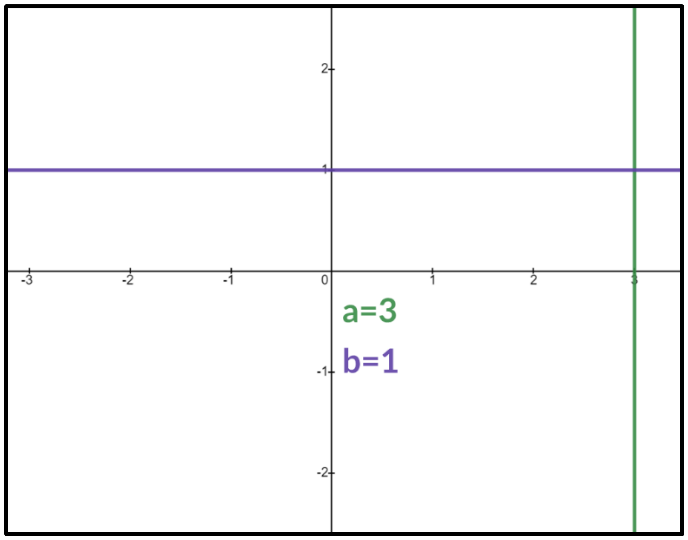
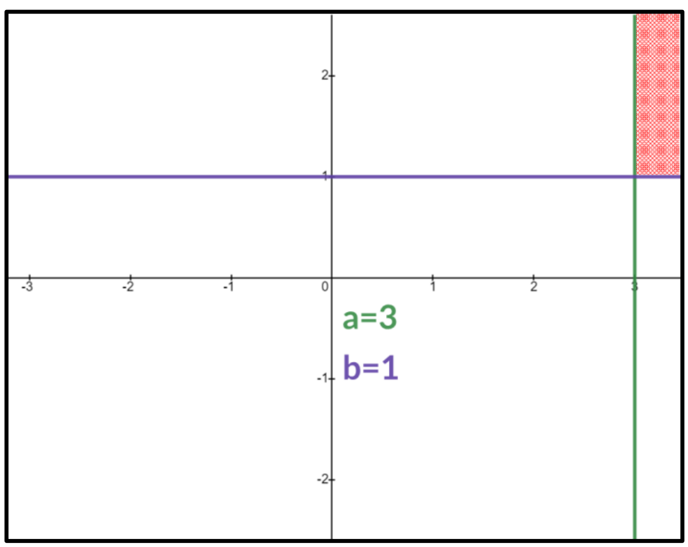
HW3 - Question 4 – Pac Learning

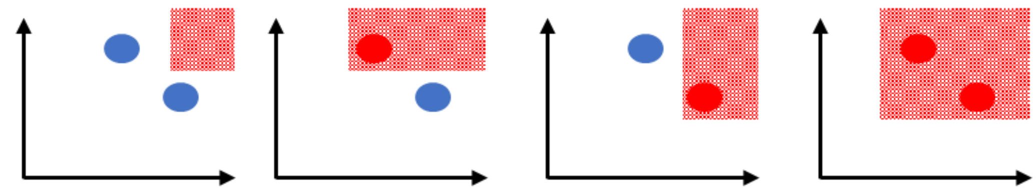
1. Suppose we have the classifier family: $H = \{h_{a,b}: a, b \in \mathbb{R}\}$ where $h_{a,b}(x, y) = 1$ iff $x \geq a \land y \geq b$. To find the VC dimension of this classifier family we need to understand it's capabilities, we find the maximum amount of data points classifiers from H can classify to every possible combination of classes. Let's take $h_{3,1}$ as an example:



In this example the area where the classifier returns 1 is when $x \geq 3 \land y \geq 1$, the area marked in red:



We will show that H shatters a group of 2 samples into every possible combination of classifications but not for every group of 3 samples, hence the VC dimension would be 2. For every 2 samples that are on a straight line with a negative slope we can classify the 2 points into every possible combination of classifications using classifiers from H.

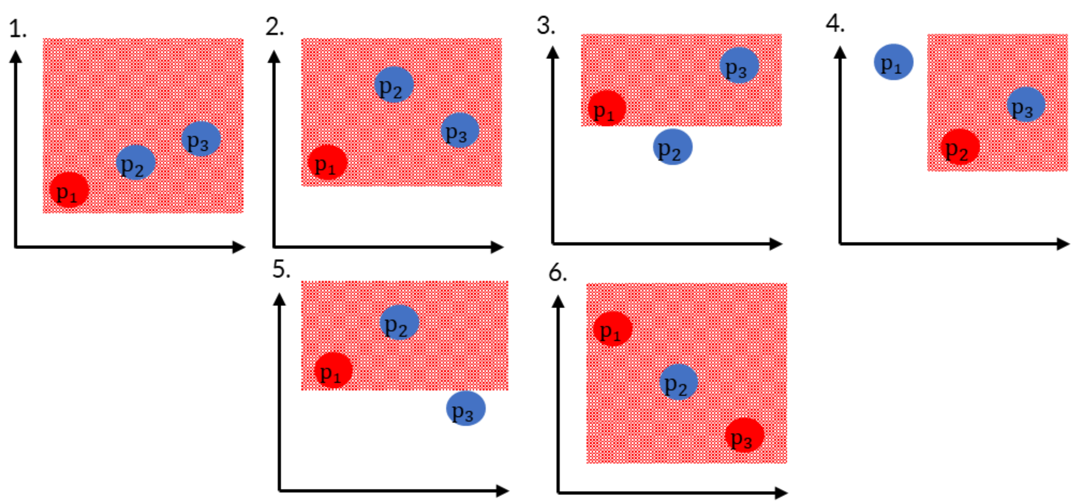


The figure shows that the statement is true; The classifier is the area marked in red. The red and blue points are two types of classes.

H shatters a group of 2 samples, hence the VC dimension of classifiers from H is at least 2. We will now show that no 3 samples can be classified into every possible combination. We have 6 cases to consider, we will assume that $x_1 \leq x_2 \leq x_3$:

1. $y_1 \leq y_2 \leq y_3$
2. $y_1 \leq y_3 \leq y_2$
3. $y_2 \leq y_1 \leq y_3$
4. $y_2 \leq y_3 \leq y_1$
5. $y_3 \leq y_1 \leq y_2$
6. $y_3 \leq y_2 \leq y_1$

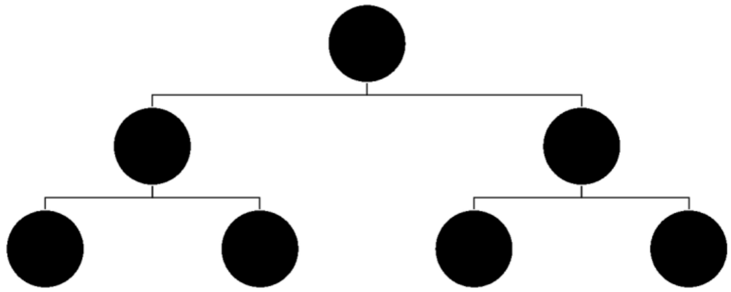
For each of these cases we will show that it is not possible to classify the points into every combination of classifications. The following scheme shows exactly that:



Each figure shows how we cannot shatter each case mentioned above.

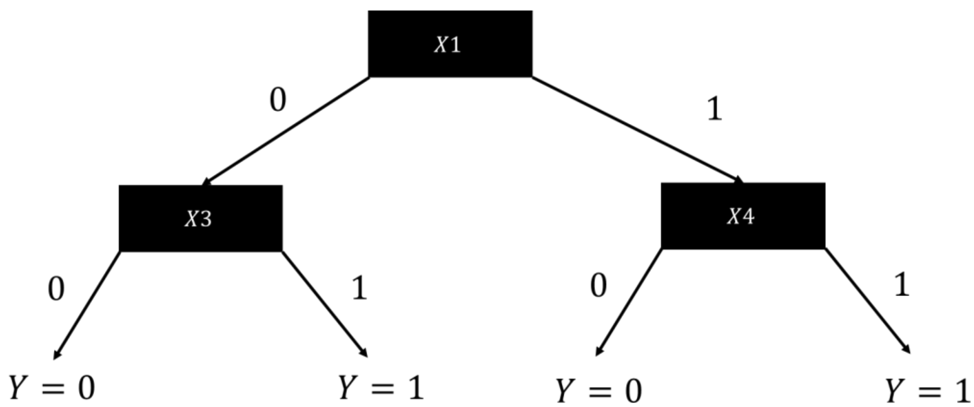
It is impossible to shatter **every** 3 points with the classifier family H, hence the VC dimension of H is 2. ▪

2. $X = (X_1, X_2, X_3, X_4)$ Boolean variables, Y is calculated by: $(X_1 \wedge X_4) \vee (\neg X_1 \wedge X_3)$, i.e Y is a binary classifier. We try to learn $f: X \to Y$ using a decision tree of depth 2. Let's define $H = \{h: h\ is\ a\ decision\ tree\ of\ depth\ 2\}$.



A decision tree of depth 2.

Since H contains a combination of decision trees, there is a finite amount of trees we can create of depth 2 and 4 leaves, so H is finite. We will show that the claim of realizability applies to the problem by showing the decision tree that classifies all points correctly. The truth table for Y is the following:



This decision tree is of depth 2 and classifies each prediction precisely as intended, there isn't any thresholds implemented but equality to 0 or 1.

Since the assumption of realizability holds, and H is finite, the lower bound of the classifier family is $m \geq \frac{\log(|H|/\delta)}{\epsilon}$ when H is the group of all decision trees of depth 2.