CS 6037 - Assignment 1 Report

Arnauld Cabello - Noah Zukowski

9/20/2025

All team members have contributed in equal measure to this effort

Starting with L2 regularization, this approach imposes a penalty thats proportional to the sum of the squared weights. This has the effect of shrinking all coefficients uniformly towards zero but rarely eliminates them entirely. This allows the model to retain all features while reducing overfitting by controlling the influence of each weight, and it also discourages large weights. Our results align with the expected outcome, as they indicate a slightly positive slope for L2. This particular method is effective in situations where many features affect the target variable, as it maintains all of them and their impacts.

Moving on to L1 regularization, this approach applies a penalty to a model's loss based on the absolute value of its weights, discouraging large values, and pushing many to exactly 0. This is an effective approach for removing less important features from the model. The general trend of reducing weight is expected, but our results differ slightly. Our results suggest that this penalty is too high due to the graphs' decline into negative values, which is unusual. This result could be because of a larger regularization parameter or a high learning rate.

In summary, L2 is better for retaining all features while controlling overfitting is important, whereas L1 is more suitable for feature selection and producing simpler, interpretable models. Each method is valid in its own way and can perform well with proper tuning of parameters and learning rates