(d) The second order polynomial non-linear transformation can be written as

$$\phi(\mathrm{x}) = \begin{bmatrix} 1 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \end{bmatrix}$$

$$\phi(X_1) = \begin{bmatrix} 1 \\ -\sqrt{2} \\ -\sqrt{2} \\ 1 \\ 1 \\ \sqrt{2} \end{bmatrix}, \phi(X_2) = \begin{bmatrix} 1 \\ -\sqrt{2} \\ \sqrt{2} \\ 1 \\ 1 \\ -\sqrt{2} \end{bmatrix}, \phi(X_3) = \begin{bmatrix} 1 \\ \sqrt{2} \\ \sqrt{2} \\ 1 \\ 1 \\ \sqrt{2} \end{bmatrix}, \phi(X_4) = \begin{bmatrix} 1 \\ \sqrt{2} \\ -\sqrt{2} \\ 1 \\ 1 \\ -\sqrt{2} \end{bmatrix}$$

(e) The optimal weight can be found by

$$\mathrm{w}^* = \sum_{n=1}^{4} y_n \alpha_n^* \phi(X_n)$$

In order to find optimal w, we can use the outputs obtained in part(c) and part(d).

$$\mathrm{w}^* = y_1\alpha_1\phi(X_1) + y_2\alpha_2\phi(X_2) + y_3\alpha_3\phi(X_3) + y_4\alpha_4\phi(X_4)$$

$$\mathrm{w}^* = \frac{1}{8}\begin{bmatrix} 1 \\ -\sqrt{2} \\ -\sqrt{2} \\ 1 \\ 1 \\ \sqrt{2} \end{bmatrix} - \frac{1}{8}\begin{bmatrix} 1 \\ -\sqrt{2} \\ \sqrt{2} \\ 1 \\ 1 \\ -\sqrt{2} \end{bmatrix} + \frac{1}{8}\begin{bmatrix} 1 \\ \sqrt{2} \\ \sqrt{2} \\ 1 \\ 1 \\ \sqrt{2} \end{bmatrix} - \frac{1}{8}\begin{bmatrix} 1 \\ \sqrt{2} \\ -\sqrt{2} \\ 1 \\ 1 \\ -\sqrt{2} \end{bmatrix}$$

$$\mathrm{w}^* = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

For finding optimal b, we can use the following equation:

$$b^* = y_s - \sum_{n=1}^{4} y_n \alpha_n^* \mathcal{K}(X_n^T, X_s), \text{ since } y_s(\mathrm{w}^{*^T}\phi(X_s) + b^*) = 1.$$

where s for support vector and $\alpha_s^* \geq 0$. In this problem any of x vectors can be used as a support vector since all of α values are greater than 0. Let $s = 1$,

$$y_1(\mathrm{w}^{*^T}\phi(X_1) + b^*) = 1.$$

$$1(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}^T \begin{bmatrix} 1 \\ -\sqrt{2} \\ -\sqrt{2} \\ 1 \\ 1 \\ \sqrt{2} \end{bmatrix} + b^*) = 1$$

The inner product gives 1, and the equations reduces to $1(1 + b^*) = 1$ which gives $b^* = 0$.