

# COMP 790-134 Machine Learning with Discriminative Methods—Homework 3

The objective of this assignment is to become more familiar with feature selection techniques and to build some intuition for the underlying optimization problems. Due Thursday, Feb 19. This is a quick write-up of what we discussed in class.

## 1 Data

- Construct a dataset with (at least) hundreds of points and three meaningful dimensions. The label variable  $y$  should be a linear combination of those three dimensions. For experiments you should add 0-7 noise dimensions. You will also make two variants of the dataset, either 1) setting one of the noise dimensions to be a “mirror” or copy of one of the meaningful dimensions, and 2) setting one of the noise dimensions to be a copy of another noise dimension. Split the data into train and test subsets (you may do this multiple times to explore how much the results change for different subsets of data).
- Fit model for predicting  $y$  for each of the dataset variants above. You should use least-squares regression, ridge regression, and lasso. Compare the results. You should fit the models on training data, and report results on train and test data. If you do parameter selection, use cross validation.
- Adjust the parameters for lasso and ridge regression to select subsets of the features. Check when these subsets are the meaningful variables. Show how this changes with more noise dimensions and larger magnitudes of noise? How do the variants with repeated variables change the

results? Compare feature selection with lasso and ridge to brute force search over all subsets of features.