ETGG 4803
Due 8 May 2020

Quiz #4
Name __Thomas Gilman____

Answer the questions in the space provided. Be neat, clear, and thorough.

1. Consider the use of *cross-validation* in a learning model.
   a. What is cross-validation?

      Cross Validation is a sampling procedure in where a given limited sample of test data can be used to create a variable and determine the accuracy of a given machine learning process.

   b. What is the purpose of cross-validation?

      The purpose of Cross validation is to evaluate a machine learning model given a limited amount of data.

   c. What is the difference between cross-validation and *k*-fold cross-validation?

      Cross Validation is a procedure that resamples data. So if you have a training set you can create a variable from it and it is then evaluated to get the accuracy of that variable. Another form of this is k-fold Cross-validation in which every test serves as a training and test data point which are split into k subsets with k test rounds; the average score should be better than the earlier cross validation.

   d. What type of cross-validation was used with the Room Occupancy data?

      Holdout cross-validation

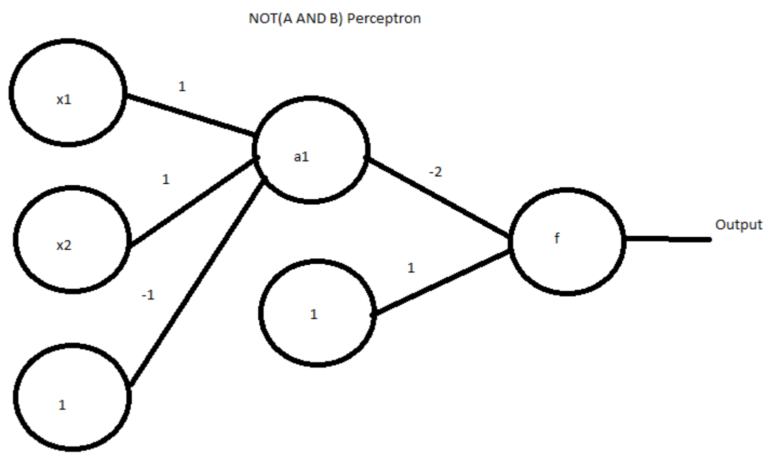
2. Explain the meaning of the term *linearly separable*. Is all data linearly separable? Why or why not?

   Being linearly separable means that there is deviation, a difference, or distance between two pieces of data if the points can be separated by a line/plane. If the data cannot be separated by a line or hyperplane, the data is not linearly separable. Given the meaning of linearly separable and the possibility that some data may not be able to be separated, this means that not all data is linearly separable.

3. A neural net is designed to recognize handwritten vowels. How many neurons should be in the output layer of the neural net?

6 Neurons total in output layer, one for each vowel (5) and a +1 bias Neuron

4. Create a perceptron that gives the correct output for the logical statement NOT (*A* AND *B*) or explain why it is not possible.



f(1, 1) = 0

f(0, 1) = 1

f(1, 0) = 1

f(0, 0) = 1

5. A robot in a lumber yard must learn to distinguish oak from pine. The following data has been obtained.

Example	Density	Grain	Hardness	Class
1	Heavy	Small	Hard	Oak
2	Heavy	Large	Hard	Oak
3	Heavy	Small	Hard	Oak
4	Light	Large	Soft	Oak
5	Light	Large	Hard	Pine
6	Heavy	Small	Soft	Pine
7	Heavy	Large	Soft	Pine
8	Heavy	Small	Soft	Pine

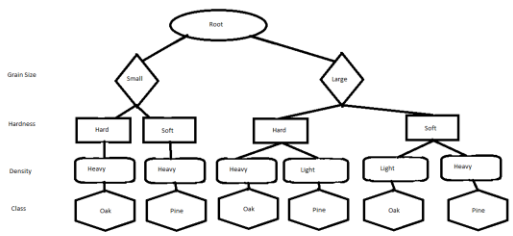
You choose a Decision Tree classifier.

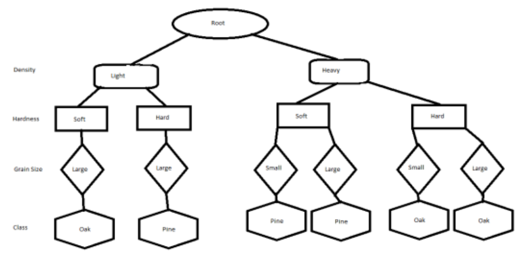
a. Which root (Density, Grain, or Hardness) gives the most information?

All three roots can give the exact same amount of information. The root that I am choosing to use for my decision tree is the density in this case.

b. Draw the decision tree with the root from **a**.

The decision tree for using the Grain as the root is on the left, while the decision tree for using the Density is on the right. If you were to use the Hardness as the root, it would look exactly like the Density decision tree but otherwise swapping places between The Density and Hardness nodes.





c. Use your decision tree to classify these new examples as Oak or Pine.

i. [Density = Light, Grain = Small, Hardness = Hard]
Pine

ii. [Density = Light, Grain = Small, Hardness = Soft]
<u>Oak</u>

6. Suppose Paul wears a kilt about once a year and Jason wears a kilt about once every five years. Jason walks down a particular sidewalk every day but Paul only walks on that sidewalk once every two days.

Paul Sidewalk Times: 365.25 / 2 = 182.625 days a year
Jason Sidewalk Times: 365.25 days a year

Total Outcomes = 365.25 + 182.625 = 547.875
Total Outcomes is the number of times Jason walks on the sidewalk in a year plus the number of times Paul walks on the sidewalk in a year.

Paul Wears a kilt: 1 / 365.25 = 0.0027378508
Jason Wears a kilt 1 / 1826.25 = 0.0005475702

a. You see one of the two walking down the sidewalk, but you are too far away to tell who it is and you cannot see what the person is wearing. What is the probability that it is Jason?

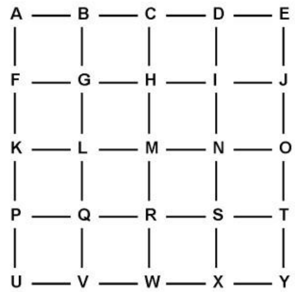
Probability of Jason walking on the sidewalk is the number of times Jason walks on the sidewalk in a year divided by the number of Total Outcomes (Paul walks + Jason walks).
The probability that it is Jason is 365.25 / 547.875 = 2/3 = .66666667 = 66.66%

b. The person gets a bit closer and you can see he is wearing a kilt. What is the probability that it is Paul?

P(Paul | kilt) = P(Paul | kilt) * P(kilt) / P(Paul)
P(kilt | Paul) = Times wears skirt / (Times Paul wears skirt + Times not Paul wears skirt) = (1/365.25) / ((1/365.25) + (1/1826.25)) = .8333
P(Paul | kilt) = ((1/365.25) / ((1/365.25) + (1/1826.25)) * (1/365.25)) / (182.625 / 547.875) = 0.00684462

The probability of the person walking down the street wearing a kilt being Paul is 0.68446269%.

7. In the following diagram, each data point *A* through *Y* has attributes (*x*, *y*) corresponding to its coordinates in the grid.



a. What are the five nearest neighbors of data point *M* using Euclidean distance? Break ties alphabetically.

G, H, L, N, R are the closest neighbors using Euclidean distance.

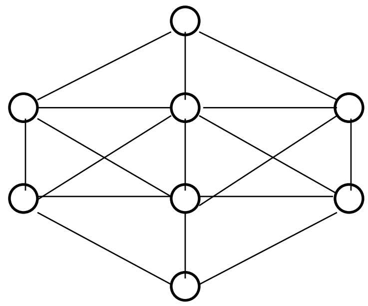
b. What are the five nearest neighbors of *M* using Manhattan distance? Again, break ties alphabetically.

C, H, L, N, R are the closest neighbors using Manhattan distance.

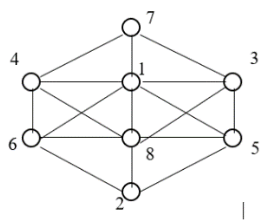
c. Suppose data points *A* through *L* are in Class 1 and data points *N* through *Y* are in Class 2. Give the classification of *M* using *k*-NN with Manhattan distance for:

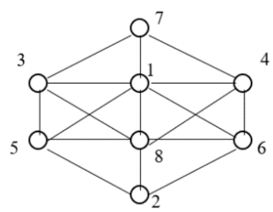
i. k = 1 Class **1**
ii. k = 3 Class **1**
iii. k = 5 Class **1**
iv. k = 7 Class **1**

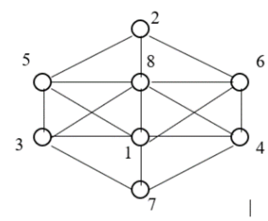
8. Each vertex in the graph below is to be labelled with a unique digit from 1 to 8 so that adjacent vertices are *not* labelled with consecutive digits.

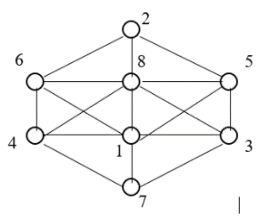


Give all possible solutions and convince me that you have found all of them.









There are only the 4 combinations as you can only have 1 and 8 in the middle, and 2 and 7 on the opposite ends as the two center nodes access all but one of the nodes which is the one opposite the other center node. This means you can flip the numbers from the left to the right side, and you can mirror the same combinations on the opposite side as shown in the pictures. (mirroring up and down in between the middle nodes).

9. Consider the following diagram. The goal is to connect each small box on the top with its same-letter box on the bottom. However, all paths must remain inside the large box without crossing each other or going through any of the boxes with letters in them. Give appropriate paths or convince me that it is not possible to connect the boxes in this way.