Name: Ritu Kumari
Roll No.: 2016078
Sec-'B'   Group-2

You have to carry out regression tasks on the following UCI Machine Repository Datasets

Housing - http://archive.ics.uci.edu/ml/datasets/Housing

Computer Hardware - http://archive.ics.uci.edu/ml/datasets/Computer+Hardware

Wine Quality - http://archive.ics.uci.edu/ml/datasets/Wine+Quality

You have to use split the dataset into three parts – approximately the first 70% of training and the last 30% for testing.

The number of variables are respectively 14 (Housing), 9 (Computer Hardware) and 12 (Wine).

Your task is to find out the top 5 most influential variables, responsible for the output.

You have to IMPLEMENT Orthogonal Matching Pursuit (OMP) for this task.

During training, you will run OMP to find out the top 3 and top 5 variables.

During testing, you will use these selected (top 3 or top 5) variables in linear regression to predict the output variable.

You have to report the prediction error in the following table.

	Top 3 Error	Top 5 Error
Housing	11.2015	11.3351
Computer Hardware	132.3807	131.4372
Wine	Red: 0.7375 White: 0.8084	0.7342 0.7528

You have report the root mean squared error (RMSE). Defined https://www.kaggle.com/wiki/RootMeanSquaredError.

NOTE:

You have to write the OMP algorithm on your own. If you copy for the internet, it will be considered plagiarism. If you copy from a fellow student, it will be considered plagiarism. Appropriate action will be taken.