**Q1. Principal Component Analysis (PCA):**

Given 3 data points in 2-d space, (2, 2), (4, 4) and (9, 9),

**A (Covariance Matrix)**  , where *X* is mxn data matrix.

(a) (2 pt) what is the first principle component? (Find unit vector)

(b) (2 pt) If we want to project the original data points into 1-d space by principle component you choose, what is the variance of the projected data? Also write the projected data.

(d) (2 pts) If we project the given 3 data examples on a vector U = , compute the projected data and also calculate the variance of the projected data?

e) [2 point] In order to find the Eigen vectors of a matrix A, why do we solve Why determinant of this matrix is set to 0.

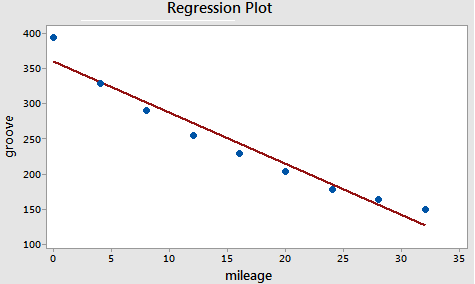
**Q2:**  A scientific laboratory conducted an experiment in order to answer the following research question: "**Is tire tread wear linearly related to mileage?"**

As a result of the experiment, the researchers obtained a data set as shown in Table below, containing the mileage (x, in 1000 miles) driven and the depth of the remaining groove (y, in mils).

|  |  |
| --- | --- |
| **mileage** | **groove** |
| 0 | 394.33 |
| 4 | 329.50 |
| 8 | 291.00 |
| 12 | 255.17 |
| 16 | 229.33 |
| 20 | 204.83 |
| 24 | 179.00 |
| 28 | 163.83 |
| 32 | 150.33 |

****

The fitted line plot of the resulting data suggests that there is a relationship between groove depth and mileage.

****

**The fitted line intercepts y-axis at 360.6**

**a)** What will be the value of θ0 and θ1 for our simple linear hypothesis: **groove = θ0 + θ1 mileage**

**b)** Plot the raw residual of the simple linear regression model of the given data set against the fitted value (predicted values). Plot the raw residuals on the Figure below, and moreover, fill in the table with Fitted (Predicted) values and Residuals.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  |  | | --- | --- | --- | |  | **Fitted value** | **Residual** | | 1 |  |  | | 2 |  |  | | 3 |  |  | | 4 |  |  | | 5 |  |  | | 6 |  |  | | 7 |  |  | | 8 |  |  | | 9 |  |  | |  |  |  | |

**c)** How does your Residual plot helps to determine if your regression function is linear or not-linear. Explain if there is any room for the improvement in the model or not?

**Q3. EDA - Data Transformation:** Suppose the relationship between X (independent variable) and Y (dependent variable) is represented by a power function **Y = 5X3**. Your task is to transform the data in such a way that we can fit a Linear Line using Linear Regression. What will be the approximated intercept term and slope of the line.

**Intercept: ------------------------- Slope: ------------------------------------**