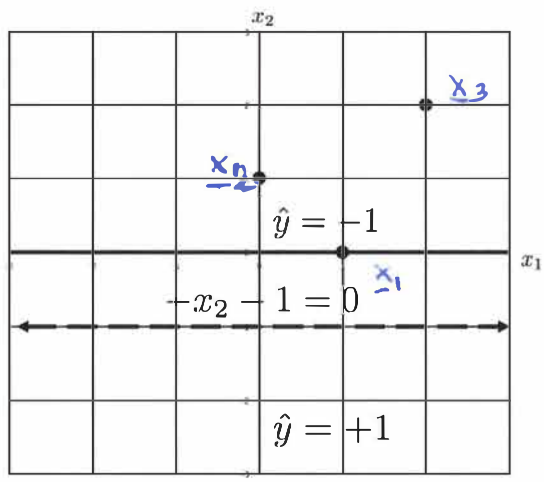
1. (10 MARKS) Consider a binary classification problem where the data points are two dimensional, i.e., $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ and the labels $y \in \{-1, 1\}$. Throughout this problem consider the following three points:

$$\mathbf{x}_1 = (1,0)^T, \quad \mathbf{x}_2 = (0,1)^T, \quad \mathbf{x}_3 = (2,2)^T.$$

4 marks (a) Suppose that the hypothesis set $\mathcal{H}$ consists of all linear classifiers whose decision boundary is a **horizontal line** in the (x_1, x_2) plane. As one example, the classifier $h(\mathbf{x}) = \text{sign}(-x_2 - 1)$ belongs to the set $\mathcal{H}$. The decision boundary of $h(\mathbf{x})$ is the dashed horizontal line shown in the figure below. Note that in this classifier all points below this line are classified as $\hat{y} = +1$ while all points above the horizontal line are classified as $\hat{y} = -1$ by the hypothesis $h(\mathbf{x})$



List all dichotomies in $\mathcal{H}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ that can be achieved. (Recall that a dichotomy in this problem will be a vector of length three whose elements are either +1 or −1, and is achieved by applying some hypothesis in $\mathcal{H}$ to the points $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ in that order.)

What is the VC dimension of $\mathcal{H}$? (No justification is needed for this)

Achievable Dichotomies:

x_1	x_2	x_3
+1	+1	+1
−1	+1	+1
−1	−1	+1
−1	−1	−1
1	−1	−1
1	1	−1

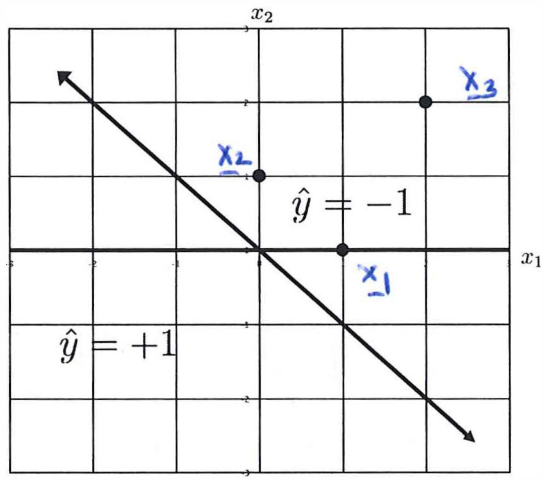
0.5 pt for each correct dichotomy
−0.5 for each incorrect

$$\underbrace{d_{vc}(H) = 1}_{1\ \text{pt}}$$

total/4

4 marks

(b) Suppose that the hypothesis set $\mathcal{G}$ consists of all linear classifiers **passing through the origin**. As one example the classifier $g(\mathbf{x}) = \text{sign}(-x_1 - x_2)$ belongs to the set $\mathcal{G}$. Its decision boundary is shown by the solid line passing through the origin in the figure below. Note that all points below the decision boundary are classified as $\hat{y} = +1$ and all points above this line are classified as $\hat{y} = -1$.



List all dichotomies in $\mathcal{G}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ that can be achieved.

State without justification the VC dimension of $\mathcal{G}$.

0.5 for each correct -0.5 for an incorrect

x_1	x_2	x_3
+1	+1	+1
-1	+1	+1
1	-1	+1
-1	-1	-1
1	-1	-1
-1	1	-1

$\underbrace{d_{vc}(\mathcal{H}) = 2}_{1 \text{ pt}}$

2 marks

(c) Suppose $\mathcal{M} = \mathcal{H} \cup \mathcal{G}$ is the union of the hypothesis classes in parts (a) and (b). What is the number of dichotomies in $\mathcal{M}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$? Provide a **brief** justification for your answer.

$|\mathcal{M}(x_1, x_2, x_3)| = 8$ — 1

the two dichotomy vector missing in $\mathcal{G}$, are found in $\mathcal{H}$. } 1

total/6

10 marks **3**. Suppose we are given a sequence of real numbers: $x_1, x_2, x_3, x_4, \ldots$ where $x_i \in \mathbb{R}$. We observe the following values: $x_1 = -1$, $x_2 = 0$, $x_3 = +1$, $x_4 = +1$.

We wish to select a prediction function of the form $\hat{x}_i = w_0 + w_1 \cdot x_{i-1}$, for $i \geq 2$ that makes a prediction of x_i from the value of x_{i-1}. Our task is to minimize the following in sample training error:

$$E_{\text{in}}(\mathbf{w}) = \frac{1}{3}\sum_{i=2}^{4}(\hat{x}_i - x_i)^2.$$

3 marks (a) Rewrite the above problem specifications to get the problem into the standard form of a least squares problem: $\mathbf{w}^\star = \arg\min_{\mathbf{w}\in\mathbb{R}^2} \frac{1}{3}\|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2$. Specifically specify the data matrix $\mathbf{X}$ and the target vector $\mathbf{y}$, where $\mathbf{w} = (w_0, w_1)^T$.

$$\underline{y} = \begin{bmatrix} x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \longrightarrow 1\ \text{pt}$$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \longrightarrow 2\ \text{pt}$$

3 marks (b) Find the least squares solution $\mathbf{w}^\star = (w_0^\star, w_1^\star)$ in part (a).

$$\underline{w}^* = (X^T X)^{-1} X^T \underline{y}$$

$$(X^T X)^{-1} = \begin{bmatrix} 1/3 & 0 \\ 0 & 1/2 \end{bmatrix} \text{ — } 1\ \text{pt}$$

$$X^T Y = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \text{ — } 1\ \text{pt}$$

$$\underline{w}^* = \begin{bmatrix} 2/3 \\ 1/2 \end{bmatrix} \text{ — } 1\ \text{pt}$$

(alternative solution accepted)

total/6

4 marks

(c) Redo parts (a) and (b) if the prediction function is of the form $\hat{x}_i = w_1 \cdot x_{i-1}$ i.e., we set $w_0 = 0$.

$$\underline{y} = \underbrace{\begin{bmatrix}0\\1\\1\end{bmatrix}}_{1\text{ pt}} \qquad X = \underbrace{\begin{bmatrix}-1\\0\\1\end{bmatrix}}_{1\text{ pt}}$$

$$\left.\begin{aligned}(X^T X)^{-1} &= \frac{1}{2}\\ X^T \underline{Y} &= \begin{bmatrix}-1 & 0 & 1\end{bmatrix}\begin{bmatrix}0\\1\\1\end{bmatrix} = 1\end{aligned}\right\} 1\text{ pt}$$

$$\underline{w}_1^* = \frac{1}{2}. \quad \text{— } 1\text{ pt}$$

(alternative solu^n to $\underline{w}_1^*$ is accepted)

total/4