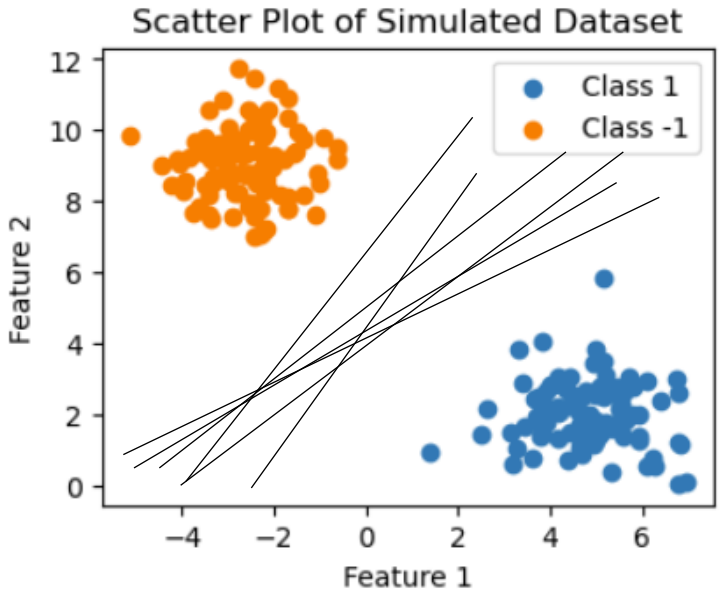
**Problem 1**

The parameter estimates for a linear regression model with an L1 loss function using the advertising dataset are intercept=3.279789, TV=0.043821, Radio=0.196741, Newspaper=0. Comparing these to the parameter estimates in the previous homework, in which we created a linear regression model with an L2 loss function, the parameter estimates are quite similar. Using a linear regression model with an L2 loss function, the parameter estimates were intercept=2.93889, TV=0.0457646, Radio=0.18853, Newspaper=-0.00103749. Comparing the two sets of parameter estimates, we see that the intercept for the L1 loss function is higher than the one for the L2 loss function, the estimate for TV is only slightly lower than the one for the L2 loss function, the estimate for Radio is only slightly higher than the one for the L2 loss function, and the estimate for Newspaper is slightly higher than the one for the L2 loss function.

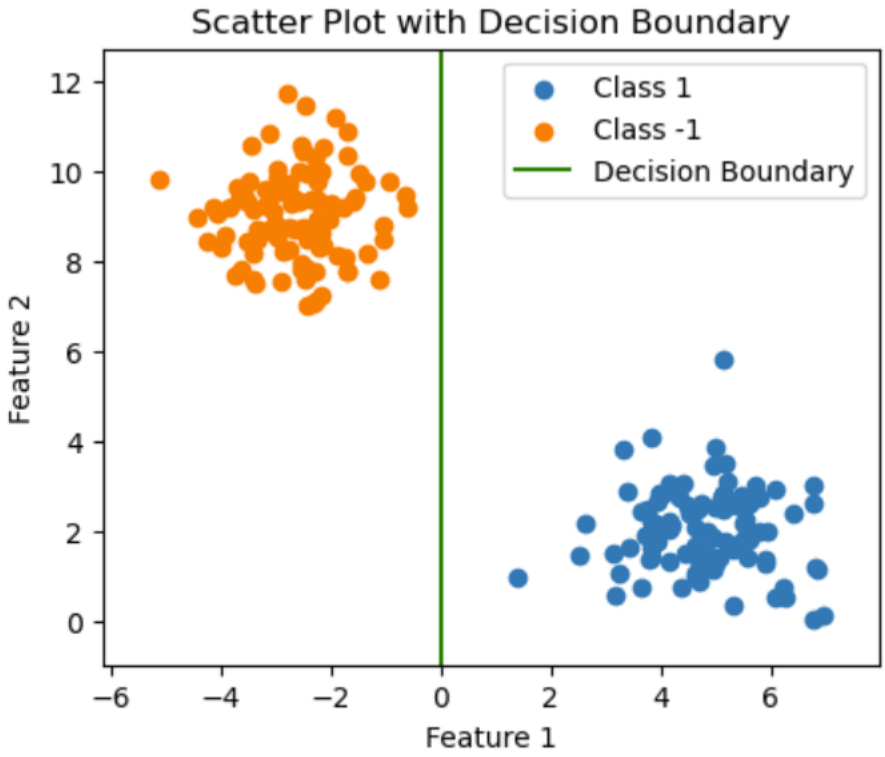
**Problem 2**

Here is the output from the make_blobs function in scikit-learn used to generate 195 data points simulated from multivariate normal distributions.




We can see from the two classes in the plot above that the observations can be separated with a linear classifier (some potential linear classifiers have been added to the plot to show this). Next, Gurobi was used to implement a linear classifier by formulating and solving an optimization problem. The optimization problem was formulated with variables representing the classifier's coefficients and intercept, along with slack variables that allow for soft-margin classification. The objective function aims to minimize the squared norm of the classifier's coefficients—contributing to the maximization of the margin—while also minimizing the sum of the slack variables, which account for any data points that cannot be perfectly classified. The classifier's coefficients and intercept are as follows: intercept= 1.3753370137125657e-11, coefficients=

[1.0752862747425174, 1.4644115138024693e-12]. These values, which define the decision boundary, were then plotted on the scatter plot as shown below.




The linear classifier in the plot above confirms my hypothesis that the simulated dataset can be classified using a linear classifier. Unfortunately, my Gurobi would not optimize the problem for a dataset with more than 195 observations, so that is what the above plots show. Due to this, the decision boundary is essentially a vertical line (which works in this case but maybe could have been improved if more points were generated).