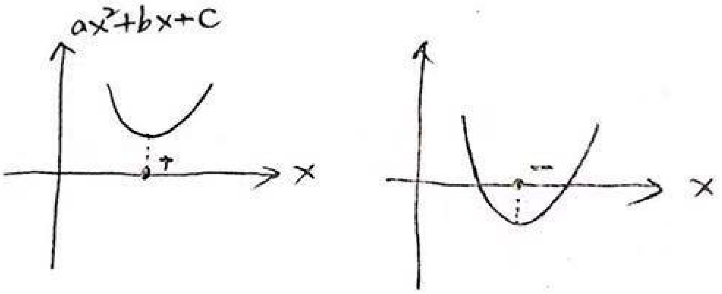
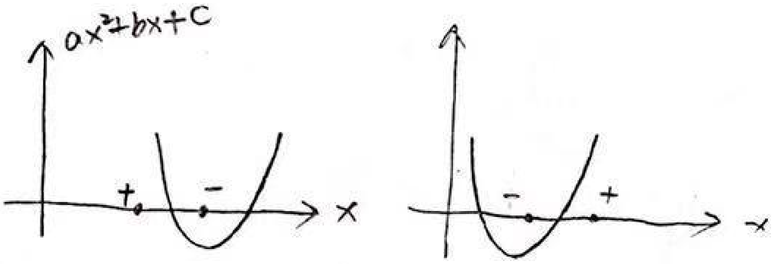
Problem 1:

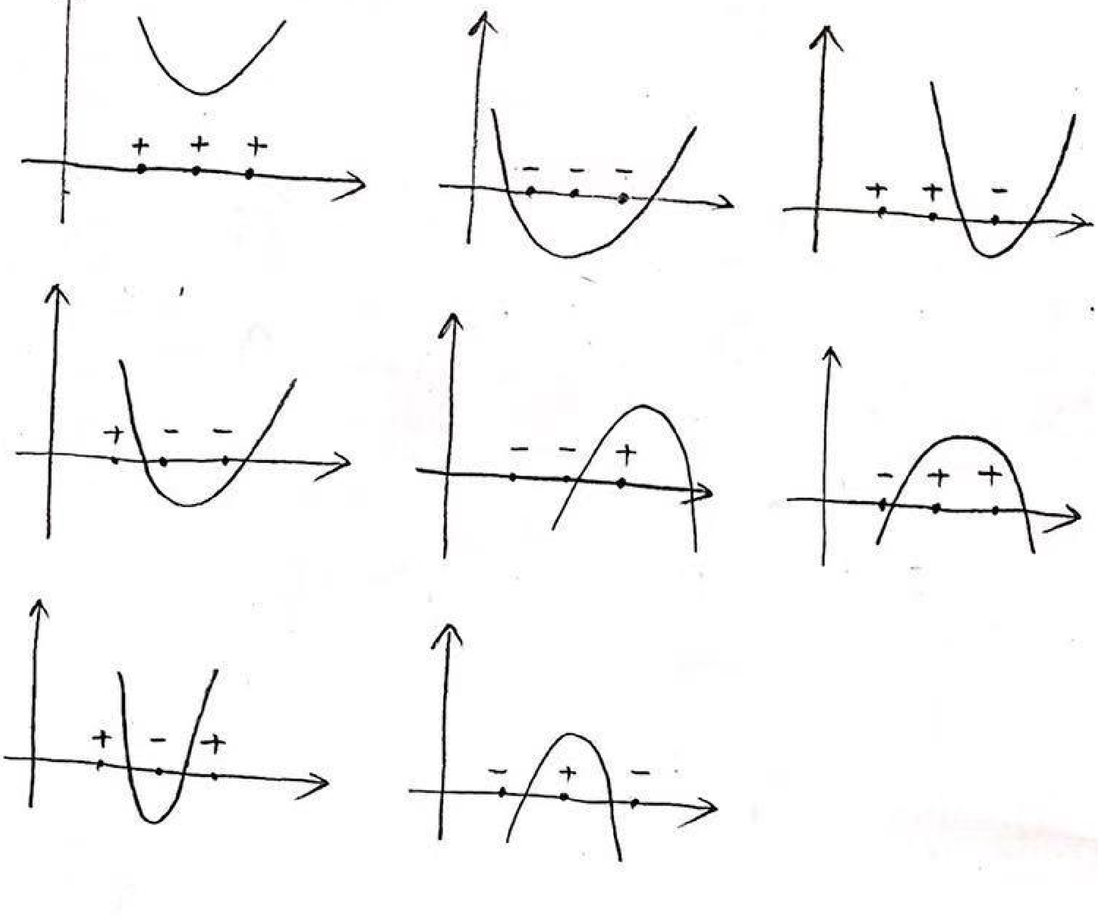
① $VC \geq 1$



② $VC \geq 2$



③ $VC \geq 3$



④ $VC \leq 4$.

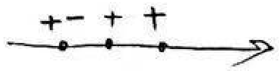
There are two classes of point distribution to represent all point distributions, and we can prove that neither of them can be shattered by H based on the result of $VC \geq 3$.

(1) 2 points are overlapped.

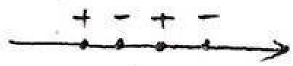
Then we can assign "+" to one point and "−" to the other.

(2) All four points have different x value.

Then we can assign their labels as "+" "−" "+" "−" from the smallest x to the biggest x



(1)



(2)

Problem 2

$$x = [x_1 \; x_2]^T, \quad z = [z_1 \; z_2]^T$$

$$K_\beta(x,z) = (1+\beta x \cdot z)^3 = \left[1+\beta(x_1 z_1 + x_2 z_2)\right]^3$$

$$= 1 + 3\beta(x_1 z_1 + x_2 z_2) + 3\beta^2(x_1^2 z_1^2 + 2x_1 x_2 z_1 z_2 + x_2^2 z_2^2)$$

$$+ \beta^3(x_1^3 z_1^3 + 3x_1^2 x_2 z_1^2 z_2 + 3x_1 x_2^2 z_1 z_2^2 + x_2^3 z_2^3)$$

$$\phi_\beta(x) = [1, \sqrt{3\beta}x_1, \sqrt{3\beta}x_2, \sqrt{3}\beta x_1^2, \sqrt{6}\beta x_1 x_2, \sqrt{3}\beta x_2^2, \beta^{\frac{3}{2}} x_1^3,$$
$$\sqrt{3}\beta^{\frac{3}{2}} x_1^2 x_2, \sqrt{3}\beta^{\frac{3}{2}} x_1 x_2^2 + \beta^{\frac{3}{2}} x^3]^T$$

β is a parameter to show the weight of each term.

if $0<\beta<1$, $\sqrt{\beta} > \beta > \beta^{\frac{3}{2}}$, the lower dimension weighs more

if $\beta = 1$, $\sqrt{\beta} = \beta = \beta^{\frac{3}{2}} = 1$, $K_\beta(x,z) = K(x,z)$

if $\beta > 1$, $\sqrt{\beta} < \beta < \beta^{\frac{3}{2}}$, the higher dimension weighs more.

Problem 3

(a). Our goal is to find $\min_{w} \frac{1}{2}\|w\|^2$

s.t. $y_n w^T x_n \geq 1$

Now we have $x_1 = (1,1)^T$ "+" $\Rightarrow y_1 w^T x_1 = 1 \cdot (w_1, w_2)\begin{pmatrix}1\\1\end{pmatrix} \geq 1$

$x_2 = (1,0)^T$ "−" $\Rightarrow y_2 w^T x_2 = (-1)(w_1, w_2)\begin{pmatrix}1\\0\end{pmatrix} \geq 1$

$$\Rightarrow \begin{cases} w_1 + w_2 \geq 1 \\ -w_1 \geq 1 \end{cases} \Rightarrow \begin{array}{l} \min \frac{1}{2}\|w\|^2 \\ \text{s.t.} \begin{cases} w_1 + w_2 \geq 1 \\ w_1 \leq -1 \end{cases} \end{array} \Rightarrow \begin{cases} w_1 = -1 \\ w_2 = 2 \end{cases} \Rightarrow w = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

(b). $x_1 = (1,1,1)^T$ $y_1 = +1 \Rightarrow y_1 w^T x_1 = (+1) \cdot (w_1\ w_2\ b^*)\begin{pmatrix}1\\1\\1\end{pmatrix} \geq 1$

$x_2 = (1,0,1)^T$ $y_2 = -1 \Rightarrow y_2 w^T x_2 = (-1)(w_1\ w_2\ b^*)\begin{pmatrix}1\\0\\1\end{pmatrix} \geq 1$

$$\Rightarrow \begin{cases} w_1 + w_2 + b \geq 1 \\ -(w_1 + b) \geq 1 \end{cases} \Rightarrow \begin{array}{l} \min_{w} \frac{1}{2}\|w\|^2 \\ \text{s.t.} \begin{cases} w_1 + w_2 + b \geq 1 \\ w_1 + b \leq -1 \end{cases} \end{array} \Rightarrow \begin{cases} w_1 = 0 \\ w_2 = 2 \\ b = -1 \end{cases}$$

$$w^* = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \quad b^* = -1$$

4.1
(d)
The dimensionality of the feature matrix is 1811.

4.2
(b)
Because we assume that training data and testing data are independent identically distributed and based on this assumption we implement our algorithms. If the class portions have a big difference between training and testing data, then they are no longer i.i.d..
(d)

C	accuracy	F1-score	AUROC
10^{-3}	0.7089	0.8297	0.8105
10^{-2}	0.7107	0.8306	0.8111
10^{-1}	0.8060	0.8755	0.8575
10^{0}	0.8146	0.8749	0.8712
10^{1}	0.8182	0.8766	0.8696
10^{2}	0.8182	0.8766	0.8696
best C	10&100	10&100	1

4.3
(c)
Accuracy: c=10 performance: 0.7429
F1-Score: c=10 performance:0.4375
AUROC: c=1 performance:0.7405