CS188: Artificial Intelligence, Spring 2009
Written Assignment 4: Classification

Due April 30 at the beginning of lecture
You can work on this in groups, but everyone should turn in his/her own work.
Don't forget your **name, login, and GSI**.

Name: **Login:** **GSI:**

1 Question 1: Course Gossip (9 points)

In this question, you will train classifiers to predict whether sentences are about CS 188 or CS 186 (databases) using a bag-of-words Naive Bayes classifier. Each sentence is labeled with the class to which it pertains.

Training set	**Held-out set**	**Test set**
(188) agents need good models. (188) agents need data. (186) buffers need memory. (186) DBs need data models.	(188) agents need memory. (186) DBs have data.	(186) data have data models.

a) Write down all of the maximum likelihood (relative frequency) parameters for a bag-of-words naive Bayes classifier trained on the *training set* above. Let Y be the class for a sentence, and W be a word. You may omit any parameters equal to 0. Ignore punctuation. *Note: Bag-of-words classifiers assume that the words at every sentence position are identically distributed. Repeated words affect both the likelihood of a word during estimation and sentence probabilities during inference.*

b) According to your classifier, what is the probability that the first held-out sentence "agents need memory" is about 188?

c) Using Laplace (i.e., *add one*) smoothing for all of your parameters, what is the probability of seeing the test sentence "data have data models": $P(W_1 = data, W_2 = have, W_3 = data, W_4 = models)$? Assume that the only words you ever expect to see are those in your training and held-out sets. *Hint: sum over Y, using the new estimates after smoothing.*

d) Using Laplace smoothing, what is the probability according to your classifier that the test sentence "data have data models" is about 186?

e) Suggest an additional feature that would allow the classifier to correctly conclude that "data have data models" is about 186 when trained on this training set.

2 Question 2: Red Light, Green Light (7 points)

You meet the chief administrative officer for intersection oversight for the California department of transportation. Her job is to report whether the stoplights at intersections are working correctly. You remark, "I bet I could automate your job." She scoffs in disbelief, but humors you by showing you some data.

X_1	X_2	Y
red (-1)	*red* (-1)	*broken* (+1)
red (-1)	*green* (+1)	*working* (-1)
green (+1)	*red* (-1)	*working* (-1)
green (+1)	*green* (+1)	*broken* (+1)

a) The data above have been annotated with feature and output values. Y is the label variable to predict. Circle all of the following feature sets make the training data linearly separable.

(i) X_1, X_2 (ii) X_1 only (iii) $\min(X_1, X_2)$, X_1, X_2 (iv) $\min(X_1, X_2)$, X_2 (v) $|X_1 + X_2|$

b) Using just X_1 and X_2 as features, fill in the resulting weight vectors after two-class perceptron updates (one vector of weights) for the first and second data points sequentially, starting with the initial vector below.

	X_1	X_2
Initial Weights	2	1
Weights after observing $(X_1 = -1, X_2 = -1, Y = 1)$		
Weights after observing $(X_1 = -1, X_2 = 1, Y = -1)$		

b) Using just X_1 and X_2 as features, fill in the resulting weight vectors after two-class MIRA updates for the first and second data points sequentially, starting with the initial vector below. Assume the maximum update size C is 5.

	X_1	X_2
Initial Weights	1	2
Weights after observing $(X_1 = -1, X_2 = -1, Y = 1)$		
Weights after observing $(X_1 = -1, X_2 = 1, Y = -1)$		

c) Is it possible in any classification problem for a perceptron classifier to reach perfect training set accuracy in fewer updates than a MIRA classifier, if they both start with the same initial weights and examine the training data in the same order? Briefly justify your answer.

3 Question 3: One-Sentence Answers (4 points)

a) What is one advantage of Naive Bayes over MIRA?

b) What is one advantage of MIRA over Naive Bayes?

c) What is one advantage of a nearest-neighbor classifier over Naive Bayes?

d) Invent a method of combining a Naive Bayes classifier with a MIRA classifier and describe it.