First Name: _____ Last Name: _____ Student ID: _____

BBM 406 – Introduction to Machine Learning (Fall 2016)

Midterm Exam

November 14th, 2016

Instructions

- Do not open this exam booklet until you are directed to do so. Read all the instructions first.
- When the exam begins, write your name on every page of this exam booklet.
- The exam contains **five multi-part problems**. You have **100 minutes** to earn 100 points.
- The exam booklet contains **6 pages** to the exam, including this one.
- This exam is a **closed book and notes exam**. But you are allowed to have an A4 sized *copy sheet* during this exam (you can use both sides of this sheet).
- Please write your answers in the space provided on the exam paper.
- Show all work, as partial credit will be given. You will be graded not only on the correctness and efficiency of your answers, but also on your clarity that you express it. Be neat.
- Good luck!

Problem	Points	Score
1	20	
2	20	
3	20	
4	20	
5	20	
Total	100	

It is a violation of the Academic Integrity Code to look at any exam paper other than your own, any other reference material (books, lecture notes), or to give inappropriate help to someone or to receive unauthorized aid by someone. Please also not discuss this exam with the students who are scheduled to take a makeup exam.

Academic Integrity is expected of all students of Hacettepe University at all times, whether in the presence or absence of members of the faculty. Do NOT sign nor take this exam if you do not agree with the honor code.

Understanding this, I declare I shall not give, use or receive unauthorized aid in this examination.

Signature: _____

First Name: _____ Last Name: _____ Student ID: _____

Q1.[20 points] **Short Questions**

(a) [5 points] Describe the *k-fold cross-validation* algorithm and how it is used for model selection.

(b) [5 points] What is the main assumption in the *Naïve Bayes* method? Describe when this assumption could be useful, and when can it lead to poor results?

(c) [5 points] As an intern in a software company, you want to classify applications to detect buggy applications using features of the source code. But this project only has a few applications to be used as training data. To create the most accurate classifier, do you recommend using a *discriminative* or *generative classifier*? Explain your reasoning.

(d) [5 points] Which datasets given below are *linearly separable*? Explain your reasoning.

Dataset A

x_1	x_2	y
1	1	+
4	2	–
4	5	–
5	5	+

Dataset B

x_1	x_2	y
1	1	+
5	5	–
4	5	–
5	5	+

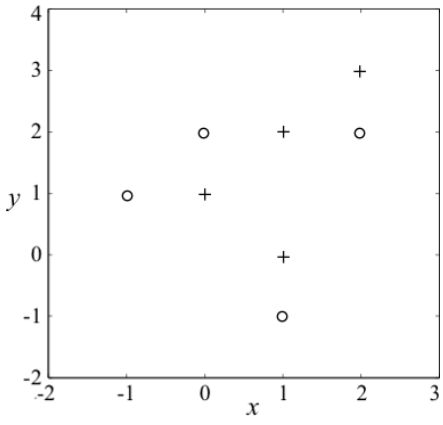
Dataset C

x_1	x_2	y
1	1	+
4	2	–
4	5	+
5	5	+

First Name: _____ Last Name: _____ Student ID: _____

Q2.[20 points] **k-Nearest-Neighbor Classification**

Consider the following dataset with two real-valued input and one binary output (+ or ○). The following questions assume that we are using k-nearest-neighbor classification with unweighted Euclidean distance to predict y for an input x.



(a) [4 points] What would be the classification of a test sample $x = (1,1)$ according to 3-NN?

(b) [4 points] What would be the classification of a test sample $x = (1,1)$ according to 5-NN?

(c) [4 points] What would be the classification of a test sample $x = (1,1)$ according to 7-NN?

(d) [8 points] What is the leave-one-out cross-validation error of 1-NN for this dataset. In your answer, please also state the classification predictions for each data point.

First Name: _____ Last Name: _____ Student ID: _____

Q3. [20 points] **Naïve Bayes Classifiers**

Suppose we are given the following dataset, where x_1, x_2, x_3 are input binary random variables, and y is a binary output whose value we want to predict.

x_1	x_2	x_3	y
0	1	1	0
1	1	1	0
0	0	0	0
1	1	0	1
0	1	0	1
1	0	1	1

(a) [8 points] Compute all the probabilities necessary for a Naïve Bayes classifier, i.e., the class probability $p(y)$ and all the individual feature probabilities $p(x_i|y)$, for each class y and feature x_i. Assume no Laplace smoothing!

(b) [4 points] Which class would be predicted for $x = (1\ 0\ 0)$? Show all your work.

(c) [4 points] What is the potential problem with using a Naïve Bayes Classifier that was learned without Laplace smoothing?

(d) [4 points] Suppose now you are told that $y = x_2$ XOR x_3. Do you think that Naïve Bayes classifier is appropriate for this problem? Why?

First Name: _____ Last Name: _____ Student ID: _____

Q4.[16 points] **Perceptron**

Consider the following supervised learning problem in which the training examples are points in 7-dimensional space.

Instance	Data	Label y
x_1	(0, 0, 0, 1, 0, 0, 1)	1
x_2	(1, 1, 0, 0, 0, 1, 0)	-1
x_3	(0, 0, 1, 1, 0, 0, 0)	1
x_4	(1, 0, 0, 0, 1, 1, 0)	-1
x_5	(1, 0, 0, 0, 0, 1, 0)	-1

Show how the perceptron algorithm without bias (*i.e.* we set the bias $b = 0$ and keep it fixed throughout the iterations) for the above sequence of samples. Assume that the initial set of weights is $w = (0, 0, 0, 0, 0, 0, 0)$.

First Name: _____ Last Name: _____ Student ID: _____

Q5. [20 points] **Neural Networks**

Draw a neural network that represents the function $f(x_1, x_2, x_3)$ defined below. In your answer, you are supposed to use only *linear units* and *sign units*. Recall that, while the linear unit takes as input weights and feature values and outputs $w_0 + \sum_i w_i x_i$, the sign unit outputs +1 if $(w_0 + \sum_i w_i x_i) > 0$ and -1 otherwise.

x_1	x_2	x_3	$f(x_1, x_2, x_3)$
0	0	0	10
0	0	1	-5
0	1	0	-5
0	1	1	10
1	0	0	-5
1	0	1	10
1	1	0	10
1	1	1	10

Note: To get full credit, you have to write down the specific numeric weights (*e.g.*, 1, 0.5, +1, etc.) as well as the specific units used at each hidden and output node.