1. (60 points) Write code to conduct regression by (a) (35 points) utilizing several regression functions: (i) LinearRegression (start line 88), (ii) RANSACRegressor (start line 193), (iii) Ridge (start line 299), and (iv) Lasso (start line 305). These functions are provided by the Python scikit-learn library. (b) (25 points) using one approach to conduct non-linear regression (nonlinear relationships using random forests), and

(2) (20 points) Each regressor needs to be tested using the California housing dataset(#32), which can be loaded using fetch\_california\_housing from sklearn.datasets. You need to use all the columns and all the instances in this dataset. During lectures, I may just use one column and a subset of the dataset for demonstration purpose. If you decide to use fewer columns/instances of the dataset, you need to show an analysis of why you chose to use only a subset of features/instances. For example, you can provide a correlation analysis to justify why you are not using all columns, or show a random sample of the instances. If you do not justify why you are using fewer columns, points will be deducted.

1. (15 points) Properly analyze the regressors’ behavior by applying the knowledge that we discussed in class. Such analysis should include at least Mean squared error (MSE) (or R2 score, or residual plots) and running time. Put the analysis to report.pdf file.

Have tested MSE, R2 score, residual plots and running time in the code. However, cannot get the value due to some unexpected error.

1. (5 points) Write a readme file readme.txt with detailed instructions to run your program.