University of Wisconsin Madison
Department of Computer Sciences

CS 760 Machine Learning
Spring 2017

Final Examination

Duration: *1 hour 15 minutes*

One set of handwritten notes and calculator allowed.

Instructions:
· Write your answers in the space provided.
· Show your calculations LEGIBLY.
· If you feel that a question is not fully specified, state any assumptions you need to make in order to solve the problem.
· Use the backs of the sheets for scratch work ONLY.
· Write all the final answers BELOW the questions. Answers written in the scratch sheets will NOT be considered.

Name:

UW ID:

Problem	Score	Max Score
1		20
2		30
3		20
4		30
Total		**100**

Problem 1: Decision trees and instance based learning   **20 points**

1. Which of the following statements are true for BOTH decision trees and Naive Bayes classifiers (you may choose more than one statement)? Explain   **(4 points)**
    a) In both classifiers a pair of features are assumed to be independent
    b) In both classifiers a pair of features are assumed to be dependent
    c) In both classifiers a pair of features are assumed to be independent given the class label
    d) In both classifiers a pair of features are assumed to be dependent given the class label

2. Consider the following training set in 2 dimensional Euclidean space: **(6 points)**

x	y	Class
-1	1	-
0	1	+
0	2	-
1	-1	-
1	0	+
1	2	+
2	2	-
2	3	+

a)   What is the prediction of a 3 nearest neighbor classifier at the point (1,1)?

b) What is the prediction of a 5 nearest neighbor classifier at the point (1,1)?

c) What is the prediction of a 7 nearest neighbor classifier at the point (1,1)?

3. What is the biggest advantage/disadvantage of decision trees when compared to logistic regression classifiers?   **(5 points)**

4. Show a decision tree that would perfectly classify the following data set: **(5 points)**

	A	**B**	**Class**
Instance 1	2	3	+
Instance 2	4	4	+
Instance 3	4	5	-
Instance 4	6	3	+
Instance 5	8	3	-

| Instance 6 | 8 | 4 | - |

Problem 2 - Neural Networks: **30 points**

a) State whether the following statements are true or false and explain why. **(12 points)**

    i)   A Perceptron can learn to correctly classify the following data, where each consists of three binary input values and a binary classification value: (111,1), (110,1), (011,1), (010,0), (000,0).

    ii)   The Perceptron Learning Rule is a sound and complete method for a Perceptron to learn to correctly classify any two-class problem.

    iii)   Training neural networks has the potential problem of overfitting the training data.

b)   Answer the following: **(12 points)**
        i)   What is the search space and what is the search method used by the backpropagation algorithm for training neural networks?

ii) What quantity does backpropagation minimize?

iii) Does the back-propagation algorithm, when run until a minimum is achieved, always find the same solution no matter what the initial set of weights are? Briefly explain why or why not.

c) Demonstrate how the perceptron without bias (i.e. we set the parameter $b = 0$ and keep it fixed) updates its parameters given the following training sequence:

$x_1 = (0,0,0,1,0,0,1)$ $y_1 = 1$
$x_2 = (1,1,0,0,0,1,0)$ $y_2 = -1$
$x_3 = (0,0,1,1,0,0,0)$ $y_3 = 1$
$x_4 = (1,0,0,0,1,1,0)$ $y_4 = -1$
$x_5 = (1,0,0,0,0,1,0)$ $y_5 = -1$

Assume initial weights to be 0 and learning rate to be 1.0. **(6 points)**

Problem 3   **20 points**

Briefly describe the following:

i) Pruning a decision tree

ii) Auto encoders

iii) Bagging

iv) Regularization

v) Markov Blanket

vi) Occam's razor

Problem 4 - Support Vector Machine   **20 points**

1. What are the advantages/disadvantages of a non-linear SVM? Give examples to justify your reasoning.

2. What is a kernel function? Why do we need it?

3. Given the following data samples (square and triangle mean two classes), which one(s) of the following kernels can we use in SVM to separate the two classes?
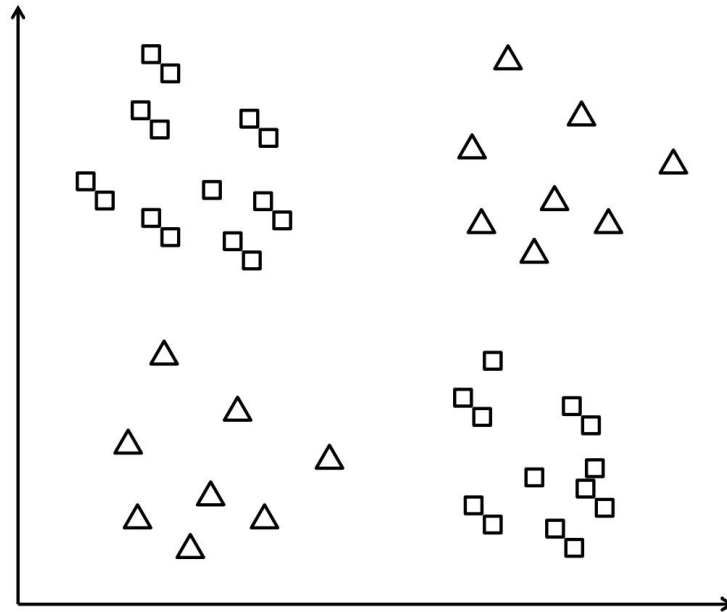
a) Linear kernel
b) Polynomial kernel
c) Gaussian RBF (radial basis function) kernel
d) None of the above

4. How does the margin ρ relate to the weight vector w? Express the relation using a formula.