CSCI 5582 Quiz 4

Name: _____

On my honor, as a University of Colorado at Boulder student, I have neither given nor received unauthorized assistance on this work. _____.

1. **(5 points) True or False:** In supervised machine learning, good performance on a training set guarantees similar performance on an unseen test set.

2. **Briefly describe** the operation of machine learning systems that are based on the following general approaches:
    a) **(5 Points) Splitting algorithms**

    b) **(5 Points) Covering algorithms**

3. **Decision trees**
    a) **(10 Points)** Given the training data in the table on the next page, which feature would the decision tree learning algorithm place at the top of the tree? Show the basis for your choice. (You can do this without computing the information gain metric).

    b) **(10 Points)** Give a complete decision tree that correctly accounts for the training data on the next page.

4. **Decision lists** (Reminder k-DL means that each test in the overall list can consist of at most k elements.)

    a) **(10 Points)** Construct a 2-DL list that correctly accounts for the data in the training set on the following page.

    b) **(5 Points)** If you were restricted to the space of 1-DL lists, what would be the optimal performance possible with our training set? Show a decision list that gives this performance.

5. **(5 Points)** A machine learning system has been built by taking a weak classifier approach and improving it with 5 rounds of boosting. Describe the steps involved in classifying a new instance using the resulting system.

Training Set

Training Instance	Label	F1 High/Low	F2 Red/Green/Blue	F3 In/Out	F4 Meat/Veg
1	Yes	High	Red	In	Meat
2	Yes	High	Green	Out	Veg
3	No	Low	Green	In	Veg
4	Yes	High	Red	In	Meat
5	No	Low	Green	Out	Meat
6	No	Low	Red	In	Veg
7	No	High	Red	Out	Meat
8	Yes	Low	Green	In	Meat
9	No	High	Red	In	Veg
10	Yes	Low	Blue	In	Meat