Assignment 03

EE-527 Machine Learning Laboratory

Due Date - 11:59 pm 30 Jan,2022

Problem 1

Write the following function in python to generate n number of points around the line $y = ax + b$

$$[y_{outlier}, y_{noisy}, y_{actual}] = generateDataSet(a, b, xmin, xmax, n, \alpha, \sigma)$$

where $x \in [x_{min}, x_{max}]$, σ is the standard deviation of additive white noise and α is the fraction of outliers present in the data ($\alpha \in (0, 1)$). The output of the function is obtained as follows

$$\begin{aligned} y_{actual}(i) &= ax(i) + b \\ y_{noisy}(i) &= y_{actual}(i) + \sigma\mathcal{N}(0, 1) \\ y_{outlier} &= outlierCorruption(y_{noisy}, \alpha) \end{aligned} \tag{1}$$

Display the scatter plot of the dataset. Plot the inliers in BLUE and outliers in RED.

Problem 2

Perform Regression Diagnostics and display the line obtained in each iteration. Please note that the outliers detected in each iteration should be marked in red color. Experiment with different values of α.

Problem 3

Perform RANSAC on the above set of points and plot the output of each trial. Identify and plot the final line.