**Problem SET 1**

Name:-Sahil Soni

Email:-Ssoni41@gatech.edu

GTID:- 903459126

Q.) Suggest a lazy version of the eager decision tree-learning algorithm ID3. What are the advantages and

Disadvantages of your lazy algorithm compared to the original eager algorithm.

**Ans.**

LazyDT is one of the lazy version of eager decision tree learning algorithm ID3. It’s one of version can be used with hill climbing approach. In LazyDT , for each instance separate single tree is built from root to leaf node. While in case of, Normal Dt only one “generalized tree” has been formed from training Data and all predications are based upon that Tree.

**Advantages:-**

* Less Computation time for Training as all decisions are made during prediction phase.
* One of main purpose of Lazy Dt is to solve Replication, fragmentation and data partition associated with Decision Tree. This is because splitting in case of Lazy Dt is corresponding to each separate test case. While in Dt splitting is done the average of each node (which new test instance will not be knowing in case of DT)
* It better for missing/noisy data because it leafs never split on missing Values also.
* Its works well when we have more training data.
* We can add more complex target functions.

**Disadvantages:-**

* More computation time required for Predictions because decision on outcome depends on new query Test Data. It is because it is isefficient memory indexing.
* If we have too many irrelevant features, predictions will not be accurate for test data because of same “curse of dimensionality”.

Q.) Imagine you had a learning problem with an instance space of points on the plane and a target function

that you knew took the form of a line on the plane where all points on one side of the line are positive and

all those on the other are negative. If you were constrained to only use decision tree or nearest-neighbor

learning, which would you use? Why?

**Ans.**

As we know target function and if target function is not complex, my preference will be to use nearest neighbor. Because its low dimensionality problem where plane has 2-d as it is forming a line .We also know one side of plane is positive so we can easily choose distance metric. So because it has 2 dimensions, easy to choose distance metric and we already know target function. So this make be believe I should try KNN here.

However, I can also try Decision tree if we have to predict lot of test data, which in case of KNN will take time. However, as data information is not present so I will stick to KNN.

Q.) Give the VC dimension of the following hypothesis spaces. Briefly explain your answers.

Ans.

1. An origin-centered circle (2D)

Although for a circle VC Dimensions is 3 because but as this is origin centered VC dimension is at least 2 because any two point on circle whose distance is not same from Origin can be shattered.

2. An origin-centered sphere (3D)

The VC dimension in this case is also 2 because of same reason as above.

References:-

<https://alliance.seas.upenn.edu/~cis520/dynamic/2018/wiki/index.php?n=Recitations.VCDim#toc3>

<https://www.ke.tu-darmstadt.de/lehre/arbeiten/bachelor/2012/Korfhage_Nikolaus.pdf>

<https://pdfs.semanticscholar.org/71e6/7f0358c76fa3950b79daf0d9ddc0b0fe9c76.pdf>