

Problem 1: You are required to use numpy for operation on matrices. Create a matrix B of dimensions $x \times y$, where $x = 300$ and $y = 80$. Initialize Matrix B. Create a vector v of dimension $y \times 1$. Initialize the matrix with a random values and vector with normal distribution using $\mu = 2$ and $\sigma = 0.01$. Perform following operation on them.

1. Iterative multiply (elementwise) each row of matrix B with vector v and sum the result of each iteration in another vector d
2. Find mean and standard deviation of the new vector d.
3. Plot histogram of vector d using 5 bins.

Problem 2: Use the housing dataset house.csv. This dataset contains information about the sales price of houses along with other attributes. Analyse this dataset using pandas library and plot some interesting information using matplotlib library.

1. Load the data using pandas.
2. Summarize each field in the data, i.e. mean, average etc.
3. Group data by the field furnishingstatus.
 - (a) Give average sqft, average price and average bedroom of each group.
 - (b) Plot for each field (sqft, bedroom, price etc). Use a boxplot that visualizes the statistical information about them.
 - (c) For each group of furnishingstatus, draw a prediction line for price vs sqft.

Problem 3: You are required to pre-process the given wine and housing dataset.

1. convert any non-numeric values to numeric values. For example, you can replace a country name with an integer value or more appropriately use hot-one encoding. [Hint: use hashmap (dict) or pandas.get_dummies]. Please explain your solution.
2. If required drop out the rows with missing values or NA.
3. Split the data into a train (80%) and test(20%).

Problem 4: Implementing Various Types of Plots using matplotlib and seaborn for the given datasets.