Problem 3:

3.1 For this problem, we implement ridge regression and test it on the curve-fitting dataset used in Problem 2. The parameters M and Lambda are taken from [1,2,4,8] and [0, 0,001, 0.01, 0.1, 1, 10], respectively. Below we show the results of the regressions for M = 2. It is clear that as regularization becomes large, the resulting fitted curve becomes increasingly flat and no longer represents a good fit for the data.

<figs>

3.2 Here we test ridge regression on the “regress” datasets. We have the following datasets: “regressA\_train”, “regressB\_train”, and “regress\_validate”. First, we use regressA as the training set and regressB as the test set, validating our trained models using the validate dataset. Call the resulting model M1. Next, we reverse the roles of A and B, using B as training data and A as the test set. Call the model produced here M2. In both cases, we take M from the set [1,2,4,8] and Lambda from the set [0, 0,001, 0.01, 0.1, 1, 10]. The following table displays the SSEs of the models generated from these pairs of values for M and L.

We found that the best parameters for M1 were (M = 2, L = 0), with an SSE of roughly 2.35, and the best parameters for M2 were (M=1, L=1), with an SSE of roughly 32.1.

Problem 4:

In this problem, we use LASSO to model the weights of a dataset generated using a basis [[[[phi(x)]]]] with a small Gaussian noise distribution. The weights are sparse, in that most of them are 0. LASSO tends to give sparse weight vectors, as a result of its use of the L1 norm for regularization.

We take Lambda from the set [0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 2, 5]. For LASSO and ridge regression, we use the given train, validation, and test datasets to generate the best model for each. Below we plot the true weights, weights given by the best LASSO model, and weights given by the best ridge regression model.

<figs>

It is readily apparent that LASSO gives weights very similar to those used in actually generating the dataset. We observe an SSE of 0.43, for L = 0.01 with LASSO. For Ridge Regression, we observe an SSE of 0.36, for L = 0.025. While the SSE of Ridge Regression is lower than that of LASSO, the weights found were very different than the true weights. It seems that while Ridge Regression may outperform LASSO on test data when the datasets are small, it is much less generalizable for sparse weight vectors.