Assignment - 7

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.

   Answer: A (From the definition of VC dimension, VC dimension is directly proportional to m (no of training samples)

2. In ensemble learning, you aggregate the predictions for weak learners, so that an ensemble of these models will give a better prediction than prediction of individual models. Which of the following statements is / are true for weak learners used in ensemble model?

   1. They don't usually overfit.
   2. They have high bias, so they cannot solve complex learning problems
   3. They usually overfit.

   A) 1 and 2
   B) 1 and 3
   C) 2 and 3
   D) Only 1

   Solution: (A) Weak learners are sure about particular part of a problem. So, they usually don't overfit which means that weak learners have low variance and high bias.

3. 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

Answer: B (Please refer to the lecture notes)

4. For a particular learning task, if the requirement of error parameter 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) 1000 times

   Answer: C (We know that, m is inversely proportional to ϵ.
   $m1 * \epsilon 1 = m2 * \epsilon 2$
   $m2 = (m1 * \epsilon 1)/ \epsilon 2$
   $m2 = (m1 * 0.1)/0.01$
   $m2= 10*m1)$
5. Data scientists always use multiple algorithms for prediction and they combine output of multiple machine learning algorithms (known as "Ensemble Learning") for getting more robust or generalized output which outperform all the individual models. In which of the following options you think this is true (Choose best possible answer)?
   A) Base models having the higher correlation.
   B) Base models having the lower correlation.
   C) Use "Weighted average" instead of "Voting" methods of ensemble.
   D) Base models coming from the same algorithm
   Answer: B (Please refer to the lecture notes)
6. 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.

   Answer: B, D (Please refer to the lecture slides)

7. Computational complexity of classes of learning problems depends on which of the following?
   A) The size or complexity of the hypothesis space considered by learner

B) the accuracy to which the target concept must be approximated

C) the probability that the learner will output a successful hypothesis

D) All of the above

Answer: D (The complexity depends on size and accuracy expected and the probability of a successful hypothesis as well as error allowed.)

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

   A) 1

   B) 2

   C) 3

   D) 4

   Answer: B (If 2 points are not equidistant from the origin, they can be shattered by a circle. Therefore, VC dimension of a circle centered at origin is 2.)

9. Which among the following prevents overfitting when we perform bagging?

   A) The use of sampling with replacement as the sampling technique

   B) The use of weak classifiers

   C) The use of classification algorithms which are not prone to overfitting

   D) The practice of validation performed on every classifier trained

   Answer: A (Bagging employs sampling with replacement to sample a subset of points from the original dataset. This means regions that are denser (and thus have more evidence) appear in every sample while outlier regions (very sparse) occur in only a fraction of classifiers, thus reducing their effect on the whole.

10. VC dimension for conjunctions of n Boolean literals is:

    A) Atleast n

    B) Atmost n

    C) Can’t say

    D) None

    Answer: A (Please refer to the lecture slides)