Ahmad Fallahpour

1 Kernels

1.1

According to mercer theorem, for any N following matrix should be positive semidefinite.

$$K = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_N) \\ k(x_2, x_1) & k(x_2, x_2) & \dots & k(x_2, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_N, x_1) & k(x_N, x_2) & \dots & k(x_N, x_N) \end{bmatrix} \tag{1}$$

Based on the the given kernel function and by picking distinct points i.e. $x_i \neq x_j$ if $i \neq j$, matrix K is simplify as follow:

$$K = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \tag{2}$$

Which is the identity matrix. In other words, it has N eigenvalue that all of them are one meaning that this matrix is positive semidefinite.

1.2

$$\frac{\partial J(\alpha)}{\partial \alpha} = K^2\alpha - Ky = 0 \tag{3}$$

$$\alpha^* = K^{-1}y \tag{4}$$

$$\omega^* = \Phi^T\alpha^* \tag{5}$$

Prediction for new data point x:

$$f(x) = (\omega^*)^T\Phi(x) = y^TK^{-1}\Phi^T\Phi(x) = y^TK^{-1}\begin{pmatrix} k(x, x_1) \\ k(x, x_2) \\ \vdots \\ k(x, x_N) \end{pmatrix} \tag{6}$$

Since $K = I$ we can rewrite the prediction for data point x as follow:

$$f(x) = [k(x, x_1), k(x, x_2), \dots, k(x, x_N)]y \tag{7}$$

According to eq. (7) for any training data point x_i the prediction $f(x_i) = y_i$

$$f(x) = [k(x_i, x_1), k(x_i, x_2), \dots, k(x_i, x_i), \dots, k(x_i, x_N)]y = [0, 0, \dots, 1, \dots, 0]y = y_i \tag{8}$$

1.3

According to eq. (7) if a new data point x_i is not equal to any given training data points, then $k(x_i, x_n) = 0$ so we have:

$$f(x) = [k(x_i, x_1), k(x_i, x_2), \dots, k(x_i, x_N)]y = [0, 0, \dots, 0]y = 0 \tag{9}$$

So, the prediction is always 0.

2 Support Vector Machines

2.1

No. They cannot be separated with a linear separator, because they are in a line and one point form the first class is located between of two points from other class.

2.2

Yes. As it can be seen in the following plot, with two dimensional feature, traning data points are not in the same line any more and they can be separated with suitable decision boundary.

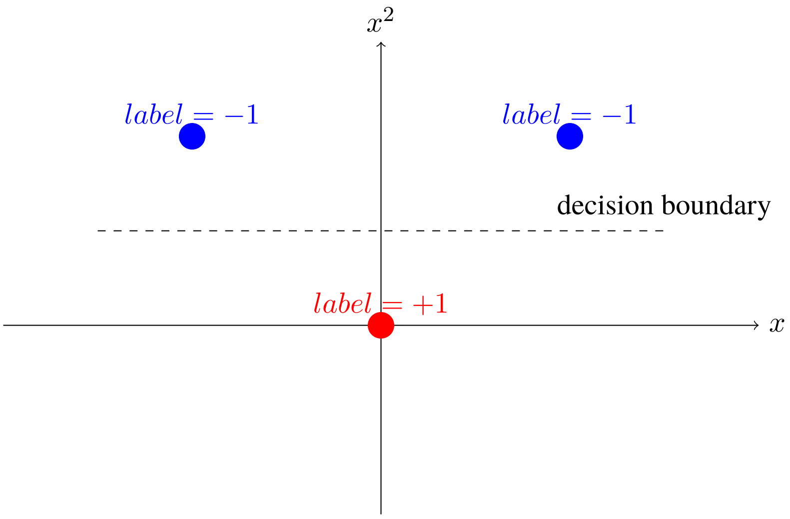

Figure 1: New feature space.

2.3

$$k(x, x') = \Phi(x)^T \Phi(x') = (x, x^2) \begin{pmatrix} x' \\ x'^2 \end{pmatrix} = xx' + (xx')^2 \tag{10}$$

$$K = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & k(x_1, x_3) \\ k(x_2, x_1) & k(x_2, x_2) & k(x_2, x_3) \\ k(x_3, x_1) & k(x_3, x_2) & k(x_3, x_3) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{11}$$

As you can see, K is a diagonal matrix. So, the eigenvalues are the entries on the main diagonal that all greater equal to zero. As a result, K is PSD.

2.4

There are three data points so $N = 3$ and $\Phi(x)$ is two dimensional; so, ω is two dimensional as $\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix}$

Primal formulation:

$$\begin{aligned}
\min_{\omega_1,\omega_2,b,\xi_1,\xi_2,\xi_3} \quad & C(\xi_1+\xi_2+\xi_3)+\frac{1}{2}(\omega_1^2+\omega_2^2)\\
\text{s.t.} \quad & 1-y_1(\omega_1 x_1+\omega_2 x_1^2+b)\leq \xi_1 => 1-\omega_1+\omega_2+b\leq \xi_1\\
& 1-y_2(\omega_1 x_2+\omega_2 x_2^2+b)\leq \xi_2 => 1+\omega_1+\omega_2+b\leq \xi_2\\
& 1-y_3(\omega_1 x_3+\omega_2 x_3^2+b)\leq \xi_3 => 1-b\leq \xi_3\\
& \xi_1\geq 0,\xi_2\geq 0,\xi_3\geq 0
\end{aligned}$$

Dual formulation:

$$\begin{aligned}
\max_{\alpha_1,\alpha_2,\alpha_3} \quad & \sum_{n=1}^{3}\alpha_n-\frac{1}{2}\sum_{n=1}^{3}\sum_{m=1}^{3}\alpha_n\alpha_m y_n y_m k(x_n,x_m)\\
\text{s.t.} \quad & 0\leq \alpha_1\leq +\infty\\
& 0\leq \alpha_2\leq +\infty\\
& 0\leq \alpha_3\leq +\infty\\
& \alpha_1 y_1+\alpha_2 y_2+\alpha_3 y_3=0
\end{aligned}$$

2.5

$$\begin{aligned}
\max_{\alpha_1,\alpha_2,\alpha_3} \quad & \alpha_1+\alpha_2+\alpha_3-\frac{1}{2}\big(\alpha_1^2 y_1^2 k(x_1,x_1)+\alpha_2^2 y_2^2 k(x_2,x_2)\big)\\
\text{s.t.} \quad & 0\leq \alpha_1\leq +\infty\\
& 0\leq \alpha_2\leq +\infty\\
& 0\leq \alpha_3\leq +\infty\\
& \alpha_1 y_1+\alpha_2 y_2+\alpha_3 y_3=0
\end{aligned} \tag{12}$$

$$\begin{aligned}
\max_{\alpha_1,\alpha_2,\alpha_3} \quad & \alpha_1+\alpha_2+\alpha_3-\frac{1}{2}\big(\alpha_1^2+\alpha_2^2\big)\\
\text{s.t.} \quad & 0\leq \alpha_1\leq +\infty\\
& 0\leq \alpha_2\leq +\infty\\
& 0\leq \alpha_3\leq +\infty\\
& -\alpha_1-\alpha_2+\alpha_3=0
\end{aligned} \tag{13}$$

Due to symmetry, we have $\alpha_1 = \alpha_2$. Then, the equations simplify as follow:

$$\begin{aligned}
\max_{\alpha_1,\alpha_2,\alpha_3} \quad & 2\alpha_1+\alpha_3-\alpha_1^2\\
\text{s.t.} \quad & 0\leq \alpha_1\leq +\infty\\
& 0\leq \alpha_3\leq +\infty\\
& \alpha_3=2\alpha_1
\end{aligned} \tag{14}$$

$$\begin{aligned}
\max_{\alpha_1} \quad & 4\alpha_1-\alpha_1^2\\
\text{s.t.} \quad & 0\leq \alpha_1
\end{aligned} \tag{15}$$

The objective function $4\alpha_1 - \alpha_1^2$ is maximized when $\alpha_1 = 2$ which leads to $\alpha_2 = 2$ and $\alpha_3 = 4$

$$\alpha = \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix} \tag{16}$$

$$\omega = \sum_{n=1}^{3} \alpha_n y_n \phi(x_n) = -2 \begin{pmatrix} -1 \\ 1 \end{pmatrix} - 2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -4 \end{pmatrix} \tag{17}$$

$$b = y_1 - \sum_{n=1}^{3} y_n k(x_n, x_1) = -1 - (-1)(2)(1) = 1 \tag{18}$$

2.6

$$\omega^T \phi(x) + b = 0 => (0, -4) \begin{pmatrix} x \\ x^2 \end{pmatrix} + 1 = -4x^2 + 1 = 0 => x^2 = \frac{1}{4} \tag{19}$$

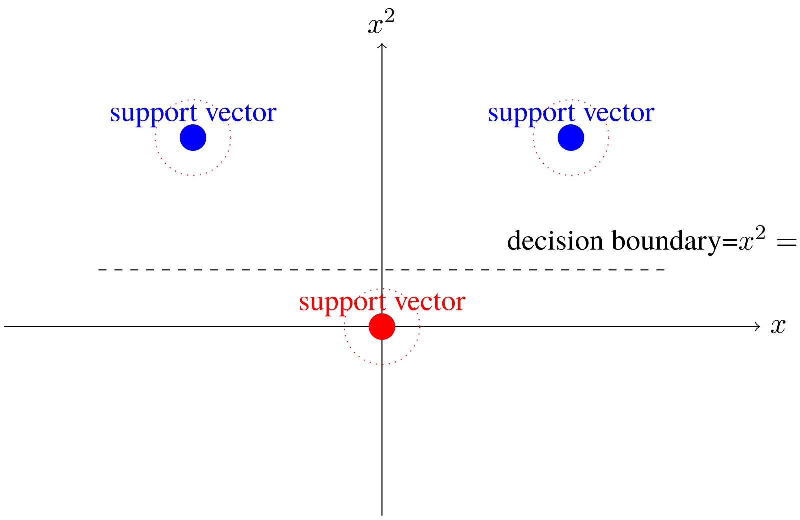

Figure 2: Two dimensional feature space.

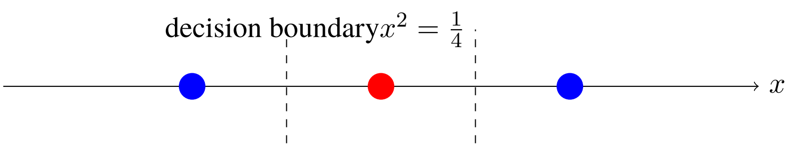

Figure 3: Original one dimensional feature space.

3 Adaboost for building up a nonlinear classifier

3.1

$$(s, b, d) = (1, 0.5, 1) \tag{20}$$

$$\epsilon_1 = 0.5 \tag{21}$$

$$\beta_1 = 0 \tag{22}$$

3.2

$$\omega_2(n) = \omega_1(n)\exp(0) = \omega_1(n) = 0.25, \forall n \tag{23}$$

3.3

$$(s, b, d) = (1, 0.5, 2) \tag{24}$$

$$\epsilon_1 = 0.25 \tag{25}$$

$$\beta_1 = 0.55 \tag{26}$$

3.4

$$\omega_2(1) = \omega_2(2) = \omega_2(4) = 0.17, \omega_2(3) = 0.50, \tag{27}$$

$$(s, b, d) = (-1, 0.5, 1) \tag{28}$$

$$\epsilon_2 = 0.17 \tag{29}$$

$$\beta_2 = 0.79 \tag{30}$$

3.5

$$\omega_3(1) = \omega_3(4) = 0.10, \omega_3(2) = 0.50, \omega_3(3) = 0.30 \tag{31}$$

$$(s, b, d) = (1, -0.5, 1) \tag{32}$$

$$\epsilon_3 = 0.10 \tag{33}$$

$$\beta_3 = 1.10 \tag{34}$$

3.6

$$F(x) = sign[0.55 \times h_{(1,0.5,2)}(x) + 0.79 \times h_{(-1,0.5,1)}(x) + 1.10 \times h_{(1,-0.5,1)}(x)] \tag{35}$$

$$F(x1) = sign[0.55 \times 1 + 0.79 \times 1 + 1.10 \times 1] = 1 \tag{36}$$

$$F(x2) = sign[0.55 \times -1 + 0.79 \times 1 + 1.10 \times -1] = -1 \tag{37}$$

$$F(x3) = sign[0.55 \times -1 + 0.79 \times 1 + 1.10 \times 1] = 1 \tag{38}$$

$$F(x4) = sign[0.55 \times -1 + 0.79 \times -1 + 1.10 \times 1] = -1 \tag{39}$$

All four are classified correctly.