ORF 363/COS 323	Problem set 4	Fall 2025, Princeton University
Instructor: A.A. Ahmadi		
TAs: Budway, Hua, Shi, Tang, Yang		
Due on Tuesday, October 28, 2025, at 1:30pm EST, on Gradescope		

For all problems that involve coding, please include your code.

Problem 1: Theory-applications split in a course. (Courtesy of Stephen Boyd)
A professor teaches a course with 24 lectures, labeled $i = 1, \ldots, 24$. The course involves some interesting theoretical topics, and many practical applications of the theory. The professor must decide how to split each lecture between theory and applications. Let T_i and A_i denote the fraction of the ith lecture devoted to theory and applications, for $i = 1, \ldots, 24$. (We have $T_i \geq 0$, $A_i \geq 0$, and $T_i + A_i = 1$.)
A certain amount of theory has to be covered before the applications can be taught. We model this in a crude way as

$$A_1 + \cdots + A_i \leq \phi(T_1 + \cdots + T_i), \quad i = 1, \ldots, 24, \tag{1}$$

where $\phi : \mathbb{R} \to \mathbb{R}$ is a given nondecreasing function. We interpret $\phi(u)$ as the cumulative amount of applications that can be covered, when the cumulative amount of theory covered is u. We will use the simple form $\phi(u) = a \max\{0, u - b\}$ with $a, b > 0$, which means that no applications can be covered until b lectures of the theory is covered; after that, each lecture of theory covered opens the possibility of covering a lectures on applications.
The theory-applications split affects the emotional state of students differently. We let s_i denote the emotional state of a student after lecture i, with $s_i = 0$ meaning neutral, $s_i > 0$ meaning happy, and $s_i < 0$ meaning unhappy. Careful studies have shown that s_i evolves via a linear recursion (dynamics)

$$s_i = (1 - \theta)s_{i-1} + \theta(\alpha T_i + \beta A_i), \quad i = 1, \ldots, 24,$$

with $s_0 = 0$. Here α and β are parameters (naturally interpreted as how much the student likes or dislikes theory and applications, respectively), and $\theta \in [0, 1]$ gives the emotional volatility of the student (i.e., how quickly he or she reacts to the content of recent lectures). The student's *cumulative emotional state* (CES) is by definition $s_1 + \cdots + s_{24}$. This is a measure of his/her overall happiness throughout the semester.
Now consider a specific instance of the problem, with course material parameters $a = 2$, $b = 3$, and three groups of students, with emotional dynamics parameters given as follows:

	Group 1	Group 2	Group 3
θ	0.05	0.1	0.3
α	-0.1	0.8	-0.3
β	1.4	-0.3	0.7

Your job is to plan (four different) theory-applications splits that respectively maximize the CES of the first group, the CES of the second group, the CES of the third group, and, finally, the minimum of the cumulative emotional states of all three groups. (Hint: you would need to appropriately reformulate constraint (1) to end up with a convex optimization problem.) Report the numerical values of the CES for each group, for each of the four theory-applications splits (i.e., fill out the following table):

	Group 1	Group 2	Group 3
Plan 1			
Plan 2			
Plan 3			
Plan 4			

For each of the four plans, plot T_i as well as the emotional state s_i for all three groups, versus i. (So you should have four figures with four curves on each.) These plots show you how the emotional states of the students change as the amount of theory varies.

Problem 2: Support Vector Machines (SVMs)

Recall our Support Vector Machines application of convex optimization from lecture. We have m feature vectors $x_1, \ldots, x_m \in \mathbb{R}^n$ with each x_i having a label $y_i \in \{-1, 1\}$. The goal is to find a linear classifier, that is a hyperplane defined by an affine function $a^T x - b$, where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$, by solving the optimization problem

$$\begin{aligned} \min_{a,b} \quad & \|a\| \\ \text{s.t.} \quad & y_i(a^T x_i - b) \geq 1, \ \forall i = 1, \ldots, m. \end{aligned} \tag{2}$$

We will then use this classifier to classify new data points. Throughout the problem, we assume that there is at least one data point with $y_i = 1$ and one with $y_i = -1$ (otherwise there is nothing to classify).

1. Uniqueness of the optimal solution.
   (a) Is the objective function $\|a\|$ convex? Strictly convex?
   (b) What about $\|a\|^2$? Is it convex? Strictly convex?
   (c) Prove that the solution to (2) is unique.
2. We would like to show that the optimization problem (2) is equivalent to

$$\begin{aligned} \max_{a,b,t} \quad & t \\ \text{s.t.} \quad & y_i(a^T x_i - b) \geq t, \ \forall i = 1, \ldots, m \\ & \|a\| \leq 1, \end{aligned} \tag{3}$$

   which is easier to interpret in terms of finding a classifier with maximum margin. Show that if (2) is feasible (with a positive optimal value), then (3) is feasible (and has a positive optimal value). Conversely, show that if (3) is feasible (with a positive optimal value), then (2) is feasible (and has a positive optimal value).
3. Assume the optimal value of (3) is positive. Show that an optimal solution of (3) always satisfies $\|a\| = 1$.

Problem 3: SVMs with linearly separable data

Open the Matlab file `HWSVM.mat`. To do this, download the file into your working directory and open it by calling `"load HWSVM"` in Matlab. This will load 6 vectors into MATLAB. You will need three of these vectors (`"x1part2"`, `"x2part2"` and `"ypart2"`) for this part of the problem. These three vectors correspond to $m = 53$ points in $\mathbb{R}^2$ whose components $(x_1, x_2)_{i, i=1,\ldots,m}$ are given in the first two vectors and whose labels y_i are given in the vector `ypart2`. For Python users, you can use the following code to load the data file.

```
import scipy
mat = scipy.io.loadmat('HWSVM.mat')
x1 = mat['x1part2']
x2 = mat['x2part2']
y = mat['ypart2']
```

1. Plot all the 53 points on a graph. We need to be able to tell the difference between points that are labelled 1 and points that are labelled -1.
2. Solve optimization problem (2) and plot on the same graph the optimal linear classifier (hyperplane) and the two shifted hyperplanes corresponding to the boundaries of the margin. Give the equations of these three lines.
3. Which points are the support vectors? Give their coordinates.

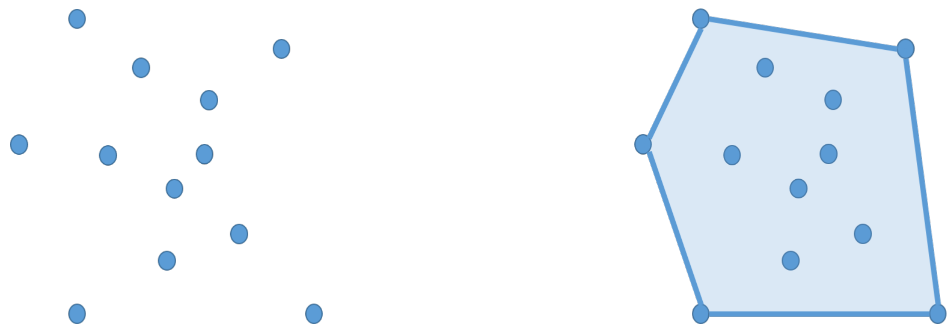
Problem 4: SVMs with data that is not linearly separable

You will now need the data vectors `"x1part3"`, `"x2part3"` and `"ypart3"` from `"HWSVM.mat"`. These three vectors correspond to $m = 100$ points $(x_1, x_2)_{i, i=1,\ldots,m}$ in $\mathbb{R}^2$ and an associated vector y which has the label of each point.

1. Let S be a set consisting of s points $z_1, \ldots, z_s$ in $\mathbb{R}^k$. The convex hull of S is defined as

$$\text{conv}(S) = \left\{ \sum_{i=1}^{s} \lambda_i z_i \mid z_i \in S, \lambda_i \geq 0, \text{ and } \sum_{i=1}^{s} \lambda_i = 1 \right\}.$$

   In words, this is the set of points that can be written as a convex combination of the points in S. A geometric interpretation of this definition is given in Figure 1.



(a) A set of points S (b) Convex hull of S

Figure 1: Convex hull of a set

Define

$$A = \{(x_1, x_2)_{i,i=1,\dots,m} | y_i = 1\}$$

and

$$B = \{(x_1, x_2)_{i,i=1,\dots,m} | y_i = -1\}.$$

We say that the sets A and B are linearly separable if there exists a hyperplane $a^T x - b$ that takes value ≥ 1 on A and ≤ -1 on B. Prove that if A and B are linearly separable, then their convex hulls do not intersect.

2. For the numerical data given, find a point that is both in $\text{conv}(A)$ and $\text{conv}(B)$ using CVX/CVXPY. Plot this point on the graph and give its coordinates.
   Hint: Write the problem as a convex optimization problem.

3. Recall the following convex optimization problem from lecture that attempts to simultaneously minimize the number of misclassified points and maximize the length of the margin:

$$\begin{aligned} \min_{a,b,\eta} \quad & \|a\| + \gamma \|\eta\|_1 \\ \text{s.t.} \quad & y_i(a^T x_i - b) \geq 1 - \eta_i, \ \forall i = 1, \dots, m \\ & \eta_i \geq 0, \ \forall i = 1, \dots, m. \end{aligned} \tag{4}$$

   Solve this problem for $\gamma = 1, 2, \dots, 10$ and generate two plots: The first one will give the length of the margin (counting both sides) as a function of γ; the second one will give the number of misclassified points as a function of γ. Discuss the overall trends of the two plots; are they what you were expecting?

Problem 5: Hillary or Bernie?

You would like to use the knowledge you've acquired in optimization over the past few weeks to see if you could have predicted the outcome of each Hillary-Bernie race in the Democratic primaries. To make things easier, you consider only the counties in the tri-state area and New England, i.e., those that belong to the states of New York, New Jersey, Maine, New Hampshire, Pennsylvania, Vermont, Massachussetts, Connecticut, or Rhode Island.

Your goal is to find a linear classifier that, for each county, labels it either as a Bernie win or as a Hillary win. To do this, you have access to a feature vector comprising the following features: mean income, percentage of hispanics, percentage of whites, percentage of residents with a Bachelor's degree or higher, and population density.

1. Load the data file `Hillary_vs_Bernie` in MATLAB/Python. This file includes 4 parts: `features_train`, `features_test`, `labels_train` and `labels_test`.

   In `features_train`, we have given you the feature vectors for 175 counties and in `label_train`, their corresponding labels (-1 is a Bernie win and 1 is a Hillary win). As there was a wide disparity in the orders of magnitude of the original data (average income is around 10^4 whereas the percentages are between 0 and 1), each feature vector has already been normalized by its standard deviation. The original data can be found at `https://www.kaggle.com/benhamner/2016-us-election` (as fact checking is popular at the moment :)). Solve problem (4) to build a linear classifier for this training set for $\gamma = 0.1, 1, 10$. For each value of γ, specify the optimal a^*, b^* obtained.

2. Test the performance of your classifier using the feature vectors from 21 other counties (given in `features_test`) by comparing the labels obtained to the ones given in `label_test`. Which γ gives you the highest success rate in terms of prediction? Take a look at the entries of a^* in this case – what does this suggest about the people who vote for Hillary compared to those who vote for Bernie?