This homework is individual work. All assignments are due at the beginning of class on the due date. Assignments turned in late will drop 10 points for each period of 24 hours for which the assignment is late. In addition, no assignments will be accepted after the solutions have been made available. **Please include your cornell net id on your homework**.

Part 1. Theory

A. Linear Classifier Margin (20 POINTS)

Given n orthogonal vectors of the form $(0, ..., 0, 1, 0, ..., 0)$, show that for any labelling $\{+1, -1\}$ of the vectors, there always exists a linear classification rule with margin at least $\frac{1}{\sqrt{n}}$.

B. SVM Hard-Margin Classification (20 POINTS)

Show that one can replace the "1" on the right hand side of the constraint in hard-margin SVMs (repeated below) with any positive number $\gamma > 0$ without changing the hyperplane that is computed.

$$\min_{\vec{w},\, b} \frac{1}{2}||\vec{w}||^2$$
$$\text{subject to } y_i(\vec{x}_i \cdot \vec{w} + b) \geq 1 \text{ for all } i = 1 \ldots n,$$

C. SVM Soft-Margin Classification (20 POINTS)

Imagine you want to train the following soft-margin SVM, where the slack variables enter the objective function squared.

$$\min_{\vec{w},\, b, \vec{\xi}} \frac{1}{2}||\vec{w}||^2 + C \sum_{i=1}^{n} \xi_i^2$$
$$\text{subject to } y_i(\vec{x}_i \cdot \vec{w} + b) \geq 1 - \xi_i \text{ for all } i = 1 \ldots n$$

However, you do not have quadratic programming software that can solve this optimization problem. All you have is a program that can solve hard-margin problems of the form.

$$\min_{\vec{w},\, b} \frac{1}{2}||\vec{w}||^2$$
$$\text{subject to } y_i(\vec{x}_i \cdot \vec{w} + b) \geq 1 \text{ for all } i = 1 \ldots n,$$

Show how you can transform the soft-margin problem into an equivalent problem that can be trained with the hard-margin software. Hint: Extend the feature vectors $\vec{x}_i$ by adding features in the appropriate way.

Part 2. Algorithm Comparison

A. Perceptron (20 POINTS)

Use the perceptron algorithm on the following two datasets:

earn: This is a dataset of documents. The task is to classify the pages into earning reports and other documents.

adult: The goal is to predict whether the income of a person exceeds \$50K/yr based on census data.

You can find an implementation of the (batch) perceptron on the course web page, as well as these data files and instructions for how to use the perceptron. You should plot the training error as well as the testing error versus the number of iterations of the algorithm for both datasets.

Discuss your results.

B. Soft margin SVM (20 POINTS)

Compare the Perceptron results to those of a soft-margin SVM on the same datasets. Plot the training error as well as the testing error versus the value of C. Make sure that you cover a reasonable range of C.

Discuss your results.

You can find an implementation of an SVM (SVM-light) on the course web page, as well as some instructions. Be aware that SVM-light might take a longer time to converge for large values of C.