Q1.1)

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

$I^T = I$ Thus K is Symmetric

$$< x, Kx > = < x, K^T Kx > = < Kx, Kx > = \sum x^2 \geq 0$$

Hence, K is PSD (Positive Semi-Definite)

By Mercer Theorem K is a valid kernel

Q1.2)

Substituting λ = 0 and K=1 in training objective we get,

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

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

$$\alpha^* = y^T$$

Substituting back in in training objective,

$J(\boldsymbol{\alpha}) = \frac{1}{2}\mathbf{y}^T\mathbf{y} - \boldsymbol{y}^T\mathbf{y} - \frac{1}{2}\boldsymbol{y}^T\boldsymbol{y}$ = 0

Q1.3)

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

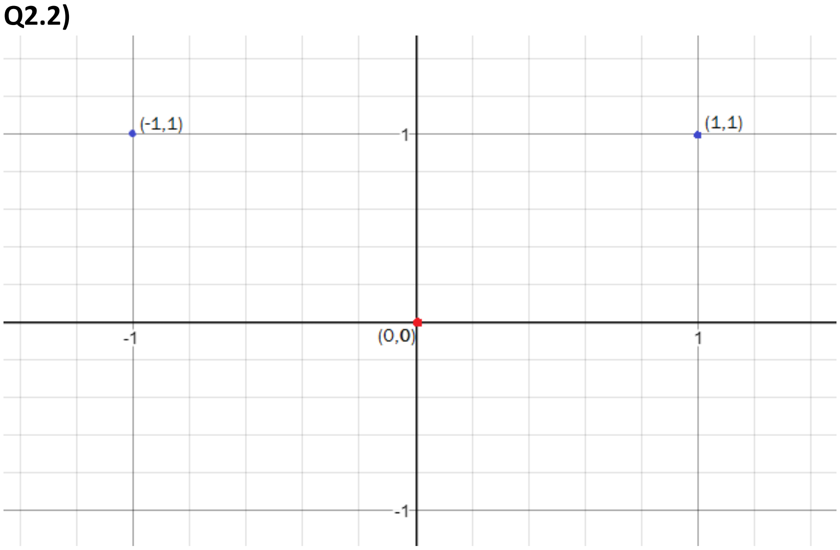
$\forall$x, x $\neq x_n$. We know, $k(x, x_n) = 0$ for the given kernel.

Therefore, $f(x) = 0$

Q2.1)

Not linearly separable because it is evident visually that there cannot exist a line that can separate the points with 100% accuracy.

Q2.2)



Yes, the data is now linearly separable in the new feature space

Q2.3)
$K(x_m, x_n) = x_m x_n + {x_m}^2 {x_n}^2$
$\boldsymbol{K} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
Let $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \alpha_3]$
$\boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha}$ = $[\alpha_1, \alpha_2, \alpha_3] \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$
$\boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} = 2{\alpha_1}^2 + 2{\alpha_2}^2 \geq 0$
Thus, $\mathbf{K}$ is PSD (Positive semi-definite)

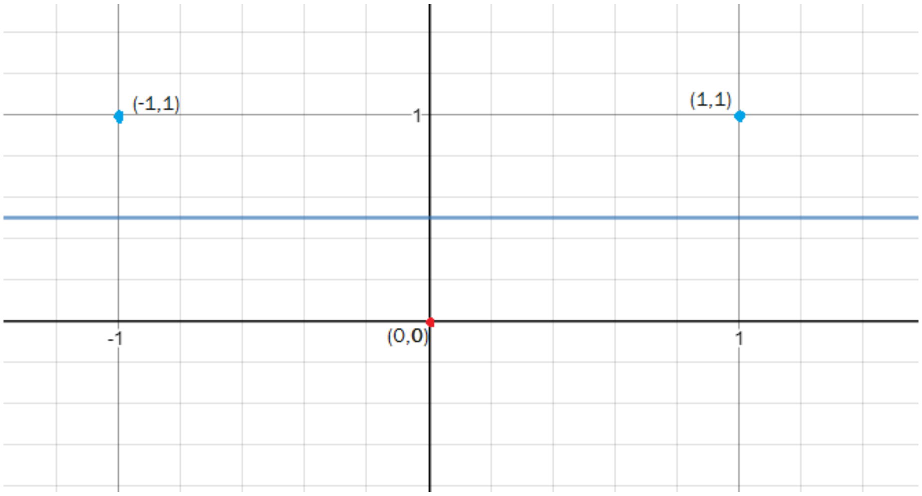
Q2.4)
$\varphi(\mathrm{x}) = [x, x^2]^T$ Let $w = [w_1, w_2]^T$
$min_{w,b,\xi} \sum_n \xi_{\boldsymbol{n}} + \frac{1}{2}||w||^2$ such that, $\forall \mathrm{n}\ (1 - y_n[w_1 x + w_2 x^2 + b]) \leq \xi_{\boldsymbol{n}}$ and $\xi_{\boldsymbol{n}} \geq 0$

Dual formulation can be obtained by substituting **K** from **Q2.3)**
$max_\alpha (\sum_n \alpha_n - \frac{1}{2} \sum_{m,n} (y_m y_n \alpha_m \alpha_n x_m x_n + y_m y_n \alpha_m \alpha_n {x_m}^2 {x_n}^2))$
such that, $\forall \mathrm{n}\ 0 \leq \alpha_n \leq c$ and $\sum_n \alpha_n \mathrm{y}_n = 0$

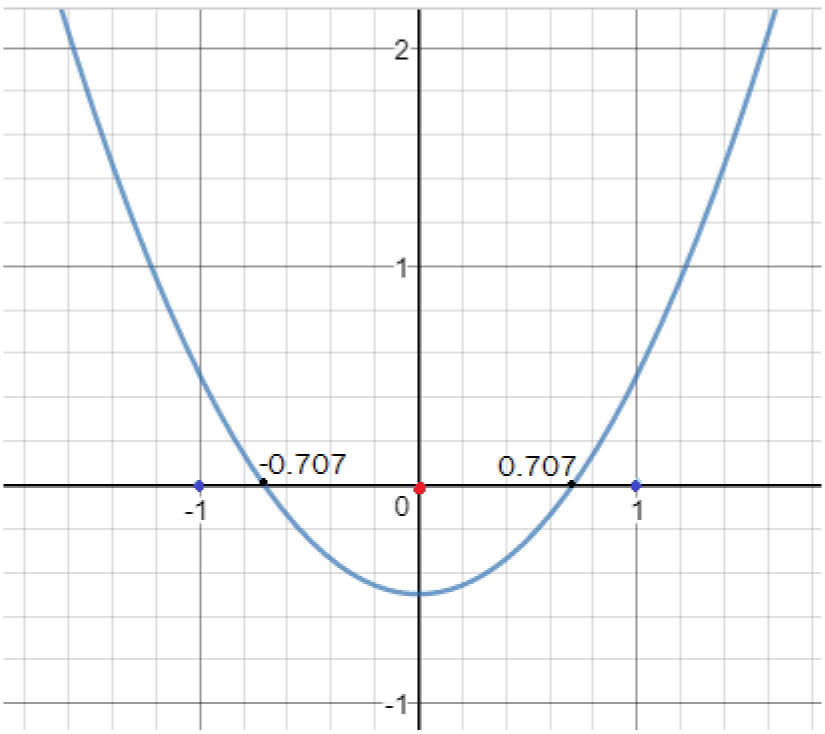
Q2.5)
The dual formulation can be re-written as,
$min_\alpha (\frac{1}{2} (\sum_{m,n} (y_m y_n \alpha_m \alpha_n x_m x_n + y_m y_n \alpha_m \alpha_n {x_m}^2 {x_n}^2) - \sum_n \alpha_n))$
such that, $\forall \mathrm{n}\ 0 \leq \alpha_n \leq c$ and $\sum_n \alpha_n \mathrm{y}_n = 0$

Expanding the inverse with K and $1 \leq m, n \leq 3$ we get,
${y_1}^2 {\alpha_1}^2 (2) + {y_2}^2 {\alpha_2}^2 (2) - (\alpha_1 + \alpha_2 + \alpha_3) = 0$
$y_1 \alpha_1 + y_2 \alpha_2 + y_3 \alpha_3 = 0$
we know, $y_1 = -1, y_2 = -1, y_3 = 1$
$\alpha_1 = \alpha_2$ since the equations are symmetrical
we then get, $\alpha_3 = 2\alpha_1$
Solution: $< \alpha_1, \alpha_1, \alpha_1 > = \{0,0,0\}$ and $\{1,1,2\}$
$W = [0, -2]$ and $b = 1$

Q2.6)



Plot in 2-Dimensional Feature Space $\hat{y} = x^2 - 1/2$



Plot in 1-Dimensional Feature Space

Q3.1)

$(s, b, d) = (1{,}0.5{,}1);$
$\epsilon_1 = 0.5$
$\beta_1 = 0$

Q3.2)

$w_2(1) = 0.25, w_2(2) = 0.25, w_2(3) = 0.25, w_2(4) = 0.25$
Hence, the algorithm is stuck at $t = 1$

Q3.3)
$(s, b, d) = (1, -0.5, 1);$
$\in_1 = 0.25$
$\beta_1 = 0.55$

Q3.4)
$w_2(1) = 0.17, w_2(2) = 0.17, w_2(3) = 0.17, w_2(4) = 0.51$
$(s, b, d) = (1, 0.5, 2);$
$\in_2 = 0.17$
$\beta_2 = 0.79$

Q3.5)
$w_3(1) = 0.1, w_3(2) = 0.1, w_3(3) = 0.49, w_3(4) = 0.3$
$(s, b, d) = (-1, 0.5, 1);$
$\in_3 = 0.1$
$\beta_3 = 1.1$

Q3.6)
$F(x) = sign[0.55h_{(1,-0.5,1)} + 0.79h_{(1,0.5,2)} + 1.1h_{(-1,0.5,1)}$
$F(x_1) = sign(0.55 + 0.79 + 1.1) = sign(2.44) = +1$
$F(x_2) = sign(-0.55 - 0.79 + 1.1) = sign(-0.24) = -1$
$F(x_3) = sign(0.55 - 0.79 + 1.1) = sign(0.86) = +1$
$F(x_4) = sign(0.55 - 0.79 - 1.1) = sign(-1.34) = -1$
All the examples are correctly classified.