COMP 7745/8745: Midterm Exam

March 2 , 2017

Allocated Time: 1 hour and 25 minutes

NAME:----------

- **Open Books and Notes**
- **Please write your answers directly below the questions in the space provided. If you run out of space, you can use the back of the exam.**
- **State any assumptions if you need to make them.**
- **Please try to attempt all questions**
- **Best of luck!**

Maximum Points: 100

Number of Pages: 12

QUESTION	POINTS
Short Questions (True/False with justification)	39
Naïve Bayes	16
Decision Trees	12
Perceptrons	10
Model Evaluation	12
General Question	11

1) State whether true or false with a brief justification (39 points) (No points will be given for random guessing without justification)
    a. Given a dataset with no noisy examples (i.e., it is never the case that for 2 examples, the attribute values match but the class value does not), the training error for the ID3 algorithm is always equal to 0.

    b. Given a linearly separable dataset, the leave-one-out cross-validation error for logistic regression is always equal to 0.

    c. Given a dataset with no noisy examples, Naïve Bayes is guaranteed to have a training error equal to 0.

d. Given any dataset, the leave-one-out cross validation error of 1-nearest neighbor is always equal to 0.

e. As we increase the number of hidden layers in the neural network, the leave-one-out cross validation error is guaranteed to decrease.

f. Given a dataset that is not linearly separable, a Neural Network trained on this dataset has lower bias as compared to a perceptron trained on the same dataset.

g. $(\neg X \vee Y) \wedge (\neg Y \vee X)$ can be represented by a perceptron

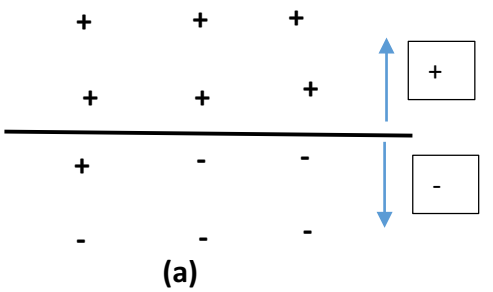
h. During the testing phase (making predictions given a trained classifier), a K-Nearest Neighbor classifier is typically far more efficient in terms of computation as compared to a Logistic Regression based classifier

i. The bias of a classifier can be reduced as follows: we average several classifiers obtained by training on random samples taken from the dataset

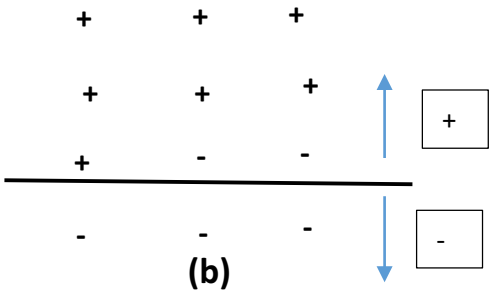
j. A Naïve Bayes classifier is more similar to a Max-likelihood estimator as compared to a Bayesian MAP estimator

k. A neural network with arbitrary number of hidden layers can represent every function that a decision tree can represent

l. Consider Adaboosting using only horizontal decision stumps. Assuming that the first decision stump is shown in Figure 1 (a), the decision stump in the next iteration is quite likely to be the one shown in Figure 1(b)



(a)



(b)

Figure 1

m. The training error of AdaBoost goes to zero, as we increase the number of iterations

2) Naïve Bayes Classification: Following is a dataset on burglaries

Alarm	Dog	Large-House	Burglary
0	1	0	0
1	0	0	0
0	0	0	1
1	1	1	0
0	1	1	1

a) Predict the value of Burglary given an instance (Alarm=0; Dog=0; Large-House = 1) (10 points)

b) The county records office told you that there are 1000 more records, but since part of the records were not available, they only told you that of the 1000 records, 800 of them were records with burglary = 1 and 200 of them were records with burglary = 0. Using this new information, what is your new prediction for (Alarm=0; Dog=0; Large-House = 1)? (6 points)

3) Decision Trees

a) Consider the following dataset,

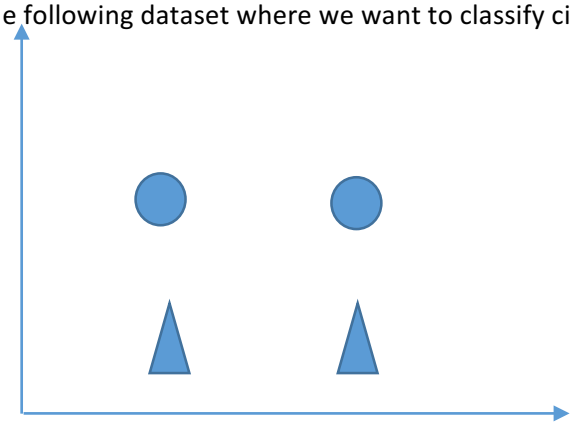
A1	A2	A3	Class
1	1	0	0
1	1	1	0
1	0	0	0
0	0	1	0
0	0	0	1
0	0	1	1
0	1	0	1
1	1	1	1

Draw the decision tree learned by ID3 for the above data? (12 points) (hint: you don't need a calculator for this)

4) Perceptrons

a)

- Consider the following dataset where we want to classify circles from triangles.



Assume in all the following cases that we have an arbitrarily small learning rate, and we are allowed to make infinitely many iterations through the examples (10 points)

i) Suppose we wish to learn a perceptron with 2 weights and no bias, using the gradient descent weight update rule for, will we converge such that we get 0 training error? Briefly explain.

ii) Suppose we wish to learn a perceptron with 2 weights and a bias, using the gradient descent weight update rule, will we converge such that we get 0 training error? Briefly explain.

5) Model Evaluation

a) Let us consider a naïve classifier which simply functions like a lookup table. That is, given a test example to classify, if the test example is present in the training dataset the classifier outputs its label as the one in the training dataset. If the test example is absent in the training dataset, the classifier outputs its label as -ve. (12 points)
    a. Given a training dataset with no duplicate entries and 50% of the dataset is of the +ve class and 50% is of the –ve class, what is the average leave-one-out cross validation error.

b. Given the following training and testing datasets, what is the precision and recall of our classifier in the test data w.r.t to the +ve class.

Feature-1	Feature-2	Class
X1	Y1	-
X2	Y2	-
X3	Y3	+
X4	Y4	-
X5	Y5	+
X6	Y6	+

Training data

Feature-1	Feature-2	Class
X1	Y1	+
X2	Y2	+
X3	Y3	+
X10	Y10	+
X11	Y11	-

Testing data

6) You get hired as a Machine learning expert at movies.com. Your task on day one is to design a classifier for predicting if a movie is a success/failure at the box office. You decide to build a logistic regression classifier as follows. You divide the data into 2 parts, train the classifier on one part and test on the other. You see that the error is quite high, then you suddenly remember that there is a learning rate parameter that you studied about in class and decide to tweak this parameter. You try several different learning rates, and realize that there is a learning rate for which the prediction accuracy on the test set is maximized. You report to your boss the prediction accuracy using the best value of the learning rate parameter.
Should your boss reward or not for this effort? If your answer is reward you, briefly explain why? If your answer is not to reward you, then come up with an alternate strategy to impress your boss! (12 points)