DA311 Machine Learning Lab

Assignment 1

1. Generate a set of points around a quadratic curve $y = ax^2 + bx + c$

   (a) Choose $a = 2$ and $b = 3$ and $c = 5$.

   (b) Select the range for x as $[-10, 10]$ and generate $n = 100$ values for x in that interval.

   (c) Compute the values of y for each x.

   (d) Plot the curve in black color.

   (e) Generate a set of n points around the curve using the equation,

   $$y_i = ax^2 + bx + c + \mathcal{N}(0, 3) \tag{1}$$

   where, $\mathcal{N}(0, 3)$ is the zero-mean unity-variance normal distribution.

   (f) Show the scatter plot of these noisy points (in red color) on the same graph generated in step (d).

2. Using Python as your programming language,

   (a) Generate a data matrix x of size $n \times d$, where n is the number of samples/rows and d is the number of features/columns.

   (b) Write a function to standardize this data. This function should work for all values of n and d.

   Note: In a dataset, features often exhibit different levels of variances. By subtracting the mean and dividing by the standard deviation, a data is standardized such that each dimension has zero mean and unit variance.

3. Plot the average error surface E for different values of a and b in the interval of $[-10 : 0.1 : 10]$.

   (a) Vary both a and b in steps of 0.1 in the interval $[-10, 10]$.

   (b) Compute the element-wise error as,

   $$e_i = y_i - \hat{y}_i \tag{2}$$

   where, $\hat{y}_i = ax_i + b$ and y_i is computed using equation (1).

   (c) Compute the average error as,

   $$E = \frac{1}{n} \sum_{i=1}^{n} {e_i}^2 \tag{3}$$

   (d) Compute the average error values for all combinations of a and b.

   (e) Plot the error surface with the values of a along x-axis, that of b along y-axis and E along z-axis.