CSCI 567 Homework 2

Shamim Samadi

October 23, 2017

Problem 1.

- **Question 1.1** Mercer's theorem: need to prove that kernel matrix K is positive semi definite:

$$K = \begin{bmatrix} k(x_1,x_1) & k(x_1,x_2) & ... & ... & k(x_1,x_N) \\ k(x_2,x_1) & k(x_2,x_2) & ... & ... & k(x_2,x_N) \\ ... & ... & ... & ... & ... \\ k(x_N,x_1) & ... & ... & ... & k(x_N,x_N) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & ... & 0 \\ 0 & 1 & ... & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & ... & ... & 1 \end{bmatrix} = I(\text{identity})$$

assume $z \in R^N$, need to show for every z, we have: $z^T K z >= 0$.

$$z^T K z = z^T I z = z^T z = ||z||^2 >= 0 \tag{1}$$

Thus k is a valid kernel.

- **Question 1.2**
$K = I$, thus:

$$J(\alpha) = \frac{1}{2}\alpha^T I \alpha - y^T \alpha + \frac{\lambda}{2}\alpha^T \alpha + \frac{1}{2} y^T y \tag{2}$$

since $\lambda = 0$:

$$J(\alpha) = \frac{1}{2}\alpha^T \alpha - y^T \alpha + \frac{1}{2} y^T y \tag{3}$$

$$\frac{\partial J(\alpha)}{\partial \alpha} = 0 \rightarrow \alpha^T - y^T = 0 \rightarrow \alpha^T = y^T \rightarrow \alpha^* = y \tag{4}$$

and we have:

$$J(\alpha^*) = \frac{1}{2} y^T y - y^T y + \frac{1}{2} y^T y = 0 \tag{5}$$

- **Question 1.3**

$$f(x) = [k(x,x_1), k(x,x_2), ..., k(x,x_N)]\alpha^* \tag{6}$$

according to the kernel definition and since $x \neq x_n, \forall n = 1, 2, ..., N$, we have:

$k(x, x_n) = 0 \forall n = 1, 2, ..., N$,

thus:

$f(x) = 0.\alpha^* = 0$.

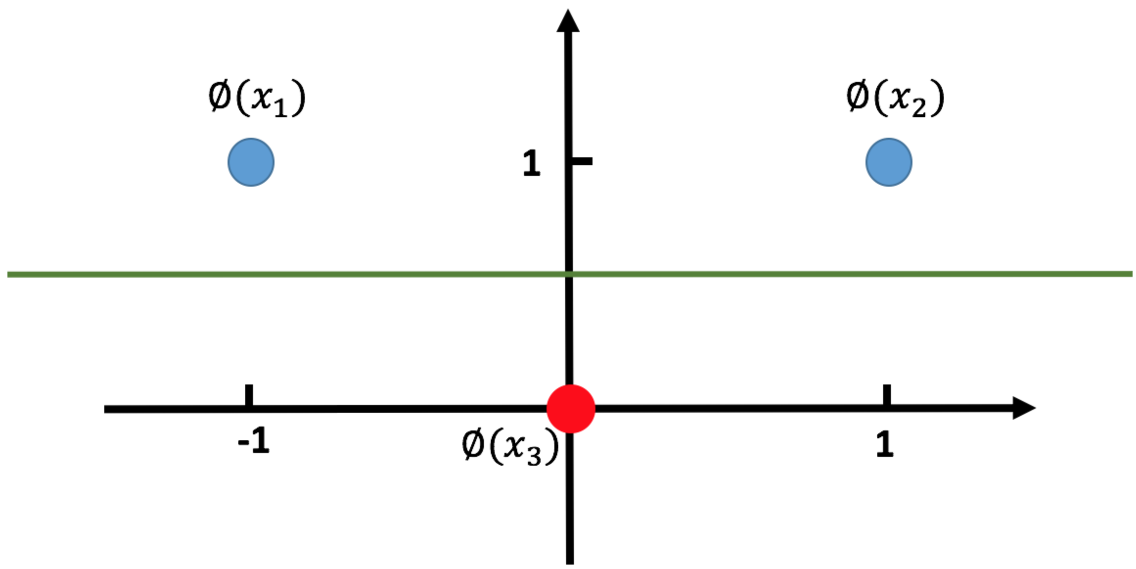

Figure 1: Transformed data points for question 2.2

Problem 2. • **Question 2.1**

No, they cannot because there exists no linear separator (no vertical line) that would separate the one-dimensional space into two parts with both blue dots (class -1) lying on the same side.

- **Question 2.2**

  See Figure 1 for the plot. In the new (transformed) space, there exists a linear separator such that points from the blue class lie on the same side (example separator drawn in green).

- **Question 2.3** • kernel function:

$$\phi(x_m)^T\phi(x_n) = \begin{bmatrix} x_m \\ {x_m}^2 \end{bmatrix}^T \begin{bmatrix} x_n \\ {x_n}^2 \end{bmatrix} = x_m x_n + {x_m}^2{x_n}^2 = x_m x_n + x_m{x_n}^2 \tag{7}$$

thus kernel function is: $k(x, x') = xx' + (xx')^2$.

• kernel matrix:

$$K = \begin{bmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{bmatrix} = \begin{bmatrix} k(x_1,x_1) & k(x_1,x_2) & k(x_1,x_3) \\ k(x_2,x_1) & k(x_2,x_2) & k(x_2,x_3) \\ k(x_3,x_1) & k(x_3,x_2) & k(x_3,x_3) \end{bmatrix} \tag{8}$$

where we have the following:

$$K_{11} = \phi(x_1)^T \phi(x_1) = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = 2 \tag{9}$$

$$K_{12} = K_{21} = \phi(x_1)^T \phi(x_2) = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \tag{10}$$

$$K_{13} = K_{31} = \phi(x_1)^T \phi(x_3) = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0 \tag{11}$$

$$K_{22} = \phi(x_2)^T \phi(x_2) = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2 \tag{12}$$

$$K_{23} = K_{32} = \phi(x_2)^T \phi(x_3) = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0 \tag{13}$$

$$K_{33} = \phi(x_3)^T \phi(x_3) = \begin{bmatrix} 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0 \tag{14}$$

Thus

$$K = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{15}$$

- show K is PSD:

assume $z \in R^3$, show $z^T K z \geq 0$ for every z:

$$z^T K z = \begin{bmatrix} z_1 & z_2 & z_3 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \tag{16}$$

$$z^T K z = 2{z_1}^2 + 2{z_2}^2 + 0 = 2({z_1}^2 + {z_2}^2) \geq 0 \tag{17}$$

thus K is PSD.

- **Question 2.4** SVM formulations:
  - primal

$$\min_{w,b} C \sum_n max(0, 1 - y_n[w^T \phi(x_n) + b]) + \frac{1}{2}||w||^2 \tag{18}$$

$$\min_{w,b} C[max(0, 1 - (-1)(-w_1 + w_2 + b)) + max(0, 1 - (-1)(w_1 + w_2 + b)) + max(0, 1 - (1)(b))] + \frac{1}{2}||w||^2 \tag{19}$$

$$\min_{w,b} C[max(0, 1 - w_1 + w_2 + b) + max(0, 1 + w_1 + w_2 + b)) + max(0, 1 - b)] + \frac{1}{2}({w_1}^2 + {w_2}^2) \tag{20}$$

- dual

$$\text{Max } \sum_n \alpha_n - \tfrac{1}{2}\sum_{m,n} y_m y_n \alpha_m \alpha_n k(x_m, x_n) = \text{Max } \sum_n \alpha_n - \tfrac{1}{2}\sum_{m,n} y_m y_n \alpha_m \alpha_n (x_m x_n + {x_m}^2 {x_n}^2)$$

subject to: $0 \le \alpha_n \le C(\forall n), \sum_n \alpha_n y_n = 0$

$$\begin{aligned} \text{Max } & \textstyle\sum_{n=1,2,3} \alpha_n - \tfrac{1}{2}(2{\alpha_1}^2 + 2{\alpha_2}^2) \\ \text{s.t. } & 0 \le \alpha_n \le C, \forall n \\ & \textstyle\sum_{n=1,2,3} \alpha_n y_n = 0 \end{aligned}$$

which simplifies to:

$$\max_{\alpha} \alpha_1 + \alpha_2 + \alpha_3 - {\alpha_1}^2 - {\alpha_2}^2 \tag{21}$$

subject to the following:
$0 \le \alpha_1 \le C, 0 \le \alpha_2 \le C, 0 \le \alpha_3 \le C$
and
$-\alpha_1 - \alpha_2 + \alpha_3 = 0$

- **Question 2.5**

solve the dual - using lagrange multipliers:

$$L = \alpha_1 + \alpha_2 + \alpha_3 - {\alpha_1}^2 - {\alpha_2}^2 - \lambda(-\alpha_1 - \alpha_2 + \alpha_3) \tag{22}$$

$$\frac{\partial L}{\partial \alpha_1} = 1 - 2\alpha_1 + \lambda = 0 \tag{23}$$

$$\frac{\partial L}{\partial \alpha_2} = 1 - 2\alpha_2 + \lambda = 0 \tag{24}$$

$$\frac{\partial L}{\partial \alpha_3} = 1 - \lambda = 0 \to \lambda = 1 \tag{25}$$

$$\frac{\partial L}{\partial \lambda} = 0 \to -\alpha_1 - \alpha_2 + \alpha_3 = 0 \tag{26}$$

from which we get: $\alpha_1 = \alpha_2 = 1$ and $\alpha_3 = 2$: $\alpha = [1 \quad 1 \quad 2]^T$

$$w = \sum_n \alpha_n y_n \phi(x_n) = \alpha_1 y_1 \phi(x_1) + \alpha_2 y_2 \phi(x_2) + \alpha_3 y_3 \phi(x_3) = [0 \quad -2]^T \tag{27}$$

All points satisfy: $0 \le \alpha_n \le C$. Thus can use any to compute the bias:

$$b = y_n - w^T \phi(x_n) = 1 \tag{28}$$

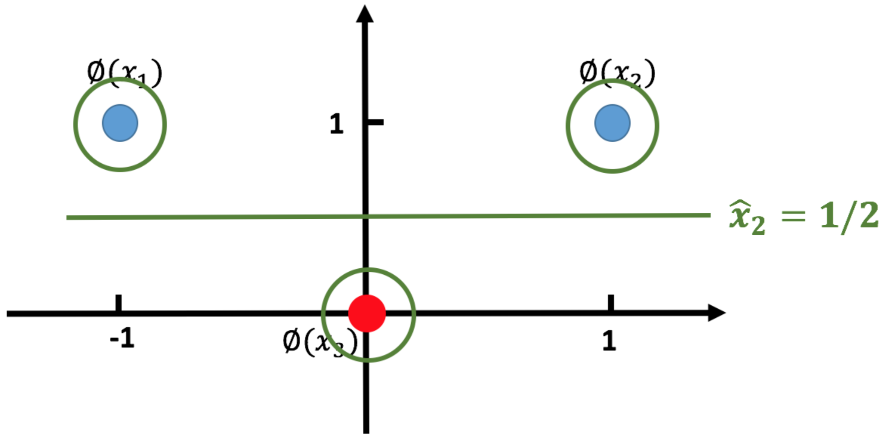

Figure 2: Linear decision boundary for kernel-SVM (question 2.6) - support vectors are circled in green.

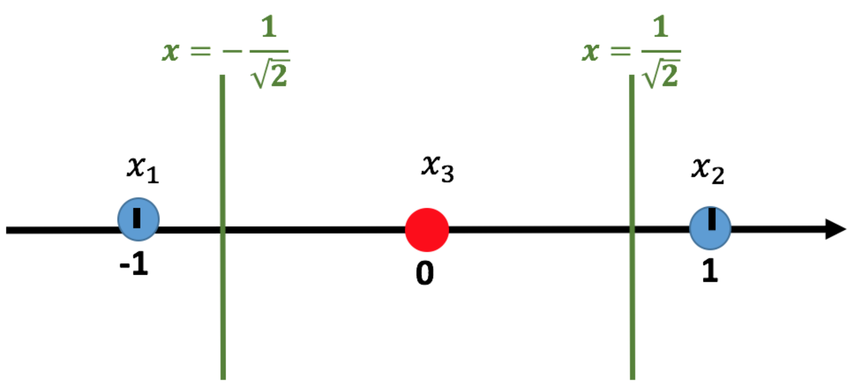

Figure 3: Nonlinear decision boundary in the original 1-D space (question 2)

- **Question 2.6**

  - transformed 2D space:

$$\hat{y} = w^T\phi(x) + b = \begin{bmatrix} 0 & -2 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} \tag{29}$$

$$\hat{y} = -2\hat{x}_2 + 1 = 0 \rightarrow \hat{x}_2 = \frac{1}{2} \tag{30}$$

  All α_i's are strictly positive, and thus all points are support vectors.

  - original 1-D space:

  Using: $\hat{x}_2 = x^2$, thus: $-2x^2 + 1 = 0 \rightarrow x^2 = \frac{1}{2} \rightarrow x = \pm\frac{1}{\sqrt{2}}$

Problem 3.

- **Question 3.1** $f_1 : (s, b, d) = (1, -0.5, 1)$

  $w_1(1) = w_1(2) = w_1(3) = w_1(4) = 0.25$, $\epsilon_1 = 0.5$, $\beta_1 = \frac{1}{2} log \frac{0.5}{0.5} = 0$

- **Question 3.2** since $\beta_1 = 0$, $e^{-\beta_1} = e^{\beta_1} = 1$, thus: $w_2(n) = w_1(n)$

  $w_2(1) = w_2(2) = w_2(3) = w_2(4) = 0.25$.

- **Question 3.3**

  $f_1 : (s, b, d) = (1, -0.5, 1)$

  $\epsilon_1 = 0.25$, $\beta_1 = \frac{1}{2} log \frac{1-0.25}{0.25} = 0.55$

- **Question 3.4** $f_2 : (s, b, d) = (1, 0.5, 2)$

  $w_2(1) = w_2(2) = w_2(3) = 0.25e^{-0.55} = 0.14$
  $w_2(4) = 0.25e^{0.55} = 0.43$

  after normalization we get:

  $w_2(1) = w_2(2) = w_2(3) = 0.25e^{-0.55} = 0.16$
  $w_2(4) = 0.25e^{0.55} = 0.51$

  $\epsilon_2 = 0.16$, $\beta_2 = \frac{1}{2} log \frac{1-0.16}{0.16} = 0.83$

- **Question 3.5**

  $f_3 : (s, b, d) = (-1, -0.5, 2)$

  $w_3(1) = w_3(2) = 0.16e^{-0.83} = 0.07$
  $w_3(3) = 0.16e^{0.83} = 0.37$
  $w_3(4) = 0.51e^{-0.83} = 0.22$

  after normalization we get:

  $w_3(1) = w_3(2) = 0.16e^{-0.83} = 0.1$
  $w_3(3) = 0.16e^{0.83} = 0.51$
  $w_3(4) = 0.51e^{-0.83} = 0.3$

  $\epsilon_3 = 0.1$, $\beta_3 = \frac{1}{2} log \frac{1-0.1}{0.1} = 1.1$

- **Question 3.6**

$$F(x) = Sign[0.55 \times h_{(+1,-0.5,1)}(x) + 0.83 \times h_{(+1,0.5,2)(x)} + 1.1 \times h_{(-1,-0.5,2)}(x)] \quad (31)$$

$$F(x_1) = Sign(0.55 + 0.83 - 1.1) = Sign(0.28) = 1 \quad (32)$$

$$F(x_2) = Sign(-0.55 - 0.83 + 1.1) = -1 \quad (33)$$

$$F(x_3) = Sign(0.55 - 0.83 + 1.1) = Sign(0.82) = 1 \quad (34)$$

$$F(x_4) = Sign(0.55 - 0.83 - 1.1) = Sign(-1.38) = -1 \quad (35)$$

All four assigned labels are correct.