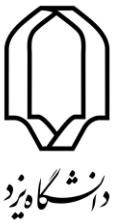
In the name of Allah



Computer Engineering Faculty of Yazd University
Artificial Intelligence

Machine Learning Course

Homework 2

(Decision Tree)

Instructor: Dr. Yazdian

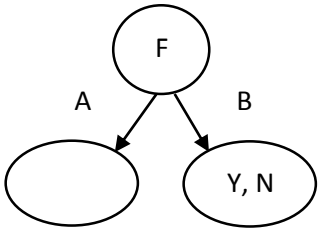
Shiva Zeymaran

Student ID: 40109434

Fall 2022

A. Theoretical Example

1. Consider the following binary decision tree which created by randomly choosing nodes with IG>0:



We assume that one of the leaves generated by using F as node, contains no training data. Now, calculate IG of node F as follows:

$$H(C) = -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1$$

$$H(C|F) = 0 * H(C|F = A) + 1 * H(C|F = B) = H(C|F = B) =$$
$$= -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1$$

$$IG(C, F) = H(C) - H(C|F) = 1 - 1 = 0$$

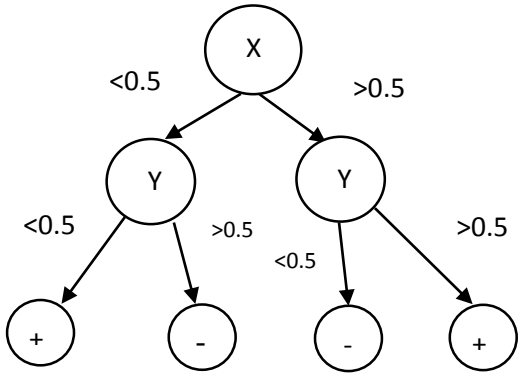
The IG in this case is equal to Zero and it is against the condition of question that was IG>0. Therefore, our assumption is wrong and each leaf of the tree in this case, contains at least one training data.

2. The maximum number of leaves in a decision tree is equal to the number of training data (that is 'n' in this problem). Because it is happen when we have one sample in each leaf of the tree. Note that this tree is over-fitted to our training data.
When we use IG for selecting nodes as described in the last question, the answer will be 'n' similarly. Because as we show in the last section, each leaf of this type of tree contains at least one training data and the maximum number of leaves will happen when we have only one sample in each of them.

3. We have two classes in this problem. So, we will have a binary decision tree. In the worst case, we can create a tree that each sample is in each leaf of the tree. If we have 'n' samples, a binary tree with 'n' leaves has

depth of $\log_2 n$. Therefore, at least we can find a binary tree the depth of $\log_2 n$ that can classify truly.

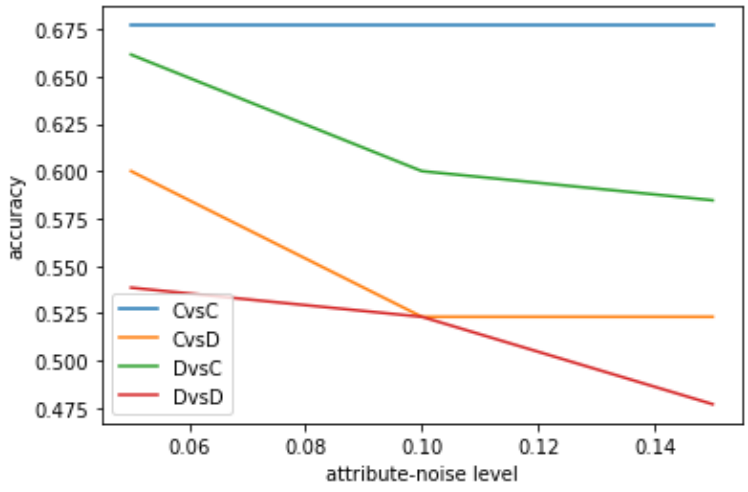
4. Assume 4 points in this space: (n = 4)
   (0.25, 0.25) – (0.25, 0.75) – (0.75, 0.25) – (0.75, 0.75)
   Also consider that (0.25, 0.25) and (0.75, 0.75) are in class '+' and two other points belongs to class '-'.
   The minimal decision tree for this example has 4 leaves that each sample is in one of these leaves. Also, the tree contains n-1 = 3 internal nodes:



B. Analysis the effect of Attribute Noise

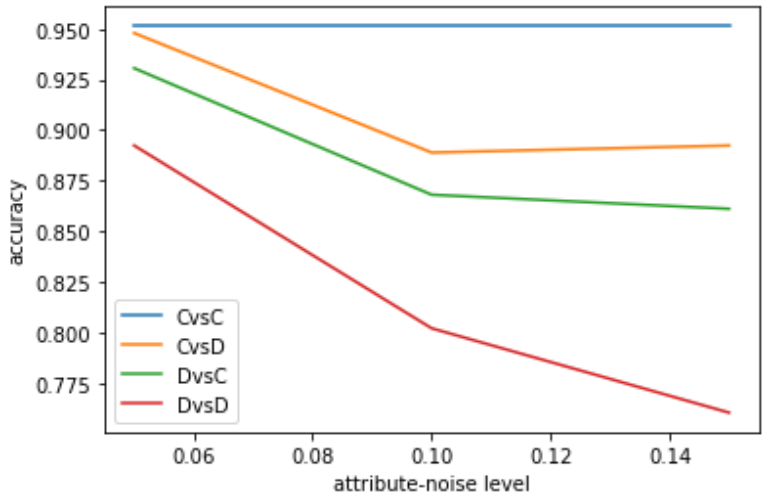
1. There is the result of plot Accuracy values by changing the attribute noise level value (5%, 10%, and 15%) for **'glass' dataset**:

```
CvsC Accuracy: [0.676923076923077, 0.676923076923077, 0.676923076923077]
DvsC Accuracy: [0.6615384615384615, 0.6, 0.5846153846153846]
CvsD Accuracy: [0.6, 0.5230769230769231, 0.5230769230769231]
DvsD Accuracy: [0.5384615384615384, 0.5230769230769231, 0.47692307692307695]
```



For **'tic-tac-toe' dataset**:

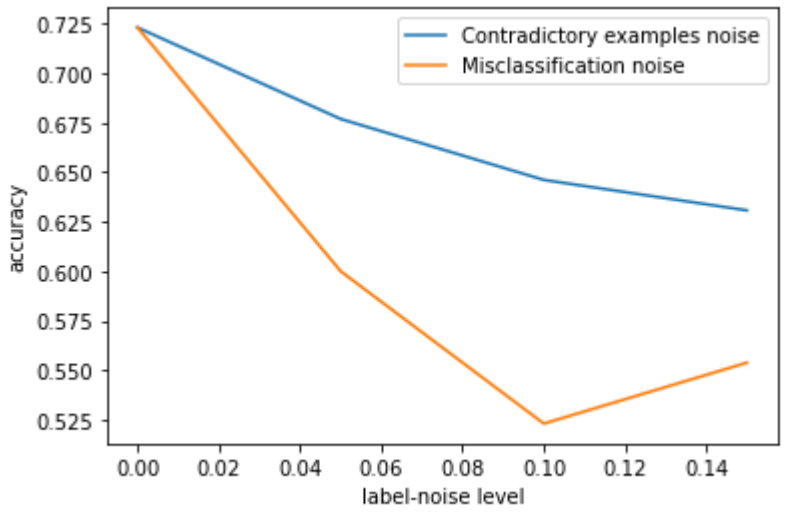
```
CvsC Accuracy: [0.9513888888888888, 0.9513888888888888, 0.9513888888888888]
DvsC Accuracy: [0.9305555555555556, 0.8680555555555556, 0.8611111111111112]
CvsD Accuracy: [0.9479166666666666, 0.8888888888888888, 0.8923611111111112]
DvsD Accuracy: [0.8923611111111112, 0.8020833333333334, 0.7604166666666666]
```



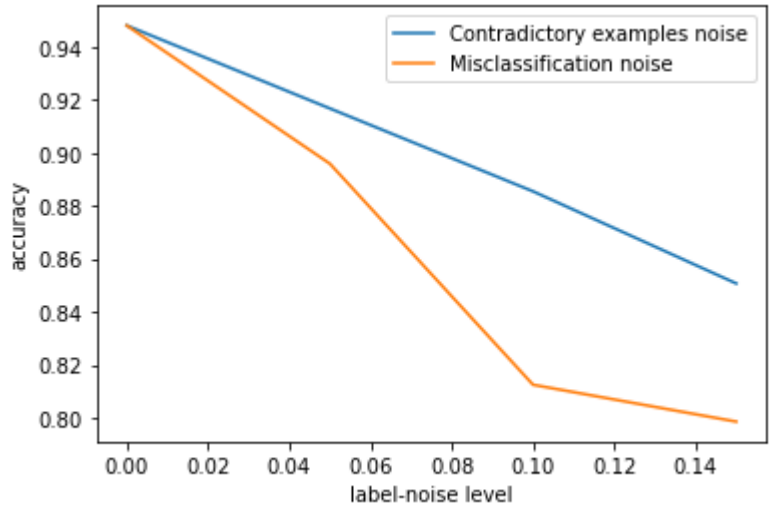
2. We understand from both plots that when we have clean training set and clean test set, the accuracy is the best possible value. Because we don't have the effect of noise in our model.
   Also in both plots, dirty training set and dirty test set is the worst case with the minimum accuracy values.
   In two other cases (CvsD and DvsC) we have some differences in two above plots. In 'glass' dataset, the DvsC has greater values for accuracy with different noise levels but in 'tic-tac-toe' dataset, the CvsD is better.
   In general, it is more sensible to have better results in CvsD because the model is created from clean dataset so it can predict better. In this case, you can see that the accuracy in CvcC for 'tic-tac-toe' is about 95% but this is 67% for 'glass' dataset. Therefore, we can say we have better dataset in 'tic-tac-tor' to predict with decision tree. Also, note that this dataset has **Categorical feature** values that is better for decision tree problem. But the 'glass' dataset has Continues feature values.
   Also note that all curves for dirty train/test cases, are decreasing by increasing the value of noise level.

C. Analysis the effect of Class-Label Noise

1. There is the result of plot accuracy values by changing the value of attribute noise level (5%, 10%, and 15%) for **'glass' dataset**:



For **'tic-tac-toe' dataset**:



2. The first value of both curves are the same because in the first step we do not have any noises. Then, in both types of class label noises, increasing the value of label noise will decrease the accuracy.
   We understand form both plots that Misclassification noise is more harmful for the dataset than contradictory examples. Because the accuracy decreases very much in this type of class noise.

3. Class noise is more harmful because having wrong labels have effect on whole model and cause wrong learning process; but, attribute noise causes the wrong learning on those specific samples and it is not harmful for the whole dataset.
   Also you can see that for example for 'tic-tac-toe' dataset, the accuracy decreased to less than 80% in the case of misclassification noise with the label-noise level of 15%; but its accuracy decreased to 86% in the case of attribute noise with the label noise level of 15% when we have dirty train set and clean test set. (We choose DvsC because it is the best case for understanding the effect of attribute noise on dataset.)