

## In-Class Computing Task 14

### Math 253: Statistical Computing & Machine Learning

#### Shrinkage and out-of-sample prediction error

In this activity, you're going to examine how a shrinkage approach can reduce *out-of-sample* prediction error.

Install the `glmnet` package.<sup>1</sup>

You're going to use the `glmnet()` function to fit ridge regression models. The data used in the examples can be loaded with

```
GenCont <- read.csv("http://tiny.cc/dcf/GenCont.csv")
```

This is a data table about the expression of 12 genes and their possible relationship to the expression of a phenotype. The variable `Phenotypes` will be the response in your model. The 12 genes will be the explanatory variables.

<sup>1</sup> Remember, *installing* a package is different than loading it with `library()`. You need to install a package only once, and it will work in any document. So don't put the installation command in your document: use the Packages tab. Your document should contain only the commands to load packages that have been installed previously.

#### Comparing OLS to ridge

Write a function with this interface:

```
compare_ols_ridge <- function(responses, predictors, lambda=1){  
  # your statements will go here  
}
```

Your `compare_ols_ridge()` should do the following:

1. Create a training data set by choosing about half the cases at random. You will need to make both the responses and the predictors.
2. Create a testing data set which will be the remaining cases. Again, you'll need both the responses and the predictors.
3. Fit an ordinary linear regression using `lm()` on the training data.
4. Fit a ridge regression for the specified `lambda` using `glmnet()`. Set `alpha=0` as an argument to get ridge regression.
5. For each of the models in (3) and (4), calculate the *in-sample* predictions using `predict()` and then turn these into an in-sample mean square prediction error.
6. Using the testing data set, calculate the *out-of-sample* mean square prediction error.
7. Your function should return a named vector with these five components: `lambda` as well as `ols_in`, `ridge_in`, `ols_out`, `ridge_out`, each of which has the corresponding mean square prediction error.
8. Use your function to create an object called `Shrink_results` with `lambda = 1` and the `GenCont` data.

#### Exploring Lambda

Use your `compare_ols_ridge()` to look at the in- and out-of-sample prediction errors for a range of  $\lambda$  from 0.1 to 100