Question 4 – PAC, VC dimension, Bias vs Variance

Section 1

A circle (r,c) is defined by its center c and its radius r. Look at the following classifiers family:

$$\mathcal{H} = \{h_{r,c}\colon r \in \mathbb{R}, c \in \mathbb{R}^2\} \text{ where } h_{r,c}(x) = 1 \text{ iff x inside the circle (r,c)}$$

Find the VCdim of this class with full proof.

Section 2

Consider a training set $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ where $x_i \in \{0,1\}^3$. In other words, each sample has 3 Boolean features $\{X_1, X_2, X_3\}$. You are also given the classification rule $Y = (X_1 \wedge X_2) \vee (\neg X_1 \wedge \neg X_2)$.

We try to learn the function $f\colon X \to Y$ using a "depth 1 decision trees". A "depth-1 decision tree" is a tree with two leaves, all distance 1 from the root.

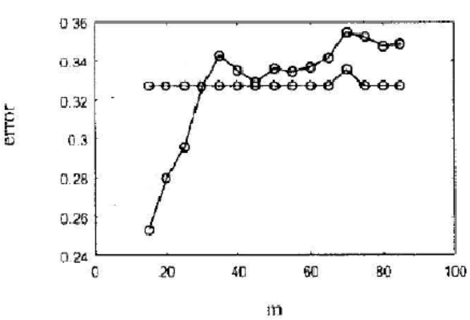
Analyze this problem and decide the appropriate sample complexity formula. Justify your answer.

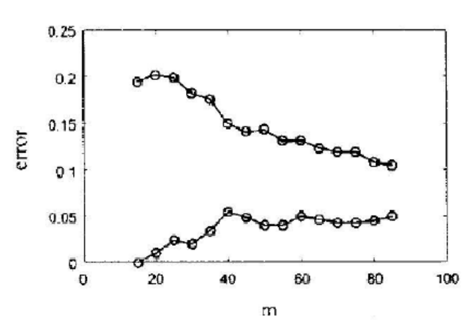
Section 3

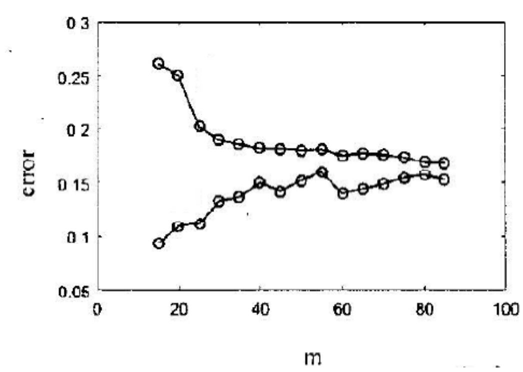
Dana was given a hard classification problem and she decided to use SVM with polynomial kernel with d=2,10,20. For each degree, she tried 15 to 85 training samples, with jumps of 5 (15, 20, ..).

The following graphs describe the train and test error for each d separately. However, she forgot which graph belongs to which d, and for each graph, what line is the train and the test.

Your task is to match each graph to the correct d and mark which lines are the test and the train. No explanation required.







Question 4 – PAC, VC dimension, Bias vs Variance

Section 1

A circle (r,c) is defined by its center c and its radius r. Look at the following classifiers family:

$$\mathcal{H} = \{h_{r,c}\colon\ r \in \mathbb{R}, c \in \mathbb{R}^2\} \text{ where } h_{r,c}(x) = 1 \text{ iff x inside the circle (r,c)}$$

Find the VCdim of this class with full proof.

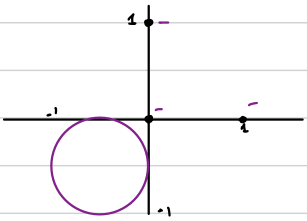
VCdim = 3.

To Prove this, we need to show that there exists a set of three points that can be shattered by $\mathcal{H}$.
consider the points: (0,0), (0,1), (1,0). there are $2^3 = 8$ different Possible labeling combinations and for each combination, ther exists a circle that can seperate the Points according to ther labels

Labeling:

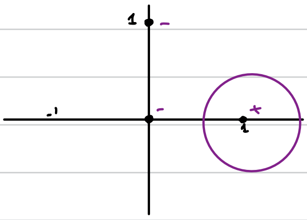
\# -1 => outside the circle
\# +1 => inside the circle

1) (0,0) = -1, (0,1) = -1, (1,0) = -1



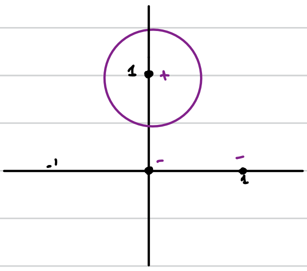
=> all the Points are outside the circle

2) (0,0) = -1, (0,1) = -1, (1,0) = 1



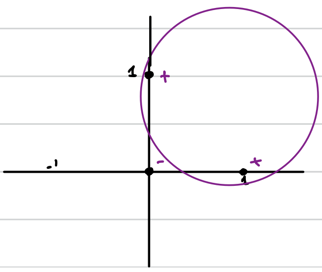
=> the circle includes (1,0) but exludes (0,0) and (0,1)

3) (0,0) = -1, (0,1) = 1, (1,0) = -1



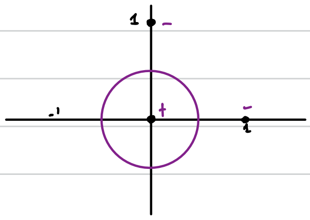
=> the circle includes (0,1) but exludes (0,0) and (1,0)

4) (0,0) = -1, (0,1) = 1, (1,0) = 1



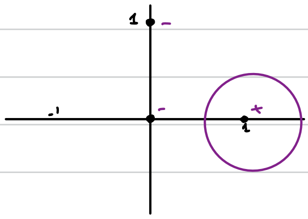
=> the circle includes (0,1) and (1,0) but exludes (0,0)

5) $(0,0)=1$, $(0,1)=-1$, $(1,0)=-1$



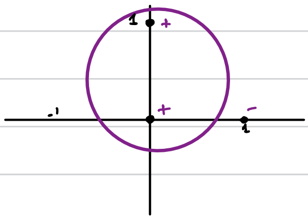
=> the circle includes $(0,0)$ but exludes $(1,0)$ and $(0,1)$

6) $(0,0)=1$, $(0,1)=-1$, $(1,0)=1$



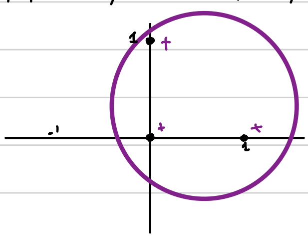
=> the circle includes $(0,0)$ and $(1,0)$ but exludes $(0,1)$

7) $(0,0)=1$, $(0,1)=1$, $(1,0)=-1$



=> the circle includes $(0,0)$ and $(0,1)$ but exludes $(1,0)$

8) $(0,0)=1$, $(0,1)=1$, $(1,0)=1$



=> the circle includes all the three points

- Now we know that three points can be shattered by circles, So $VCdim \geq 3$. We have to prove that $VCdim < 4$. We need to show that no set of 4 points can be shattered by circle.
  - consider any four points in $\mathbb{R}^2$: x, y, z, w

Point forming a Quadrilateral $(xyzw)$: its immposible to find a circle that can include any three points and exclude the fourth because a circle that includes three vertices of quadrilateral will necessarily include the forth vertex.

example: if $X:1$, $Y:1$, $Z:-1$, $W:-1$ => its immposible to draw a circle that includede X,Y and exlude Z and W.

So: $3 \leq VCdim < 4 \implies$ $VCdim = 3$

Section 2

Consider a training set $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ where $x_i \in \{0,1\}^3$. In other words, each sample has 3 Boolean features $\{X_1, X_2, X_3\}$. You are also given the classification rule $Y = (X_1 \land X_2) \lor (\neg X_1 \land \neg X_2)$.

We try to learn the function $f: X \to Y$ using a "depth 1 decision trees". A "depth-1 decision tree" is a tree with two leaves, all distance 1 from the root.

Analyze this problem and decide the appropriate sample complexity formula. Justify your answer.

- The classification rule $Y = (X_1 \land X_2) \lor (\neg X_1 \land \neg X_2) \Longrightarrow$ if X_1 and X_2 are true or X_1 and X_2 are false so it depends on X_1 and X_2
- A depth-1 discision tree is a simple decision tree with a root and two leaves so it makes a decision based on one feature only. it can be X_1 or X_2 not both
- ther are two possible depth-1 decision trees, so the size of the hypothesis class is 2 $\Longrightarrow |H| = 2$

$$m_H = \Theta\left(\frac{\log(|H|/\delta)}{\epsilon}\right)$$

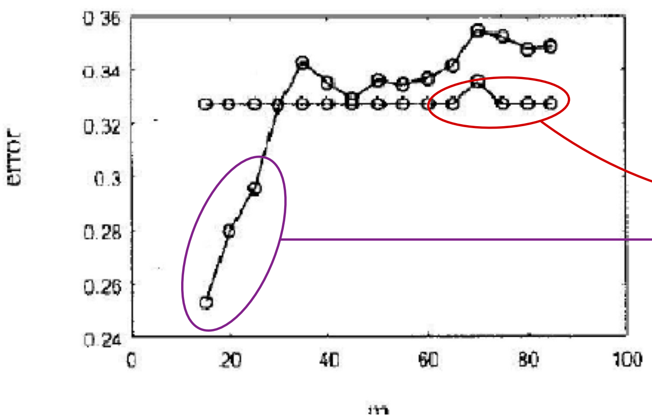
$|H| = 2$ So, $m_H = \Theta\left(\frac{\log(2/\delta)}{\epsilon}\right)$

Section 3

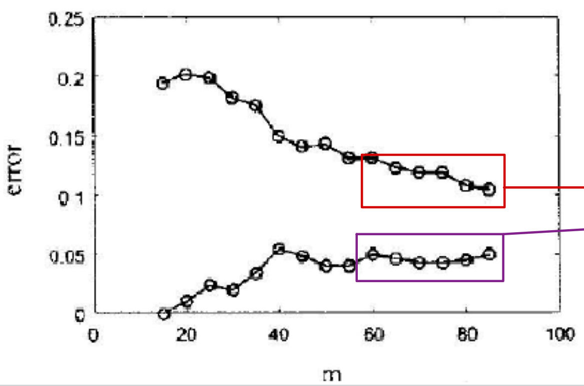
Dana was given a hard classification problem and she decided to use SVM with polynomial kernel with d=2,10,20. For each degree, she tried 15 to 85 training samples, with jumps of 5 (15, 20, ..).

The following graphs describe the train and test error for each d separately. However, she forgot which graph belongs to which d, and for each graph, what line is the train and the test.

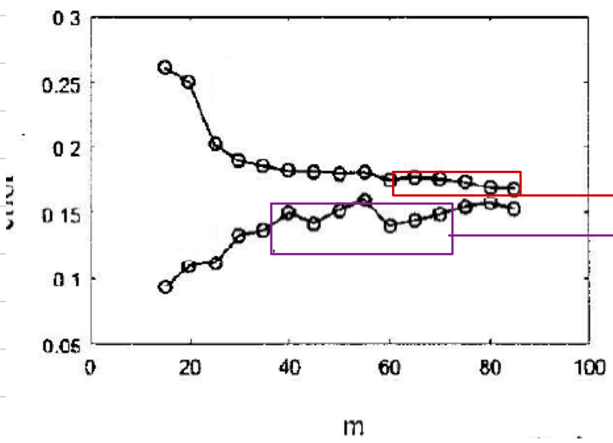
Your task is to match each graph to the correct d and mark which lines are the test and the train. No explanation required.



d = 2
Test error
train error



d = 20
Test error
Train error



d = 10
Test error
Train error