


Question 4 - PAC, VC dimension, Bias vs Variance

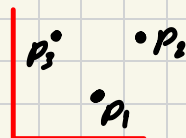
Section 1:

A circle (r, C) is defined by its center C and its radius r . Look at the following classifiers family: $H = \{h_{r,C} : r \in \mathbb{R}, C \in \mathbb{R}^2\}$ where $h_{r,C}(x) = 1$ iff x inside the circle (r, C) .

Find the VCdim of this class with full proof.

The VC-dim of this class is 3, first let's prove that we can shatter 3 points then we will show that we can't shatter 4 points.

given the following points p_1, p_2, p_3 as seen in the visual representation:

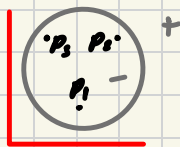
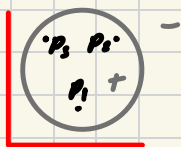


we have the following possibilities:

p_1	1	1	1	1	0	0	0	0
p_2	1	1	0	0	0	1	0	1
p_3	1	0	1	0	0	0	1	1

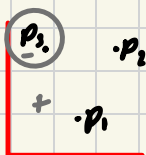
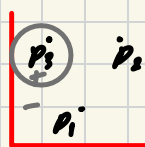
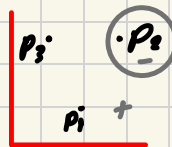
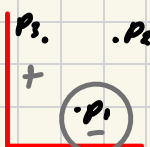
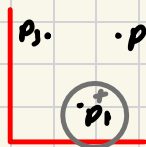
now in order to show how can we shatter the 3 points we will visualize 2 cases that can be generalized to include all the 8 possibilities that we listed before.

Case 1: Cover all 3 points



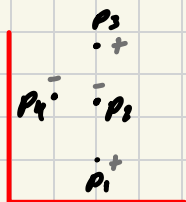
this case covers two possibilities from the grid.

Case 2: Cover only 1 point



as we can see in the graphs, we managed to shatter 3 points, and this applies to any 3 points

now we will show a case when we can't shatter 4 points given the points p_1, p_2, p_3, p_4 and the labels 1, 0, 1, 0 :



now as we can see in the graph, if we draw a circle with radius r that will include p_1 and p_3 then it will have to include p_2 , because any 2 point on a circle

the line between them will be inside the circle, and since point p_2 is on the line from p_1 to p_3 , and p_1, p_3 are inside the circle then p_2 is also inside the circle and therefore the combination $p_1=1, p_2=0, p_3=1, p_4=0$ is impossible to shatter with a circle

hence $VC-dim = 3$

Section 2:

Consider a training set $S = \{(x_1, y_1), \dots, (x_n, y_n)\}$ 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 $f: X \rightarrow Y$ 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.

We are given a depth-1 decision tree which can only check one feature at a time, now let's analyze the function Y .

According to its definition, the function returns true only if x_1 and x_2 have the same value and it doesn't consider x_3 . Hence the function $f: x \rightarrow Y$ cannot be used therefore if

we need to determine the complexity we need to use what we learned about PAC to learn a function with error

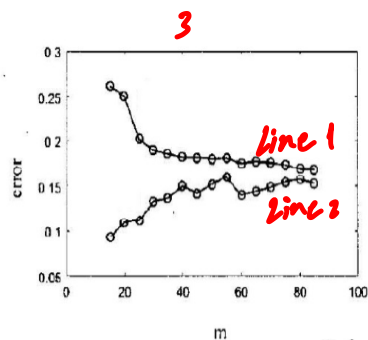
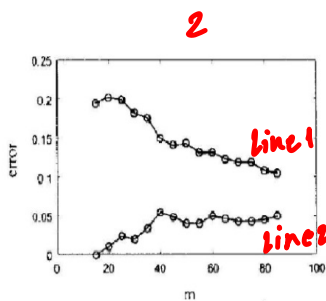
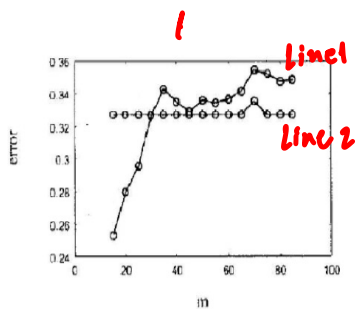
ϵ and confidence $1 - \delta$ we use the $m(\epsilon, \delta) = O\left(\frac{d + \log(1/\delta)}{\epsilon}\right)$

formula and now we substitute $d=3$ because the VC-dimension

is 3 in conclusion the complexity is: $m(\epsilon, \delta) = O\left(\frac{3 + \log(1/\delta)}{\epsilon}\right)$

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.



d	graph	train	test
2	2	line 2	line 1
10	3	line 2	line 1
20	1	line 1	line 2