R·I·T	**Rochester Institute of Technology** **Golisano College of Computing and Information Sciences** **Department of Information Sciences and Technology**

Lab 4 (2 points)
Linear Model Regularization

This lab consists of two parts, which use R to compare basic and regularized linear models.

Part I

In this exercise, we will predict the number of applications received using the other variables in the College.csv data set.

1. Split the data set into a training set and a test set.
2. Fit a linear model using least squares on the training set, and report the test error obtained.
3. Fit a ridge regression model on the training set, with λ chosen by cross-validation. Report the test error obtained.
4. Fit a lasso model on the training set, with λ chosen by cross-validation. Report the test error obtained, along with the number of non-zero coefficient estimates.
5. Comment on the results obtained. How accurately can we predict the number of college applications received? Is there much difference among the test errors resulting from these five approaches?

Part II

We will now try to predict per capita crime rate in the Boston data set, which is part of ISLR package.

1. Try out some of the regression methods explored in this week, such as best subset selection, the lasso, and ridge regression. Present and discuss results for the approaches that you consider.
2. Propose a model (or set of models) that seem to perform well on this data set, and justify your answer. Make sure that you are evaluating model performance using validation set error, cross-validation, or some other reasonable alternative, as opposed to using training error.
3. Does your chosen model involve all of the features in the dataset? Why or why not?