1. The VC dimension of hypothesis space H1 is larger than the VC dimension of hypothesis space H2. Which of the following can be inferred from this?

A. The number of examples required for learning a hypothesis in H1 is larger than the number of examples required for H2.

B. The number of examples required for learning a hypothesis in H1 is smaller than the number of examples required for H2.

C. No relation to number of samples required for PAC learning.

$$m \geq \left(4 \log_2\left(\frac{2}{\delta}\right) + 8\, VC(H) \log_2\left(\frac{13}{\epsilon}\right)\right).$$

Where m is the number of examples sufficient for PAC learning and $VC(H)$ is the VC dimension of hypothesis space H.

Keeping ϵ and δ constant. $m \propto VC(H)$.

$\therefore$ If m_1 is the number of examples reqd. for learning a hyp. in H_1 and m_2 for H_2. Then $m_1 > m_2$. Since, $VC(H_1) > VC(H_2)$.

$\boxed{\text{Option : A}}$

2. The Bayes Optimal Classifier

A. is an ensemble of some selected hypotheses in the hypothesis space.

B. is an ensemble of all the hypotheses in the hypothesis space.

C. is the hypothesis that gives best result on test instances.

D. none of the above

Refer to lecture notes and videos.

$\boxed{\text{Option B}}$

3. For a particular learning task, if the requirement of error prameter ϵ changes from 0.1 to 0.01. How many more samples will be required for PAC learning?

A. Same

B. 2 times

C. 10 times

D. 100 times

$m \geq \frac{1}{\epsilon}\left[\ln(|H|) + \ln\left(\frac{1}{\delta}\right)\right]$ labelled examples are sufficient so that with prob. $1-\delta$, all $h \in H$ with $error_D(h) \geq \epsilon$ have ~~errors~~ $error_S(h) > 0$.

$\therefore\ m \propto \frac{1}{\epsilon} \Rightarrow \frac{m_1}{m_2} = \frac{\epsilon_2}{\epsilon_1} \Rightarrow m_2 = m_1 \cdot \frac{\epsilon_1}{\epsilon_2} = \frac{0.1}{0.01} = 10$. times.

$\boxed{\text{Option - C}}$

4. Suppose the VC dimension of a hypothesis space is 4. Which of the following are true?

A. No sets of 4 points can be shattered by the hypothesis space.
B. Atleast one set of 4 points can be shattered by the hypothesis space.
C. All sets of 4 points can be shattered by the hypothesis space.
D. No set of 5 points can be shattered by the hypothesis space.

Options B and D.

Refer to lecture notes and definition of VC dimension as given in the book "Machine Learning" by Tom Mitchell.

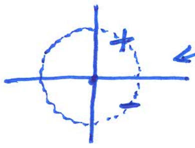
5. Consider a circle in 2D whose center is at the origin. What is its VC dimension?

A. 1
B. 2
C. 3
D. 4

Option B.

For a one point set, we can shatter any configuration perfectly using a circle centered at the origin.

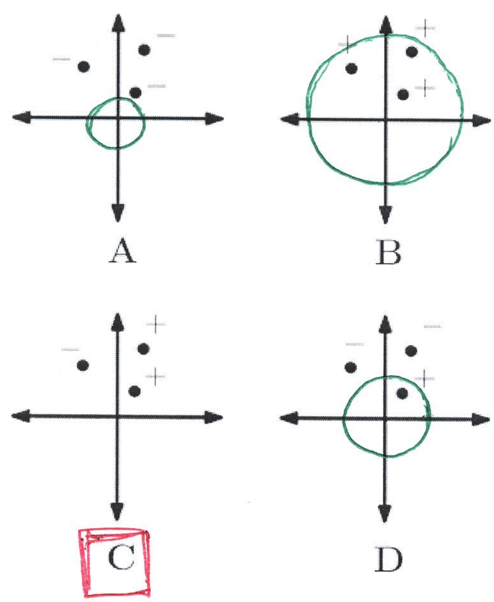
However, given a ~~po~~ set of a pair of points where the points belong to two different classes cannot be shattered by a circle centered at the ~~g~~ origin if the points are equidistant from the center.



The dotted line shows a circle centered at the origin. ~~Howev~~ However, if the points are not equidistant from the center, they can be shattered perfectly.

Hence, there exists some configuration of 2 points that can be shattered by a ~~circle~~ circle centered at the origin. ~~Hence,~~ Hence, VC dim of a circle centered at origin is 2.

6. Under a binary classification setting, which of the following sets of three labeled points cannot be shattered by a circle centered at the origin?



Assuming that the points outside the circle are -ve and those inside are positive and ~~so~~ since this is a binary classification problem, without any loss of generality, we find the the answer is $\boxed{C}$

The circles centered at the origins in A, B and D shows that the points are shattered by circles at ~~the~~ centered at the origin.

7. Given a set of 4 points (x1,-), (x2,-), (x3,+) and (x4,+), Adaboost algorithm is used to train a weak classifier on this data. In the first iteration, the weak classifier wrongly classifies x3 and correctly classifies the other three points. In the second iteration it wrongly classifies x1 and correctly classifies the other points. Assuming uniform initial weight distribution (D1) over the data points, what is the weight distribution for the 3rd iteration (D3).

A. 0.5, 0.3, 0.1, 0.1
B. 0.1, 0.3, 0.5, 0.1
✓ C. 0.5, 0.1, 0.3, 0.1
D. 0.1, 0.1, 0.3, 0.5

Steps:

Let $m = 4$, x_i = data point and y_i = class label of x_i and $i \in \{1, 2, 3, 4\}$.
h = weak hypothesis.

1. $D_1 = 0.25, \; 0.25, \; 0.25, \; 0.25$

$t = 1.$

2. $\epsilon_1 = \sum_{i=1}^{m} D_1(i)\, \delta(h_1(x_i) \neq y_i)$ $\left[\begin{array}{l} \delta(x \neq y) = 1, \text{ if } x \neq y \\ \quad\quad\quad\;\; = 0, \text{ if } x = y \end{array} \right]$ $m = 4$

3. $\alpha_1 = \frac{1}{2} \ln \left(\frac{1 - \epsilon_1}{\epsilon_1} \right)$

4. $Z_1 = \sum_{i=1}^{m=4} D_1(i)\, e^{-\alpha_1 y_i h_1(x_i)}$

5. $D_2(i) = \dfrac{D_1(i) \cdot e^{-\alpha_1 y_i h_1(x_i)}}{Z_1}$; for all $i \in \{1, 2, 3, 4\}$

After you get D_2 repeat the steps from 1. to 5. to get D_3.

Option C