ETGG 4803 **Quiz #4**
Due 8 May 2020 **Name** _____

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?

b. What is the purpose of cross-validation?

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

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

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

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

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

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?

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

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

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

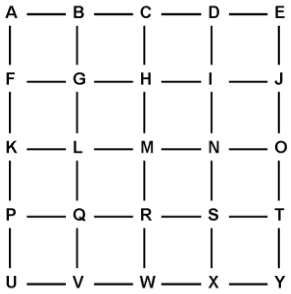
ii. [Density = Light, Grain = Small, Hardness = Soft] _____

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.

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?

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

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



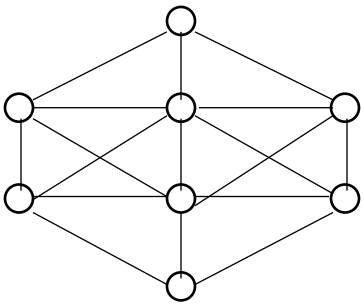
a. What are the five nearest neighbors of data point *M* using Euclidean distance? Break ties alphabetically.

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

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 _____
ii. $k = 3$ Class _____
iii. $k = 5$ Class _____
iv. $k = 7$ Class _____

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.

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.