Problem 1: Support Vector Machines (SVMs)

Part I: 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 \( a^T x - b \), where \( a \in \mathbb{R}^n \) and \( b \in \mathbb{R} \), by solving the optimization problem

\[
\begin{align*}
\min_{a,b} & \quad \|a\| \\
\text{s.t.} & \quad y_i(a^T x_i - b) \geq 1 \text{ for } i = 1, \ldots, m.
\end{align*}
\]  

We will then use this classifier to classify new data points.

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 (1) is unique.

2. We would like to show that the optimization problem (1) is equivalent to

\[
\begin{align*}
\max_{a,b,t} & \quad t \\
\text{s.t.} & \quad y_i(a^T x_i - b) \geq t \text{ for } i = 1, \ldots, m \\
& \quad \|a\| \leq 1,
\end{align*}
\]

which is easier to interpret in terms of finding a classifier with maximum margin.

Show that if (1) is feasible (with a positive optimal value), then (2) is feasible (and has a positive optimal value). Conversely, show that if (2) is feasible (with a positive optimal value), then (1) is feasible (and has a positive optimal value). You can assume that there is at least one data point with \( y_i = 1 \) and one with \( y_i = -1 \) as otherwise there is nothing to classify.
3. Assume the optimal value of \( (2) \) is positive. Show that an optimal solution of \( (2) \) always satisfies \( ||a|| = 1 \).

**Part II: SVMs - Linearly separable data.**

Open the Matlab file \texttt{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 \( (x_1, x_2)_{i=1,...,m} \) are given in the first two vectors and whose labels \( y_i \) are given in the vector \texttt{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 \( (1) \) 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.

**Part III: SVMs - 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 \( (x_1, x_2)_{i=1,...,m} \) 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 \( \ref{fig:convex_hull} \).

Define

\[
A = \left\{ (x_1, x_2)_{i=1,...,m} \mid y_i = 1 \right\}
\]

and

\[
B = \left\{ (x_1, x_2)_{i=1,...,m} \mid y_i = -1 \right\}.
\]
We say that the sets $A$ and $B$ are linearly separable if there exists a hyperplane $a^T x - b$ that takes value $\leq -1$ on $A$ and $\geq 1$ 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. 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:

   \[
   \min_{a, b, \eta} \|a\| + \gamma \|\eta\|_1 \\
   \text{s.t. } y_i(a^T x_i - b) \geq 1 - \eta_i \text{ for all } i = 1, \ldots, m \\
   \eta_i \geq 0 \text{ for all } i = 1, \ldots, m.
   \]

   Solve this problem for $\gamma = 1, 2, \ldots, 10$ and generate two plots: The first one will give the length of the margin (counting both sides) as a function of $\gamma$; the second one will give the number of misclassified points as a function of $\gamma$. Discuss the overall trends of the two plots; are they what you were expecting?
Problem 2: (II Problem 4.3) Prove that $x^* = (1, 1/2, -1)$ is optimal for the optimization problem:

$$\begin{align*}
\text{minimize} & \quad \frac{1}{2}x^T Px + q^T x + r \\
\text{subject to} & \quad -1 \leq x_i \leq 1, i = 1, 2, 3,
\end{align*}$$

where

$$P = \begin{pmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{pmatrix}, \quad q = \begin{pmatrix} -22.0 \\ -14.5 \\ 13.0 \end{pmatrix}, \quad r = 1.$$ 

References