Linear and Polynomial Regression

You are given with data file atmosphere_data.csv that contains the readings from various sensors installed at 10 locations around IIT Mandi. These sensors measure the different atmospheric factors like temperature, humidity, atmospheric pressure, amount of rain, average light, maximum light and moisture content. The goal of this dataset is to model the atmospheric temperature.

1. Find the pearson correlation coeficient between the attributes
   i) temperature and pressure ii) temperature and humidity
   Comment on the nature of correlation.

2. Split the data into train (70%) and Test (30%).

3. Find the pearson correlation coeficient between the attributes temperature and pressure.

4. Build the simple linear regression (straight-line regression) model to predict temperature given pressure.

   a. Plot the best fit line on the training data where x-axis is pressure value and y-axis is temperature.

   b. Find the prediction accuracy on the training data using root mean squared error (RMSE).

   c. Find the prediction accuracy on the test data using root mean squared error.

   d. Plot the scatter plot of actual temperature (x- axis) vs predicted temperature (y-axis) on the test data. Comment on the scatterplot.

3. Repeat (2) by considering both humidity and pressure as input to predict temperature.

4. Build the simple regression model using polynomial curve fitting to predict temperature given pressure.

   a. Find the prediction accuracy on the training data for the different values of degree of polynomial (p = 2, 3, 4, 5) using root mean squared error (RMSE).
   b. Find the prediction accuracy on the test data for the different values of degree of polynomial (p = 2, 3, 4, 5) using root mean squared error (RMSE).
   c. Plot the scatter plot of actual temperature (x-axis) vs predicted temperature (y-axis) on the test data for the best degree of polynomial (p). Comment on the scatter plot and compare with that of in 2(d).

5. Repeat (4) by considering both humidity and pressure as input to predict temperature (This is called Polynomial regression).

6. **(Assignment)** Find the correlation of temperature with all other attributes. Build linear and polynomial regression models by using attributes having correlation > 0.2 as input to predict temperature.