

Lab 14

Linear regression on a real dataset (part two: Polynomial regression)

In this assignment, you will be using Python to perform a more powerful regression for the air quality dataset. You need to use Numpy and pandas libraries to complete this assignment.

Step 1:

Download the two csv files, ozone.csv and solar.csv

Step 2:

In Python, use pandas to read in the two csv files obtained from Step 1. There are NaNs in both files. You will need to remove those NaNs. However, you will need to combine the two data frames together before you can drop the NaNs.

Step3:

You will use Numpy's polyfit function to perform a polynomial regression:

```
import numpy as np
```

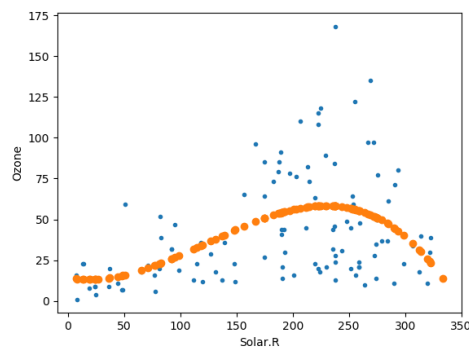
```
z = np.polyfit(x,y,3)
```

```
p = np.poly1d(z)
```

Where x should be a 1-D numpy array holding the values of solar radiation; y is a 1-D numpy array holding the values of ozone; 3 is the order of the polynomial. Once p is formed, you can predict y's value using an x's value. For example, you want to predict the ozone's value when solar radiation is 3, you can just call p(3) and the predicted value will be returned to you.

Step 4:

You will need to plot the graph for the visualization. The expected graph is showing below:



Submission requirement:

Submit a Python script file that contains the solution to all of the steps. Make sure to take a screenshot of the plotted figure.