

## Lab 3

The goal of this lab is to provide exercises that will help students reinforce and improve knowledge of Linear Regression analysis, and usage of Matplotlib's Basemap toolkit and mplot3d toolkit. In this lab, we will work with the earthquake dataset (database.csv).

Note: For each exercise enclose your source code and provide a short description of your solution. Motivate your interpretation of pairplot and linear regression analysis.

### Exercise 1

Read in the Date, Latitude, Longitude, Magnitude, Depth, and Type columns from the provided database.csv file.

- Format the Date column so it only contains the year part of the date.
- Use pivot table to show different types of seismic activity (Type) and corresponding number of occurrences (rows) for each type.
- Remove all rows from the input dataset where type of activity is not "Earthquake". In the exercises 2-5 you will be working with this dataset

### Exercise 2

In this exercise you will be working with Matplotlib's Basemap toolkit.

- Review "Geographic Data with Basemap" chapter from your course book. You may use cylindrical map projection to draw your map.
- Over the map drawn in part "a" above, plot longitudes and latitudes as dots marking the map locations where the earthquakes have occurred. Tip: look at the documentation of the Basemap plot function to understand how to incorporate longitude, latitude information into the map.
- Repeat the step b, but now plot only the locations of the earthquake whose magnitude is greater or equal to 8 (Severe).
- What are the min, max and mean values for the Magnitude?

### Exercise 3

- Remove all the rows that contain NaN values. How many (if any) rows were removed?
- During this week's lecture we have worked with seaborn function pairplot. Run this function on your dataset and examine the output plots. Can you see any potential correlation in data? Please motivate your answer.

### Exercise 4

Now that you have gained some insight into data, we are ready to perform a simple linear regression.

- Create a linear regression model where Magnitude is a function of Date (year). Print out the simple regression coefficient and model intercept.
- Use the model created under "a" part to check how well it predicts the values of magnitude for the given input dataset of years. For this purpose calculate also mean square error.
- Plot the actual data for Date and Magnitude (scatter plot) versus the projected Magnitude values (simple line plot).
- Based on analysis done in a-c, how well does your model predict Magnitude? Please motivate your answer.
- Repeat the same analysis a-d, but this time having Depth as independent variable and Magnitude as dependent one.
- Are Date or Depth a good predictor for magnitudes of earthquakes? Please motivate your answer.

### Exercise 5

This exercise is optional and you may work on it once you have completed the exercises 1-4.

- Create linear regression model where Magnitude is a function of two variables Longitude and Latitude.
- Calculate mean square error.
- Use 3d plotting features of mplot3d Axes3D, to show a three d representation of you input data set (x axes Longitude, y axes Latitude, z axes Magnitude). Show the actual data and predicted values.
- Based on analysis done in a-c, how well does your model predict Magnitude? Please motivate your answer