The Ohio State University
Computer Science and Engineering

CSE 3521– Final Exam
Survey of Artificial Intelligence I: Basic Techniques

Instructor: Wuwei Lan

2019/04/19

Student Name: _____

LastName.Number: _____

This exam contains 6 pages (including this cover page) and 4 questions. Total of points is 100. Good luck and Happy reading work!

Distribution of Marks

Question	Points	Score
1	60	
2	10	
3	10	
4	20	
Total:	100	

1. This part will test your basic knowledge about machine learning, just give short and clear answer for each question.

   (a) (10 points) Please list three classifiers that can be used for text classification.

   (b) (10 points) Compared to Perceptron, what is the benefit of Logistic Regression?

   (c) (10 points) Please explain kernel trick with your own words.

   (d) (10 points) If we want to build very deep Convolutional Neural Network (CNN), e.g. 150 layers, what technique is critical to make this model work?

   (e) (10 points) What will happen if we have vanishing gradients in Recurrent Neural Network (RNN)?

   (f) (10 points) If we have balanced dataset, can we use accuracy, instead of precision/recall/F1, for evaluation?

2. Perceptron

You have decided to become a teacher. The only issue is that you don't want to spend lots of time grading essays, so instead you decide to grade them all with a linear classifier. Your classifier considers the number of 7-letter (f_7) and 8-letter words (f_8) in an essay and then assigns a grade, either A or F, based on those two numbers. You have four graded essays to learn from:

$BIAS$	f_7	f_8	grade
1	2	1	A (+)
1	0	2	F (-)
1	1	2	A (+)
1	1	0	F (-)

(a) (3 points) You decide to run perceptron and being optimistic about the students essay writing capabilities, you decide to initialize your weight vector as $(1, 0, 0)$. If the score from your classifier is greater than 0, it gives an A, if it is 0 or lower, it gives an F. Fill in the resulting weight vector after having seen the first training example and after having seen the second training example.

	BIAS	f_7	f_8
Initial	1	0	0
After first training example			
After second training example			

(b) (3 points) Please select and justify your answer.

◯True ◯False The training data is linearly separable with the given features.

Justify:

(c) (4 points) For each of the following decision rules, indicate whether there is a weight vector that represents the decision rule. If "Yes" then include such a weight vector.

1. A paper gets an A if and only if it satisfies $(f_7 + f_8 \geq 7)$.

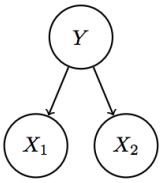
◯Yes $w =$ ◯No

2. A paper gets an A if and only if it satisfies $(f_7 \geq 5 \; AND \; f_8 \geq 4)$.

◯Yes $w =$ ◯No

3. Naive Bayes

You are given a naive bayes model, shown below, with label Y and features X_1 and X_2. The conditional probabilities for the model are parametrized by $p1$, $p2$ and q.



X_1	Y	$P(X_1\|Y)$
0	0	p_1
1	0	$1-p_1$
0	1	$1-p_1$
1	1	p_1

X_2	Y	$P(X_2\|Y)$
0	0	p_2
1	0	$1-p_2$
0	1	$1-p_2$
1	1	p_2

Y	$P(Y)$
0	$1-q$
1	q

Note that some of the parameters are shared (e.g. $P(X_1 = 0|Y = 0) = P(X_1 = 1|Y = 1) = p_1$).

(a) (5 points) Given a new data point with $X1 = 1$ and $X2 = 1$, what is the probability that this point has label Y = 1? Express your answer in terms of the parameters $p1$, $p2$ and q (you might not need all of them).

$$P(Y = 1|X_1 = 1, X_2 = 1) = \_\_\_\_\_$$

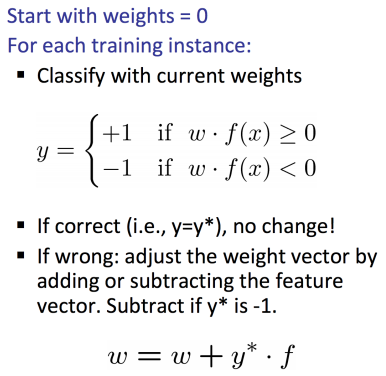
(b) (5 points) If the model is trained with the following data, what are the maximum likelihood estimates for $p1$, $p2$ and q?

sample number	1	2	3	4	5	6	7	8	9	10
X_1	0	0	1	0	1	0	1	0	1	1
X_2	0	0	0	0	0	0	0	1	0	0
Y	0	0	0	0	0	0	0	1	1	1

$p_1 =$ _____ $p_2 =$ _____ $q =$ _____

4. Python Programming

   Write down Python code for the following Binary Perceptron (You can use NumPy array for vector representation).



   (a) (10 points) Python code for training Perceptron.

   (b) (10 points) Python code for testing Perceptron.

This page is intentionally left blank to accommodate work that wouldn't fit elsewhere and/or scratch work.