CSCI 567 HW 3

Amir Erfan Eshratifar
USC ID: 2875431498

1 Kernels

1. Given the assumption that all the data points are distinct, the kernel matrix becomes an identity matrix:

$$K = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & \\ \vdots & \ddots & & \\ k(x_n, x_1) & & & k(x_n, x_n) \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots \\ \vdots & \ddots & \\ 0 & & 1 \end{bmatrix}$$

   Now, we have to show that identity matrix is positive semi-definite. Consider an arbitrary vector z:

$$z^T K z = z^T I z = z^T z = \|z\|_2^2 \geq 0$$

   $\|z\|_2^2$ is always non-negative, therefore, K is positive semi definite.

2. First, we need to find the optimal α (lec10):

$$\alpha^* = (\mathbf{K} + \lambda \mathbf{I})^{-1}\mathbf{y} = (\mathbf{I} + \mathbf{0I})^{-1}\mathbf{y} = \mathbf{y}$$

   By substituting α and K in $J(\alpha)$ equation:

$$\begin{aligned} J(\alpha) &= \frac{1}{2}\alpha^T K^T K \alpha - y^T K \alpha + \frac{\lambda}{2}\alpha^T K \alpha + \frac{1}{2} y^T y \\ &= \frac{1}{2} y^T I^T I y - y^T I y + \frac{1}{2} y^T y \\ &= \frac{1}{2} y^T y - y^T y + \frac{1}{2} y^T y = 0 \end{aligned}$$

3. For any x with $x \neq x_n, \forall n = 1, 2, \dots, N$, the kernel function $k(x, x_n) = 0$. Therefore:

$$f(x) = [k(x, x_1), k(x, x_2), \dots, k(x, x_n)]\alpha^* = 0$$

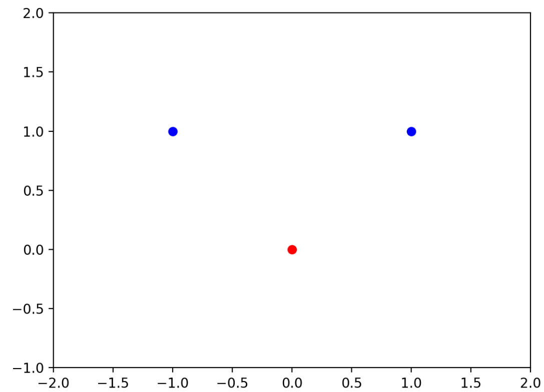
2 Support Vector Machines

1. No. Assume there is a linear classifier with w and b parameters:

$$w * (-1) + b < 0 \qquad w * (+1) + b < 0 \qquad w * (0) + b > 0$$

   , which are conflicting with each other and inferring b or w should be positive and negative simultaneously.

2. As depicted in figure below, date points can be separated using a linear classifier, for example $w^T = (0, 1), b = -0.5$, which represents a line at $x_2 = 0.5$.



3.
$$k(x,x') = \begin{pmatrix} x & x^2 \end{pmatrix} \begin{pmatrix} x' \\ x'^2 \end{pmatrix} = xx' + x^2 x'^2 \Rightarrow K = \begin{pmatrix} k(x_1,x_1) & k(x_1,x_2) & k(x_1,x_3) \\ k(x_2,x_1) & k(x_2,x_2) & k(x_2,x_3) \\ k(x_3,x_1) & k(x_3,x_2) & k(x_3,x_3) \end{pmatrix}$$
$$\Rightarrow z^T K z = \begin{pmatrix} z_1 & z_2 & z_3 \end{pmatrix} \begin{pmatrix} x_1x_1 + x_1^2x_1^2 & x_1x_2 + x_1^2x_2^2 & x_1x_3 + x_1^2x_3^2 \\ x_2x_1 + x_2^2x_1^2 & x_2x_2 + x_2^2x_2^2 & x_2x_3 + x_2^2x_3^2 \\ x_3x_1 + x_3^2x_1^2 & x_3x_2 + x_3^2x_2^2 & x_3x_3 + x_3^2x_3^2 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}$$
$$= \begin{pmatrix} z_1 & z_2 & z_3 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = 2z_1^2 + 2z_2^2 \geq 0$$

4. Primal formulation (lec11):
$$min_{w,b,\xi} \frac{1}{2}\|w\|_2^2 + C\sum_n \xi_n = min_{w,b,\xi} \frac{1}{2}(w_1^2 + w_2^2) + C(\xi_1 + \xi_2 + \xi_3)$$
such that:
$$\begin{aligned} (-1)[w_1x_{11} + w_2x_{12} + b] &\geq 1 - \xi_1 \\ (-1)[w_1x_{21} + w_2x_{22} + b] &\geq 1 - \xi_2 \\ (0)[w_1x_{31} + w_2x_{32} + b] &\geq 1 - \xi_3 \\ \xi_1 \geq 0, \xi_2 \geq 0, \xi_3 \geq 0 & \end{aligned}$$
Dual formulation (lec11):
$$O = max_\alpha[\alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}(\alpha_1^2\alpha_2^2)]$$
such that:
$$0 \leq \alpha_1 \leq C, 0 \leq \alpha_2 \leq C, 0 \leq \alpha_3 \leq C$$
$$-\alpha_1 - \alpha_2 + \alpha_3 = 0$$
Because, we are looking for a hard margin classifier we assume that C is infinite.

5. By applying the last constraint to objective function:
$$J = max_\alpha[\alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}(\alpha_1^2 + \alpha_2^2)]$$

$$\frac{\partial J}{\partial \alpha_1} = 0 \Rightarrow \alpha_1 = \alpha_2 = 1, \alpha_3 = 2$$

$$w = \sum_n y_n \alpha_n \Phi(x_n) = -1 * 1 * (-1, 1) + -1 * 1 * (1, 1) + 1 * 2 * (0, 0) = (0, -2)$$

$$b = [y_n - w^T \Phi(x_n)] = [-1 - (-2 * 1)] = 1$$

6.

$$\widehat{y} = w^T \phi(x) + b = 0 \Rightarrow -2x^2 + 1 = 0 \Rightarrow x^2 = \frac{1}{2}$$

Therefore, the decision boundary is a horizontal line at $\frac{1}{2}$. For one dimensional case, $x^2 = \frac{1}{2} \Rightarrow x = \pm\frac{sqrt(2)}{2}$

(a) 2D

(b) 1D

Figure 1: Decision boundaries and support vectors

3 Adaboost for building up a nonlinear classifier

1. At this step, there is no difference between any stump function. I chose $(s, b, d) \sim (1, 0.5, 1)$, which is a vertical line at $x_1 = 0.5$ and any point greater than 0.5 is labeled +1. Two data points are mis-classified from each label (log is natural logarithm function).

$$\epsilon_1 = \frac{1}{4} + \frac{1}{4} = 0.5$$

$$\beta_1 = \frac{1}{2}\log(\frac{1-\epsilon_1}{\epsilon_1}) = \frac{1}{2}\log(\frac{1-0.5}{0.5}) = 0$$

2.

$$w_2(n) = \begin{cases} w_1(n)exp(-\beta_1) = w_1(n) & y_n = f_1(x_n) \\ w_1(n)exp(\beta_1) = w_1(n) & otherwise \end{cases}$$

As we see, $w_2(n), \forall n \in \{1, 2, 3, 4\}$, does not get updated.

3. I chose $(s, b, d) \sim (1, 0.5, 2)$, which is a horizontal line at $x_2 = 0.5$ and any point greater than 0.5 is labeled +1. In this case x_3 is mis-classified.

$$\epsilon_1 = \frac{1}{4} = 0.25$$

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

4.

$$w_2(1) = w_1(1)exp(-\beta_1) = 0.25 * 0.58 = 0.14$$

$$w_2(2) = w_1(2)exp(-\beta_1) = 0.25 * 0.58 = 0.14$$

$$w_2(3) = w_1(3)exp(\beta_1) = 0.25 * 1.73 = 0.43$$

$$w_2(4) = w_1(4)exp(-\beta_1) = 0.25 * 0.58 = 0.14$$

Next we have to normalize the weights:

$$w_2(1) = \frac{0.14}{3 * 0.14 + 0.43} = 0.16$$

$$w_2(2) = \frac{0.14}{3 * 0.14 + 0.43} = 0.16$$

$$w_2(3) = \frac{0.43}{3 * 0.14 + 0.43} = 0.50$$

$$w_2(4) = \frac{0.14}{3 * 0.14 + 0.43} = 0.16$$

I chose $(s, b, d) \sim (-1, -0.5, 2)$, which is a horizontal line at $x_2 = -0.5$ and any point greater than 0.5 is labeled -1. In this case x_1 is mis-classified. The reason is to avoid mis-classifying a high weighted data point.

$$\epsilon_1 = 0.16$$

$$\beta_1 = \frac{1}{2}\log(\frac{1-\epsilon_1}{\epsilon_1}) = \frac{1}{2}\log(\frac{1-0.16}{0.16}) = 0.83$$

5.

$$w_3(1) = w_2(1)exp(\beta_2) = 0.16 * 2.29 = 0.37$$

$$w_3(2) = w_2(2)exp(-\beta_2) = 0.16 * 0.44 = 0.07$$

$$w_3(3) = w_2(3)exp(-\beta_2) = 0.50 * 0.44 = 0.22$$

$$w_3(4) = w_2(4)exp(-\beta_2) = 0.16 * 0.44 = 0.07$$

Next we have to normalize the weights:

$$w_3(1) = \frac{0.37}{2 * 0.07 + 0.37 + 0.22} = 0.51$$

$$w_3(2) = \frac{0.07}{2 * 0.07 + 0.37 + 0.22} = 0.10$$

$$w_3(3) = \frac{0.22}{2 * 0.07 + 0.37 + 0.22} = 0.30$$

$$w_3(4) = \frac{0.07}{2 * 0.07 + 0.37 + 0.22} = 0.10$$

I chose $(s, b, d) \sim (1, -0.5, 1)$, which is a vertical line at $x_1 = -0.5$ and any point greater than -0.5 is labeled +1. In this case x_4 mis-classified. The reason is to avoid mis-classifying a high weighted data point.

$$\epsilon_1 = 0.10$$

$$\beta_1 = \frac{1}{2}\log(\frac{1-\epsilon_1}{\epsilon_1}) = \frac{1}{2}\log(\frac{1-0.10}{0.10}) = 1.10$$

6.

$$F(x) = sign[0.55 * h_{(1,0.5,2)}(x) + 0.83 * h_{(-1,-0.5,2)}(x) + 1.10 * h_{(1,-0.5,1)}(x)]$$
$$F(x_1) = sign[0.55 * 1 + 0.83 * (-1) + 1.10 * 1] = sign[0.82] = 1$$
$$F(x_2) = sign[0.55 * (-1) + 0.83 * (-1) + 1.10 * (-1)] = sign[-2.48] = -1$$
$$F(x_3) = sign[0.55 * (-1) + 0.83 * 1 + 1.10 * 1] = sign[1.38] = 1$$
$$F(x_4) = sign[0.55 * (-1) + 0.83 * (-1) + 1.10 * (1)] = sign[-0.28] = -1$$

All the training examples are labeled correctly.