Machine Learning First Project

1. (Maximum Likelihood Estimation). Suppose the observed data $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid) and follow a normal distribution. Initially, consider the mean and variance of these variables according to the theoretical distribution. Now, consider the data below to estimate the maximum likelihood.

X_1	1
X_2	1.1
X_3	1.2
X_4	1.5
X_5	0.9
X_6	0.7
X_7	0.5
X_8	1.005

a) Store these data in an array.

b) Obtain the mean and variance from this array based on the estimations.

c) Plot the histogram with the calculated mean and variance.

d) A data point $X_{\text{new}} = -1$ is entered: where does this value fall on the normal distribution curve? (Can you calculate the histogram with this data?)

e) (Law of Large Numbers): Generate 1000 random samples from this histogram. What results can

2. Obtain the Iris flower dataset from the scikit-learn library. This dataset contains 5 features.

a) Create a new dataset including the features Sepal length, Sepal width, Petal length, and Petal width.

b) Plot the features sepal length and sepal width in two dimensions.

c) Obtain the mean of each row and store it in an array. Then, find the centralized covariance matrix.

d) Using the centralized matrix, find the variance.

e) Using the criteria $[Q1 - 1.5\,\mathrm{IQR}, Q3 + 1.5\,\mathrm{IQR}]$, remove outliers and recalculate the covariance matrix for the new dataset.

f) Calculate the Pearson correlation for two datasets with their respective outputs and compare them using `` `df.corr(method = 'pearson')` ``.

g) Obtain the covariance matrix for each part separately using the PCA method while maintaining 95% of the data. Afterward, note that initially, the special values (eigenvalues) should be computed. Compare the results with the example of book 7.5.

3. Store four numerical features from the Iris flower dataset. (Remove the fifth feature related to classification.)

a) Obtain the kernel matrix using a square kernel for this dataset.

b) Then, obtain $\phi(x_i)$s using this matrix.

c) Perform an internal product on the $\phi(x_i)$s and rebuild the kernel matrix. Is this matrix identical to the matrix obtained in part a?

d) Implement the kernel PCA algorithm on the centralized kernel matrix while preserving 90% of the total variance. How many features do we need?