CSCI 567 - Home Work - 3

Sai Swarna Latha Vempaty - 5824899937

1.1)

computing kernel matrix

$$K = \begin{pmatrix} k(x_1,x_1) & k(x_1,x_2) & \dots & k(x_1,x_N) \\ k(x_2,x_1) & k(x_2,x_2) & \dots & k(x_2,x_N) \\ \dots & \dots & \dots & \dots \\ k(x_N,x_1) & k(x_N,x_2) & \dots & k(x_N,x_N) \end{pmatrix}$$

but given that $k(x_i, x_j) = 1$ if i=j ,= 0if i $\neq j$ substituting in the above, we get

$$K= \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix} = I_{NxN}$$

to show that K is positive semi definite (PSD),consider any non zero vector,a $\in R^N$,

$$a^T.K.a = (a_1 a_2 .... a_N).I. \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_N \end{pmatrix} = a_1^2 + a_2^2 + ...a_N^2 >= 0$$

Therefore, according to mercer's theorem, k is valid kernel

note:K is also symmetric

1.2)

$$\partial J(\alpha)/\partial\alpha = K^2\alpha^* - Ky + \lambda K\alpha^* = 0 => \alpha^* = K^{-1}.y$$

consider a training example x_1, f(x_1)=[$k(x_1, x_1)$ $k(x_1, x_2)$...$k(x_1, x_n)$].K_{NxN}. $\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{pmatrix}$

by substituting k(.,.) and K_{NxN} from Q(1.1), we get

f(x_1)=[1 0 0 ...].I_{NxN}. $\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{pmatrix} = y_1$

Similarly for all training examples x_n ,n $\in$ {1,2,3,...}, we can prove that f(x_n)=y_n, therefore, λ=0 leads to training objective of 0

1.3)

$$f(x) = [k(x,x_1)\ k(x,x_2) \ldots\ k(x,x_N)].K_{NxN}.\begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_N \end{pmatrix}$$

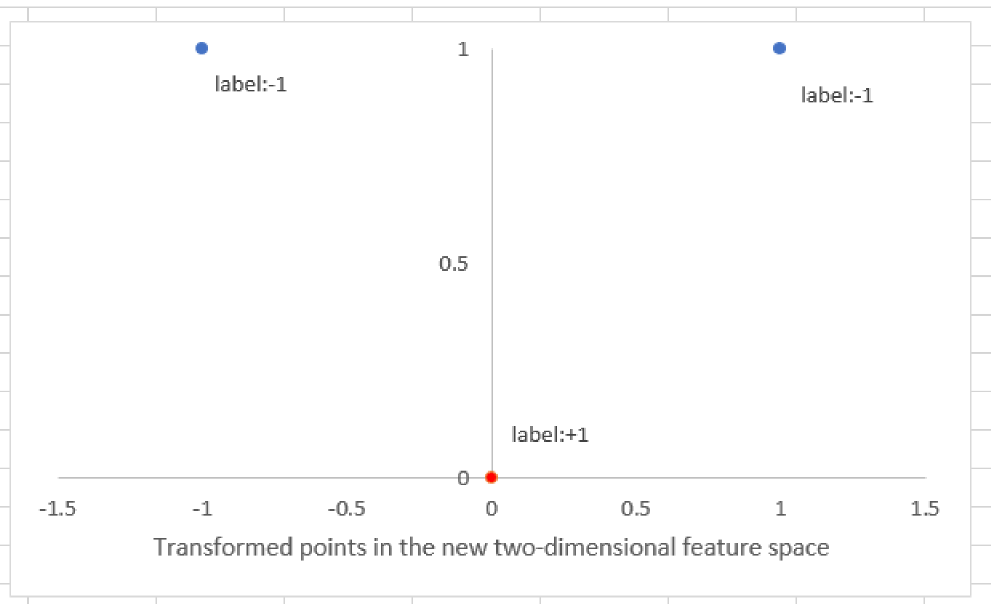
for x such that x $\neq x_n$, n $\in$ 1,2,...N; $k(x, x_n) = 0$ and $K_{NxN} = I_{NxN}$ - from Q(1.1)

$$=> f(x) = [0\ 0\ 0\ ..0].I_{NxN}.\begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_N \end{pmatrix} = 0$$

Implies for all x not in training set, first vector is zero, so predicted value is always 0.

2.1) As, all the three points are on a same line, so we cannot separate blue points that are on either side of red point with a single linear hyper plane. We will need a non-linear to classify all the points correctly.

2.2)



Yes, there is a linear boundary that can separate the points in the new dimensional feature space. Visually, there are many lines parallel to x-axis that can separate the classes, but the maximum margin classifier would be $x^2 = 0.5$ (visually).

2.3)

k(x,x')=$\phi^T(x).\phi(x')$ $\quad = xx' + (xx')^2$

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

to show that K is positive semi definite (PSD),consider any non zero vector,z $\in R^3$,

$$z^T.K.z = (z_1 z_2 z_3).K.\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = 2(z_1^2 + z_2^2) >= 0$$

Therefore K is PSD.

2.4)

<u>Primal Formulation:</u>

$min_w \ \frac{1}{2}.||\omega||_2^2$

s.t. $y_n \ [w^T\phi(x_n) + b]$ >= 1 $\forall$ n.

upon substituting $\phi(x_n) = [x \ x^2]^T$, $\omega = [\omega_1 \quad \omega_2]^T, (x_1, y_1) = (-1, -1)$
$(x_2, y_2) = (1, -1)$ $(x_3, y_3) = (0, 1)$ the constraint statement becomes:

$\omega_1 - \omega_2 - b >= 1$

$-\omega_1 - \omega_2 - b >= 1$

$b >= 1$

<u>Dual Formulation:</u>

$min_\alpha \ \alpha_1^2 + \alpha_2^2 - \alpha_1 - \alpha_2 - \alpha_3$

s.t 0<=α_1 , 0<=α_2 , 0<=α_3

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

2.5)

$min_\alpha \ \alpha_1^2 + \alpha_2^2 - \alpha_1 - \alpha_2 - \alpha_3$

using symmetric property, $\underline{\alpha_2 = \alpha_1}$ and $\alpha_3 = \alpha_1 + \alpha_2 =>$ $\underline{\alpha_3 = 2\alpha_1}$

=>$min_\alpha \ 2\alpha_1^2 - 4\alpha_1$

Taking derivative and equating it to zero, $\underline{\alpha_1 = 1}$

$\alpha_2 = 1, \alpha_3 = 2$

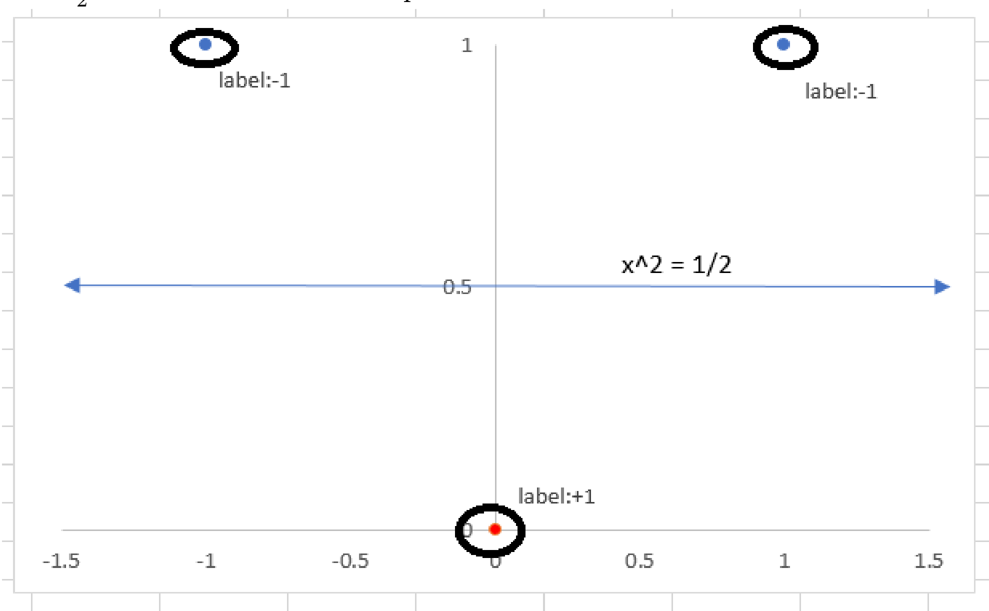
$$\alpha = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

$\omega = \sum_n .\alpha_n.y_n.\phi(x_n)$ = (1)(-1)(-1 1)^T+ (1) (-1) (1 1)^T + (2) (1) (0 0)^T = $\underline{\text{(0 -2)}^T}$

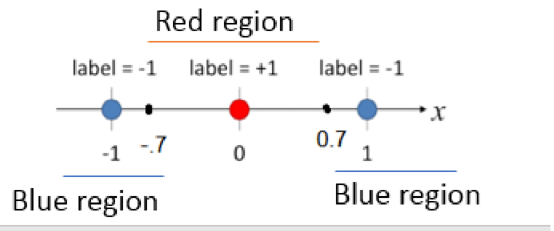
$b = y_n - \sum_m y_m \alpha_n k(m, n)$ = $y_2 - y_2.\alpha_2.k(2, 2)$ = 1

2.6)

$x^2 = \frac{1}{2}$. Plot in 2-dimensional space:



for 1-D, x=+$\frac{1}{\sqrt{2}}$ x=-$\frac{1}{\sqrt{2}}$



3.1)

All the 16 linear classifiers have 50% accuracy, hence selecting one in random:

$f_1 = h_{(1,-0.5,2)}$

$\epsilon_1 = 0.25 * 1 + 0.25 * 1 + 0 + 0 = 0.50$

$\beta_1 = 0.5 * log\frac{0.5}{0.5} = 0$

3.2)

$\omega_2(1) = 0.25 * e^{-0} = 0.25$

$\omega_2(2) = 0.25 * e^{0} = 0.25$

$\omega_2(3) = 0.25 * e^{0} = 0.25$

$\omega_2(4) = 0.25 * e^{-0} = 0.25$

objective function remains constant even after this step, as weights did not change.

3.3)

four classifiers classified only one point in wrongly, selecting one of them randomly.

$f_1 = h_{(1,-0.5,1)}$

$\epsilon_1 = 0.25 * 1 = 0.25$ (f_1 classified only one point namely, x_4 wrong.)

$\beta_1 = 0.5 * log\frac{0.75}{0.25} = 0.55$

3.4)

At t=2,

$w_2(1) = 0.25 * e^{-0.55} = 0.14, w_2(2) = 0.25 * e^{-0.55} = 0.14,$

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

After normalization:

$w_2(1) = 0.17, w_2(2) = 0.17,$

$w_2(3) = 0.17, w_2(4) = 0.49$

out of two classifiers that had minimum error rate selected one randomly.

$f_2 = h_{(1,0.5,2)}$

$\epsilon_2 = 0.17 * 1 = 0.17$ (f_1 classified only one point namely, x_3 wrong.)

$\beta_2 = 0.5 * log\frac{0.83}{0.17} = 0.79$

3.5)

At t=3,

$w_3(1) = 0.17 * e^{-0.79} = 0.08, w_3(2) = 0.17 * e^{-0.79} = 0.08,$

$w_3(3) = 0.17 * e^{0.79} = 0.37, w_3(4) = 0.49 * e^{-0.79} = 0.22$

After normalization:

$w_3(1) = 0.10, w_3(2) = 0.10,$

$w_3(3) = 0.49, w_3(4) = 0.31$

$f_3 = h_{(-1,0.5,1)}$

$\epsilon_3 = 0.10 * 1 = 0.10$ (f_1 classified only one point namely, x_2 wrong.)

$\beta_3 = 0.5 * log\frac{0.90}{0.10} = 1.10$

3.6)

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

$F(1) = sign[0.55 * 1 + 0.79 * 1 + 1.10 * 1] = 1$

$F(2) = sign[0.55 * -1 + 0.79 * -1 + 1.10 * 1] = -1$

$F(3) = sign[0.55 * 1 + 0.79 * -1 + 1.10 * 1] = 1$

$F(4) = sign[0.55 * 1 + 0.79 * -1 + 1.10 * -1] = -1$

All the 4 points in transformed system are properly labelled.