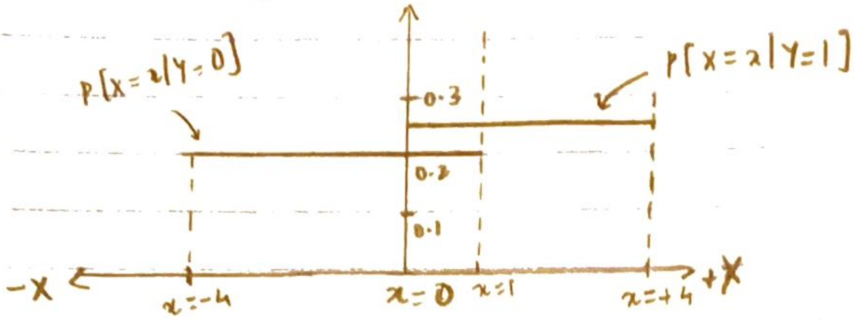
Problem 1

a) $P[X=x \mid Y=0] = \frac{1}{5}$ for $-4 \le x \le 1$; 0 elsewhere

$P[X=x \mid Y=1] = \frac{1}{4}$ for $0 \le x \le 4$; 0 elsewhere



b) $P[Y=y \mid X=x] = ?$ in terms of $P[X=x \mid Y=y]$, $P[Y=y]$, $P[X=x]$

Using Bayes rule, we can rewrite

$$P[Y=y \mid X=x] = \frac{P[Y=y, X=x]}{P[X=x]} = \boxed{\frac{P[X=x \mid Y=y]\, P[Y=y]}{P[X=x]}}$$

c) $h^*(x) = \underset{y \in \{0,1\}}{\operatorname{argmax}}\ \mathbb{P}[Y=y \mid X=x]$

$P[Y=0] = p$

$P[Y=1] = 1-p$

$$= \operatorname{argmax} \frac{P[X=x \mid Y=y]\, P[Y=y]}{P[X=x]}$$

3 main segments in piecewise $h^*(x)$

= ① $\underline{0}$ if $X \in [-4, 0]$ $\because \frac{1}{5} \cdot p \ge 0$

② $\underline{0}$ if $X \in [0, 1]$ & $p \ge 5/9$ (depends on p)

$\underline{+1}$ else if $p < 5/9$ & $x \in [0,1]$

③ $\underline{+1}$ if $X \in [1, 4]$ $\because 0 \le \frac{1}{4} - \frac{p}{4}$

$(\because)$

$$\frac{p}{5} \ge \frac{1}{4} - \frac{p}{4} \Rightarrow \frac{9p}{20} \ge \frac{1}{4} \Rightarrow 9p \ge 5 \Rightarrow \boxed{p \ge \frac{5}{9}}$$

d) $P_{X,Y}[h^*(X) \neq Y]$ depends on p. because if we choose 0 or 1, the ~~other option being true~~ probability of other is the error rate of the Bayes optimal classifier

Thus: error rate:

$$p \cdot \frac{1}{5} \cdot P[X \in [0,1]] \quad \text{if} \quad \frac{p}{5} < \frac{(1-p)}{4}$$

$$(1-p) \cdot \frac{1}{4} \cdot P[X \in [0,1]] \quad \text{if} \quad \frac{p}{5} \geq \frac{(1-p)}{4}$$

Either way: $P[X \in [0,1]] = P[X = x \mid Y = 0] \cdot P[Y=0] + P[X = x \mid Y = 1] \cdot P[Y=1]$

$$= \frac{1}{5} \cdot 1 \cdot p + \frac{1}{4} \cdot 1 \cdot (1-p)$$

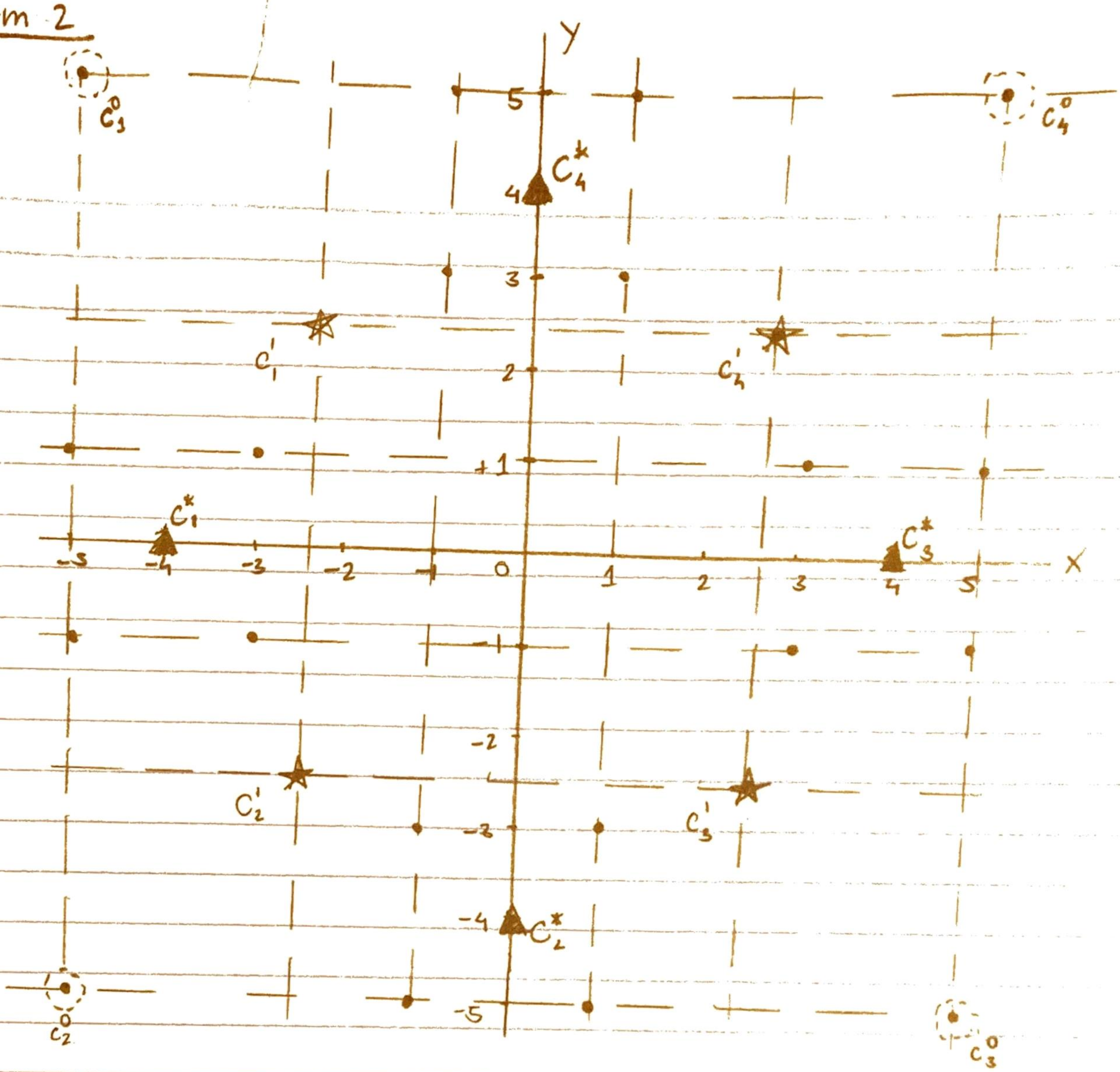
$$= \frac{1}{4} + \frac{4p - 5p}{20} = \frac{1}{4} - \frac{p}{20}$$

$$\Rightarrow P_{X,Y}[h^*(X) \neq Y] = \frac{p}{5}\left(\frac{1}{4} - \frac{p}{20}\right) \text{ if } p < \frac{5}{9}$$

$$\text{OR} \quad \frac{(1-p)}{4}\left(\frac{1}{4} - \frac{p}{20}\right) \text{ if } p \geq \frac{5}{9}$$

Problem 2

a)



b) > Forming clusters using C_0:

	Point	Distance from center
$C_1^0(-5,5)$:	$(-1,5)$	4
	$(-1,3)$	$\sqrt{4^2+2^2} = 4.24$
	$(-5,1)$	4
	$(-3,1)$	4.24
$C_3^0(5,-5)$:	$(1,-5)$	4
	$(1,-3)$	4.24
	$(3,-1)$	4.24
	$(5,-1)$	4
$C_2^0(-5,-5)$:	$(-1,-5)$	4
	$(-1,-3)$	4.24
	$(-3,-1)$	4.24
	$(-5,-1)$	4
$C_4^0(5,5)$:	$(1,5)$	4
	$(1,3)$	4.24
	$(5,1)$	4
	$(3,1)$	4.24

> New centroid locations based on above clustering

$C_1'(-2.5, 2.5)$ $C_2'(-2.5,-2.5)$ $C_3'(2.5,-2.5)$ $C_4'(2.5,2.5)$

c) The algorithm terminates once no more centers change locations

This does occur after the iteration we solved on the previous question because no points get reassigned to a different cluster after we reccompute centroids

Hence final output is plotted c_i^1 as marked & compilled and the same point to cluster groupings as compilled for the initial assignments

d) Optimal cluster centers

$c_1^* (-4, 0)$
$c_2^* (0, -4)$
$c_3^* (4, 0)$
$c_4^* (0, 4)$

marked on graph w/ ▲ symbol

summed (distances between centroid & points)

OPT error of k-means = $\sqrt{2} \cdot 4 \cdot 4 = 22.63$

Current error of k-means (c_i^1 assignments) = $4 \cdot (1.58 \cdot 2 + 2.91 \cdot 2) = 35.92$

Since our assignment of points to cluster results in a greater value of the error metric than OPT (described above), we conclude that this iteration yielded sub-optimal clusters.

PROBLEM 3

$y_i = \beta^T x_i + \epsilon_i$ where $\beta = (\beta_1, \ldots, \beta_d)$; $x_i \in \mathbb{R}^d$
$y_i \in \mathbb{R}$

a) β is a fixed d-dimensional vector
$\epsilon_i \overset{iid}{\sim}$ Laplace $(0,1)$ (i.e. $\epsilon_i \sim f_{\epsilon_i}(z) = \frac{1}{2} e^{-|z|}$)

Then, prove: likelihood $f(y_1, \ldots, y_n | \beta) = \left(\frac{1}{2}\right)^n \prod_{i=1}^{n} e^{-|y_i - \beta^T x_i|}$

We know: $\epsilon_i = y_i - \beta^T x_i$
& $f_{\epsilon_i}(z) = \frac{1}{2} e^{-|z|}$ $\Rightarrow f_{\epsilon_i}(y_i - \beta^T x_i) = \frac{e^{-|y_i - \beta^T x_i|}}{2}$

Likelihood of all the y_i's given β = product of individual likelihoods because y_i is dependent on x_i, ϵ_i which are both IID.

$$\Rightarrow f(y_1, \ldots, y_n | \beta) = \prod_{i=1}^{n} \left(\frac{1}{2}\right) e^{-|y_i - \beta^T x_i|}$$
$$= \left(\frac{1}{2}\right)^n \prod_{i=1}^{n} e^{-|y_i - \beta^T x_i|}$$

$\because$ Given (x_i, y_i) pairings & a coefficient vector β, the only random element is the noise ϵ_i which is sampled IID

b) $\hat{\beta}_{MLE} \overset{\Delta}{=} \underset{\beta}{\text{argmax}}\, f(y_1, \ldots, y_n | \beta)$

is equivalent to $\hat{\beta}_{MLE} = \underset{\beta}{\text{argmin}} \sum_{i=1} |y_i - \beta^T x_i|$

To maximize $f(y_1, \ldots y_n | \beta)$, we first convert to log-likelihood

$$\Rightarrow \underset{\beta}{\text{argmax}} \log\left(\left(\frac{1}{2}\right)^n \prod_{i=1}^{n} e^{-|y_i - \beta^T x_i|}\right)$$

$$\Rightarrow \underbrace{n \cdot \log\left(\frac{1}{2}\right)}_{\text{eliminate from argmax, because not affected by } \beta} + \sum_{i=1} \log e^{-|y_i - \beta^T x_i|}$$

$$\Rightarrow \text{argmax} \sum_{i=1}^{n} -|y_i - \beta^T x_i|$$

Since argmax of a sum of negative quantities is the same as argmin of the same positive quantities,

we can rewrite the last step as

$$\underset{\beta}{\operatorname{argmax}} \sum_{i=1}^{n} -|y_i - \beta^T x_i| \Rightarrow \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} |y_i - \beta^T x_i|$$

Hence Proved.

PROBLEM 4

$\mathcal{H}$ contains 3 disjoint intervals of $\mathbb{R}$ unioned together

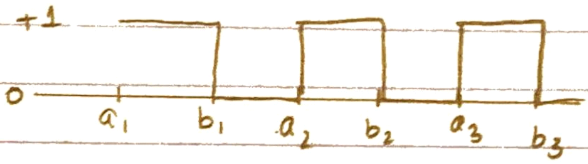
$$\mathcal{H} \triangleq \{h_{a_1,a_2,a_3,b_1,b_2,b_3} : a_i < b_i < a_{i+1} \ \forall i\}$$

where $h_{a_1,a_2,a_3,b_1,b_2,b_3}(x) = \begin{cases} 1 \text{ if } x \in [a_1,b_1] \cup [a_2,b_2] \cup [a_3,b_3] \\ 0 \text{ if } x \notin \rightarrow \text{ ranges} \end{cases}$

Show VC-dimension of $\mathcal{H} = 6$

a) Set $\{1,2,3,4,5,6\}$ is shattered by $\mathcal{H}$

Hint: Take Binary 6-tuple $\{c_1,c_2,c_3,c_4,c_5,c_6\}$ where $c_i \in \{0,1\}$ & provide function $h \in \mathcal{H}$ such that $h(i) = c_i$ for $i \in \{1,\dots,6\}$



The function $h \in \mathcal{H}$ which provides $h(i) = c_i \ \forall i \in \{1,\dots,6\}$ is one where we choose to include or exclude each (c_i, y_i) point by moving around the closest segment $[a_i, b_i]$, achieving any labeling for all 6 points (0 or 1)

We can think of the 3 intervals $[a_i, b_i] \ \forall i \in \{1,2,3\}$ as a set of axis-parallel rectangles which can be moved around to include or exclude c_i points

b) To prove: no set of 7 points in $\mathbb{R}$ that can be shattered by $\mathcal{H}$

Consider the following labeling of any 7 points in $\mathbb{R}$:

1	0	1	0	1	0	1
c_1	c_2	c_3	c_4	c_5	c_6	c_7

With this labeling, any set of 7 points in $\mathbb{R}$ cannot be shattered by $\mathcal{H}$ since we only have 3 intervals which can be mapped to +1 but 4 non-adjacent +1 labels to be classified for any set of points.

Thus, 6 is the largest k such that there exists a set C of size k which is shattered by $\mathcal{H}$

$\Rightarrow$ VC-dimension of given $\mathcal{H} = 6$