Homework 3

***Due date**: Sunday, Feb 25 11:00 pm*

Submit your solutions to Canvas. You may type your answers in a digital file, or upload a scanned/photographed copy of your handwritten homework

The textbook problems are from Russell and Norvig's AI book **3rd edition**

Decision Tree:

1. Problem 18.6 from textbook

2. Consider the following *Play Tennis* dataset (adapted from: Quinlan, "Induction of Decision Trees", Machine Learning, 1986) :

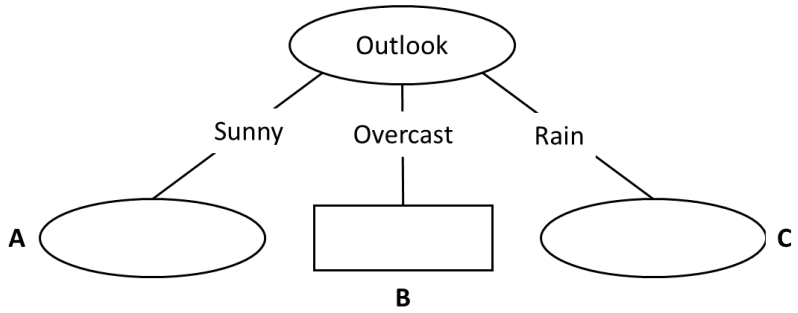
ATTRIBUTES:	POSSIBLE VALUES:	
Outlook	{Sunny, Rain, Overcast}	
Temperature	{Hot, Mild, Cool}	
Humidity	{High, Normal}	
Wind	{Strong, Weak}	
PlayTennis	{Yes, No}	<- classification target

ID	Outlook	Temperature	Humidity	Wind	PlayTennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Sunny	Hot	Normal	Weak	Yes
4	Sunny	Mild	Normal	Strong	Yes
5	Sunny	Mild	Normal	Weak	No
6	Rain	Mild	High	Strong	No
7	Rain	Mild	Normal	Weak	Yes
8	Rain	Cool	Normal	Weak	Yes
9	Rain	Mild	Normal	Strong	Yes
10	Rain	Cool	Normal	Strong	No
11	Overcast	Hot	High	Weak	Yes
12	Overcast	Cool	High	Strong	Yes
13	Overcast	Mild	High	Strong	Yes
14	Overcast	Hot	Normal	Weak	Yes

The machine learning task is to predict whether to play tennis or not, based on the data about the weather. Answer the questions in the following page.

a) Which of the major machine learning categories (supervised, unsupervised, or reinforcement) does this problem fall under? Explain.

Suppose we are in the middle of inducing the decision tree. The current state of the decision tree is given below:



b) What should the tree output in the leaf box labelled B?
c) Which data instances should be considered in node A? Write down the relevant instance ID numbers from the table.
d) Calculate the **entropy** at node A for the **Humidity** attribute. Show all the steps of your calculation. For your convenience, the logarithm in base 2 of selected values are provided.

x	1/2	1/3	2/3	1/4	3/4	1
$\log_2(x)$	-1	-1.6	-0.6	-2	-0.4	0

e) We have already calculated the entropies of Temperature and Wind at node A, and they are both 0.96. Based on these values, and based on the result you obtained previously, which attribute should be used to split node A? If needed, break any ties arbitrarily.
f) Write your chosen attribute in the blank space of node A in the tree given in the previous page. Now extend the appropriate number of branches from node A and label them properly. For each branch that leads to a leaf, draw the leaf in the tree on the previous page and mark it with the output that the leaf should produce. Show your work on the tree on the previous page.

Naïve Bayes:

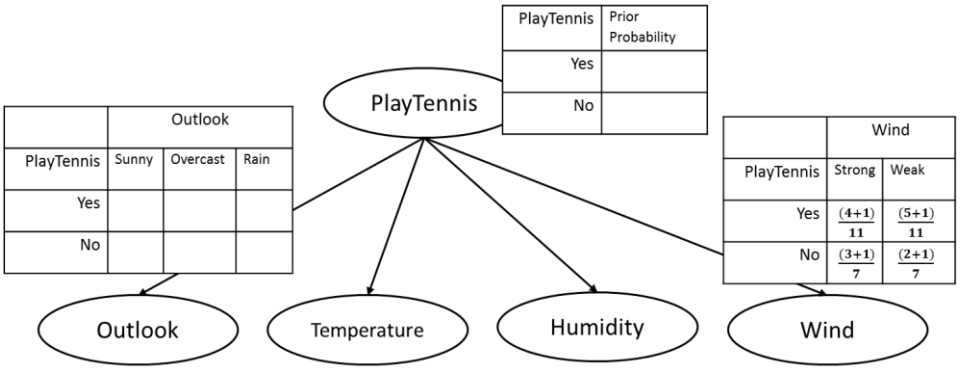
3. Use the Naive Bayes network in the class handout to predict the Risk of the following data instance. Show your work.

   Credit History = bad; Debt = high; Collateral = adequate; and Income = >35.

4. Consider the same *PlayTennis* dataset of Problem 2, which is reproduced here for your convenience (though the data instances appear in a different order):

New ID	Outlook	Temperature	Humidity	Wind	PlayTennis
1	Sunny	Hot	Normal	Weak	Yes
2	Sunny	Mild	Normal	Strong	Yes
3	Rain	Mild	Normal	Weak	Yes
4	Rain	Cool	Normal	Weak	Yes
5	Rain	Mild	Normal	Strong	Yes
6	Overcast	Hot	High	Weak	Yes
7	Overcast	Cool	High	Strong	Yes
8	Overcast	Mild	High	Strong	Yes
9	Overcast	Hot	Normal	Weak	Yes
10	Sunny	Hot	High	Weak	No
11	Sunny	Hot	High	Strong	No
12	Sunny	Mild	Normal	Weak	No
13	Rain	Mild	High	Strong	No
14	Rain	Cool	Normal	Strong	No

a) From the dataset above, the following Naïve Bayes model is built. Your job is to **fill out the likelihood tables** given below. Some tables have already been filled for your convenience. Remember that 1 is added to all the counts to avoid the problem of having a probability that is equal to 0 (Laplace smoothing).



	Temperature		
PlayTennis	Hot	Mild	Cool
Yes	$\frac{(3+1)}{12}$	$\frac{(4+1)}{12}$	$\frac{(2+1)}{12}$
No	$\frac{(2+1)}{8}$	$\frac{(2+1)}{8}$	$\frac{(1+1)}{8}$

	Humidity	
PlayTennis	High	Normal
Yes		
No		

b) Use the Naïve Bayes model constructed in the previous page to classify the following new instance. That is, determine which of the two values of PlayTennis (Yes or No) has the highest probability given the outlook, temperature, humidity and wind values of the given instance. **Show all your work and explain your answer.**

Outlook	Temperature	Humidity	Wind	PlayTennis
Sunny	Hot	Normal	Strong	?

Artificial Neural Networks

5. Investigate and explain the differences among the following types of "units" (i.e., perceptrons) in terms of the activation functions that they use. In addition, provide a graphical depiction of these activation functions.

    - Step
    - Sigmoid
    - ReLU
    - Tanh
    - Softmax

6. Problem 18.19 from textbook

7. Problem 18.22 (a, c) from textbook [Here h = No. of nodes in hidden layer]

Evaluating Machine Learning Models

8. Assume that the following confusion matrix was obtained from using a machine learning model to predict the classifcation of 24 test instances, where the target attribute is Risk, with possible values low, moderate, and high:

```
a  b   c   <-- classified as
4  0   1 |  a = low
0  1   3 |  b = moderate
1  2  12 |  c = high
```

    For each of the following metrics used in project 3, provide a formula that defines the metric, and calculate its value based on the confusion matrix above. Show your work.

    a) The model's classification accuracy.
    b) The model's classification error (= 100% - model's classification accuracy).
    c) The model's precision for class=low.
    d) The model's recall for class=moderate.