CS 559: Homework Set 4
Due: April 23, 6:15pm

Philippos Mordohai
Department of Computer Science
Stevens Institute of Technology
Philippos.Mordohai@stevens.edu

Collaboration Policy. Homeworks will be done individually: each student must hand in their own answers. It is acceptable for students to collaborate in understanding the material but not in solving the problems. Use of the Internet is allowed, but should not include searching for previous solutions or answers to the specific questions of the assignment. We will assume that, as participants in a graduate course, you will be taking the responsibility of making sure that you personally understand the solution to any work arising from collaboration.

Late Policy. The penalty for late submission is 20% of the grade per day, enforced at 6:15 each day after the due date. If urgent or unusual circumstances prohibit you from submitting a homework assignment in time, please send me an e-mail explaining the situation.

Problem 1. (5 points) What is the effect of minimizing the norm of the weight vector w in Support Vector Machines? Why is this a good objective function?

Problem 2. (15 points) In a two-class dataset in 1D, class 1 consists of {-1, 0, 1} and class 2 of {-3, -2, 2}.

(a) Find a mapping from 1D to 2D that makes this dataset linearly separable. Show the selected mapping function and plot the points in the 2D space.

(b) In you plot, draw the decision boundary according to a linear discriminant function computed according to the SVM criteria. No need to compute the decision boundary, just draw it. Which points are the support vectors?

(c) For this mapping, what is the corresponding kernel $K(x_i, x_j)$?

Problem 3. (25 points) Consider the following data drawn from two distributions in 2D.
Class 1: $D_1 = \{[2\ \ 7], [5\ \ 2], [4\ \ 4]\}$
Class 2: $D_2 = \{[1\ \ 4], [4\ \ 1]\}$
Train a strong classifier using AdaBoost and the following two hypothesis. Show all steps, especially the weight distribution of the samples at the beginning and the end of each stage, the error rate (see Notes 10, slide 23) and the weight of the weak learner. (Slide 26 in Notes 10 may be useful for computing the normalization constant at each iteration.)

The hypothesis are:

- h_1

$$h_1 : \begin{cases} x_1 > 3 & \text{assign to } \omega_1 \\ x_1 < 3 & \text{assign to } \omega_2 \end{cases}$$

- h_2

$$h_2 : \begin{cases} x_2 > 3 & \text{assign to } \omega_1 \\ x_2 < 3 & \text{assign to } \omega_2 \end{cases}$$

Finally, show the learned classifier that combines these two hypotheses. What is the decision rule?

Problem 4. (5 points) Write the joint probability distribution that corresponds to the following Bayes network. Make sure all symbols are properly defined.