CSCI567 Homework 3

Question 1.1

Since the points are distinct:

$$x_i \neq x_j$$

The kernel matrix would become simply the identity matrix, because the only place where x=x' is on the diagonal

$$K = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

So with an arbitrary vector x, we get:

$$x^{-1}Kx = x^{-1}x = \left|\left|x\right|\right|^2 \geq 0$$

Which is always positive, thus K is positive semi-definite as is k(x,x') through Mercer's theorem

Question 1.2

Since we chose the lambda to be zero:

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

$$\frac{\partial J(\alpha)}{\partial \alpha} = \alpha^T K^T K - y^T K = 0$$

$$\alpha^T K^T K = y^T K$$

$$a^* = K^{-1}y$$

$$F = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_N) \end{bmatrix} a^* = Ka^* = KK^{-1}y = y$$

Question 1.3

$$f(x) = [k(x, x_1), k(x, x_2), \dots, k(x, x_N)]K^{-1}y$$

Since x is not in the training set, the first vector becomes a vector of all 0s.

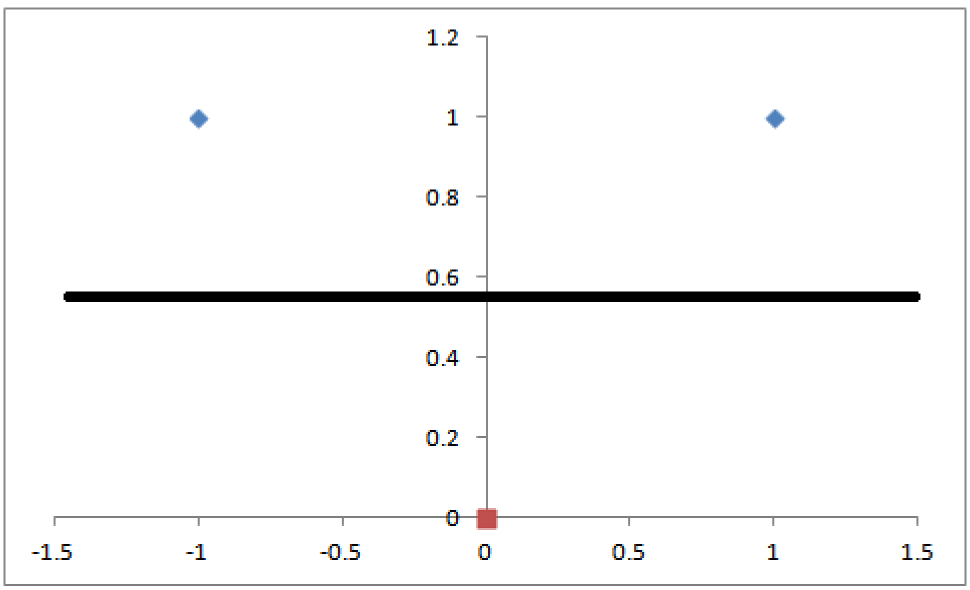
$$f(x) = [0,0,\dots,0]K^{-1}y = 0$$

Question 2.1

No, it cannot be separated. A separator can only be placed in four ways in this 1D example: on top of all 3points, on one point, between points, or on the other side of the blue points. Each of these cases, it can be shown that there is no way to classify all 3 points correctly.

Question 2.2

Yes there are now linear decision boundaries. One example is shown on the plot (the black horizontal line).



Question 2.3

$$k(x,x') = \phi(x)^T\phi(x') = \mathrm{xx'} + \mathrm{x}^2\mathrm{x}'^2$$

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

$$z^T K z = [z_1 \quad z_2 \quad z_3] \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = [2z_1 \quad 2z_2 \quad 0] \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = 2z_1^2 + 3z_2^2 \geq 0$$

Question 2.4

Primal:

$$\min_{w,b,\{\xi_n\}} C(\xi_1 + \xi_2 + \xi_3) + \frac{1}{2}\left||\mathrm{w}|\right|_2^2$$

$$s.t. \qquad 1 + w_1 - w_2 - b \le \xi_1$$

$$1 - w_1 - w_2 - b \le \xi_2$$

$$1 + b \le \xi_3$$

$$\xi_n \ge 0, \qquad \forall \mathrm{n}$$

Dual:

$$\max_{\alpha} \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}(2\alpha_1^2 + 2\alpha_2^2) = \max_{\alpha} \alpha_1 + \alpha_2 + \alpha_3 - \alpha_1^2 - \alpha_2^2$$

$$s.t. \qquad 0 \le \alpha_n \le C, \qquad \forall n$$

$$\alpha_1 + \alpha_2 - \alpha_3 = 0$$

Question 2.5

From symmetry we see:

$$\alpha_1 = \alpha_2$$

$$\min_{\alpha} \alpha_1^2 + \alpha_2^2 - \alpha_1 - \alpha_2 - \alpha_3 = \min_{\alpha} 2\alpha_1^2 - 2\alpha_1 - \alpha_3$$

Finding the vertex of the parabola:

$$\alpha_1 = \alpha_2 = \frac{2}{4} = \frac{1}{2}$$

Plugging it into the last equation of the dual formalism:

$$\alpha_3 = 1$$

$$w = \frac{1}{2}\begin{bmatrix}-1\\1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}0\\0\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix}$$

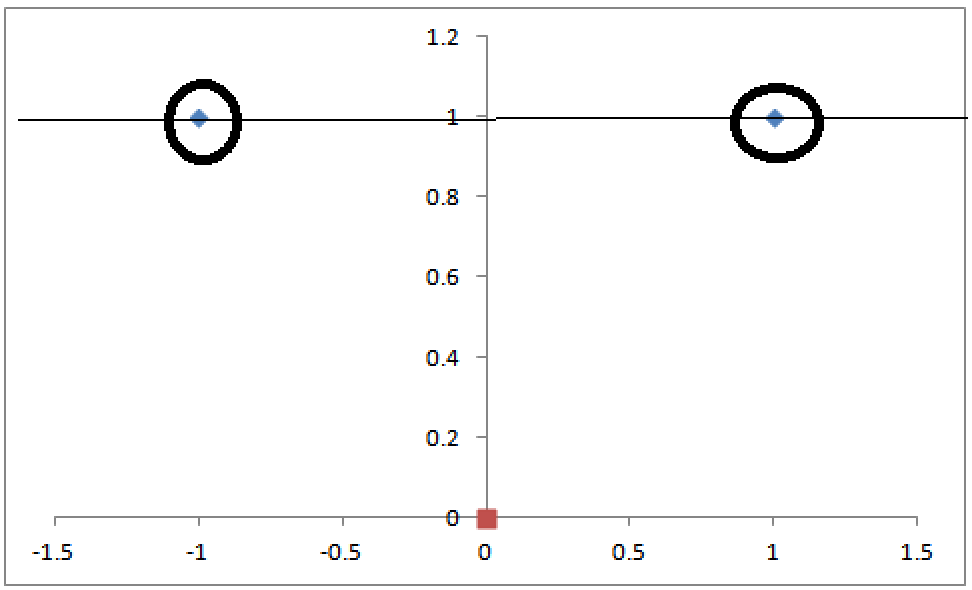
$$w^T \phi(\mathrm{x_n}) + \mathrm{b} = \mathrm{y_n}$$

$$\begin{bmatrix}0 & 1\end{bmatrix}\begin{bmatrix}0\\0\end{bmatrix} + b = 1$$

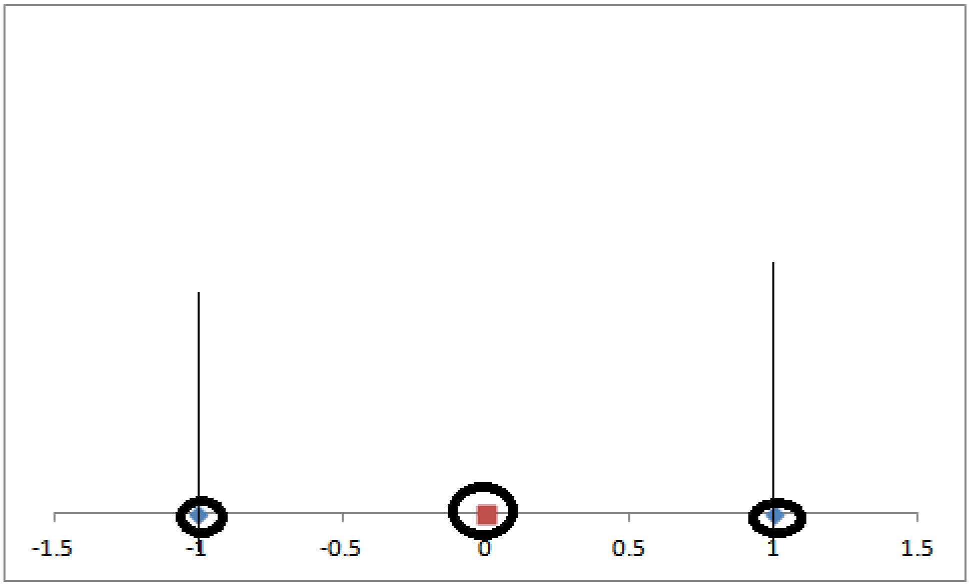
$$b = 1$$

Question 2.6

In the transformed space:



In the original space:



Question 3.1

Because of the symmetry of the problem, we only need to check two cases, if the decision boundary is between the points (b=±0.5) or if it is outside (b=±2).

$$\epsilon_{b=0.5} = \frac{1}{4} * 2 = 0.5$$

$$\epsilon_{b=2} = \frac{1}{4} * 2 = 0.5$$

Thus, all the stumps have the same error. The following values were chosen

$$f_1 = (s = +1, \qquad b = 0.5, \qquad d = 1)$$

$$\epsilon_1 = 0.5$$

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

Question 3.2

$$w_2(1) = \frac{1}{N}\exp(0) = \frac{1}{N} = \frac{1}{4} = 0.25$$

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

Question 3.3

$$f_1 = (s = +1, \qquad b = \ 0.5, \qquad d = 2)$$

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

$$\beta_1 = \frac{1}{2}\log\left(\frac{1-0.25}{0.25}\right) = 0.24$$

Question 3.4

$$w_2(1) = w_2(2) = w_2(4) = 0.25\exp(-0.24) = 0.14$$

$$w_2(3) = 0.25\exp(0.24) = 0.43$$

Normalizing:

$$w_2(1) = w_2(2) = w_2(4) = \frac{0.14}{3(0.14) + 0.43} = 0.16$$

$$w_2(3) = 0.5$$

$$f_2 = (s = -1, \qquad b = \ -0.5, \qquad d = 2)$$

$$\epsilon_2 = 0.16$$

$$\beta_2 = \frac{1}{2}\log\left(\frac{1-0.16}{0.16}\right) = 0.36$$

Question 3.5

$$w_3(2) = w_3(4) = 0.16\exp(-0.36) = 0.07$$

$$w_3(1) = 0.16\exp(0.36) = 0.37$$

$$w_3(3) = 0.5\exp(-0.36) = 0.22$$

Normalizing:

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

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

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

$$f_3 = (s = -1, \qquad b = \; 0.5, \qquad d = 1)$$

$$\epsilon_3 = 0.10$$

$$\beta_3 = \frac{1}{2}\log\left(\frac{1 - 0.1}{0.1}\right) = 0.48$$

Question 3.6

$$F(x) = sign\left(0.24 \times h_{+1,0.5,2} + 0.36 \times h_{-1,-0.5,2} + 0.48 \times h_{-1,0.5,1}\right)$$

$$F(x_1) = sign(0.24 - 0.36 + 0.48) = sign(0.36) = +1$$

$$F(x_2) = sign(-0.24 - 0.36 + 0.48) = sign(-0.12) = -1$$

$$F(x_3) = sign(-0.24 + 0.36 + 0.48) = sign(0.6) = +1$$

$$F(x_4) = sign(-0.24 - 0.36 - 0.48) = sign(-1.08) = -1$$

It predicted all 4 correctly!