Data Driven Optimization - Tutorial 2

A.K. Cherukuri

April 25, 2023

1 Support Vector Machine

Go to Brightspace to find the *'Iris.xlsx'* dataset. This dataset contains information regarding the iris flowers, listing their sepal length, sepal width, petal length, petal width, and the type of iris flower. There are two flower types, setosa (class 1) and versicolor (class -1). The classes can be found in the last column.

The objective of the assignment is to find a decision boundary that that has the maximum margin, i.e the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

For this assignment, we will use the sepal length and petal length of the flowers. Steps:

1. Load the Iris dataset into Matlab.

2. Define the independent (input) and dependent (output) variables and store them under x_i and y_i.

3. Visualize the data by making a scatter plot (put sepal length on the x-axis and petal length on the y-axis).

4. This is where we'll construct the optimization problem, and hence need the CVX toolbox. Go back to the lecture slides and find which optimization problem we have to solve. During the tutorial, it will be shown how to use the CVX toolbox and calculate the optimal values of w and b. Solve the optimization problem and report the optimizers.

5. Include the decision boundary in the scatter plot.

6. Calculate the width of the margin (see the slides), and plot the support vector lines.