Richard Vargas

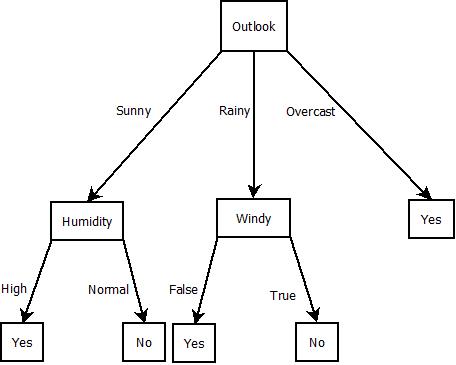
CSE 516 Machine Learning

005859278

Dr. Zhang

Assignment #2

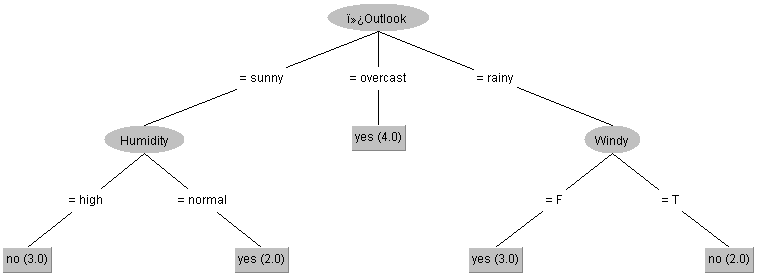
1.

Here I decided to set tree root node to Outlook because I saw that instances of Overcast all led to a Yes.

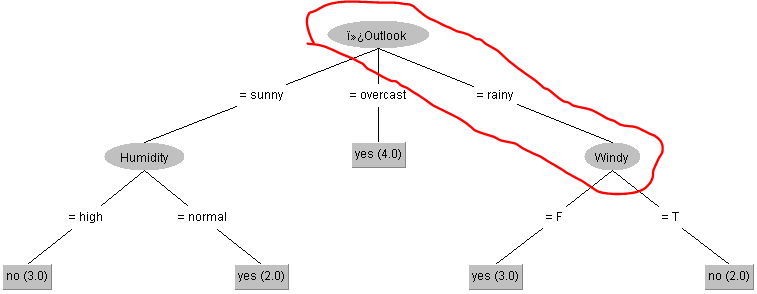


This is the Gini of the root node (parent node) where is the amount of total No’s in the set, and are the amount of Yes’s out of the whole set. The Gini of the branches from the root node are the following:

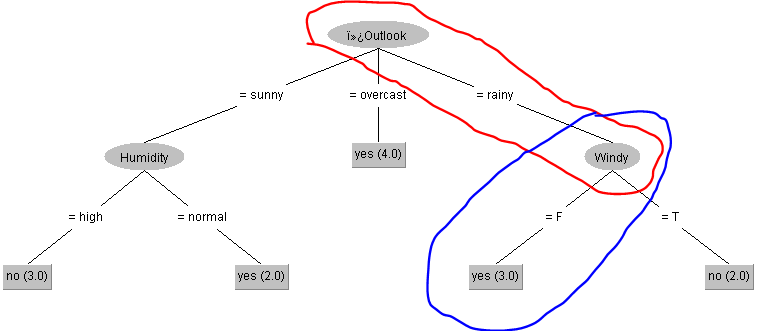
b.



X = (Outlook = rainy)



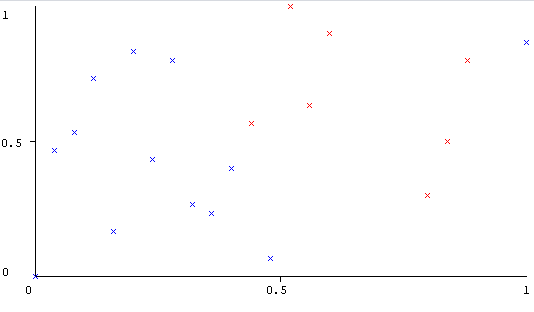
X = (Windy = False)

Therefore, X is a day to play.

2. The new instance will compute the distance from the set of test instances and based on some value K, the label is decided based on the majority label from K. Let be the new instance and be the vector of attributes, such that . For each instance in the training set, we shall compare the Euclidean distance of our test instance by using the following equation,

.

Here is the normalized scatter plot where X-axis is speed and the Y-axis is the weight. The No label is blue and the Yes label is red



Ran out of time for #3, Apologies