Machine Learning (Problem set 2)

Ali Alipour, *University of Tehran*

This report explores the implementation of various machine learning techniques. In Question 1, we analyze the performance of linear models under noise by calculating MSE, bias, and variance. Question 2 involves classifying random data points using logistic regression and feature mapping. In Question 3, we implement Parzen window estimation for kernel density estimation. These exercises provide insights into the trade-offs between model complexity, bias, and variance.

Question 1

1. First, we define the functions and libraries as required.

2. Next, we generate the required dataset by applying Poisson and normal distributions and introducing noise to the data. After applying the noise, we define a function that creates a linear model with varying degrees. Then, we calculate the Mean Squared Error (MSE) for each model and record the errors.

3. After obtaining the models, we plot the data and the corresponding models.

4. Then, we calculate the bias and variance for the models using the `bias_variance_decomp` function and display the results. We observe that as the model complexity increases, bias decreases and variance increases.

Question 2

1. First, we define the necessary functions and libraries. We generate random data points within a specified radius and display the corresponding images.

2. Next, we repeat the process for a second scenario, generating and displaying the random data points again.

3. Using the generated data, we organize them into a DataFrame and assign labels to them.

4. We then write a manual code to split the data into training and testing sets.

5. Using the Logistic Regression class, we manually implement a logistic regression algorithm to classify the data.

6. Afterward, we increase the features using a mapping function to enhance the classification process.

7. Finally, we implement a classification algorithm using radial data generated from uniform and normal distributions and report the results.

Question 3

In this example, we aim to implement the Parzen algorithm.

1. First, we load the dataset. We define both the Gaussian kernel and the estimated kernel function.

2. We extract the "duration" column from the dataset and represent it as a list. We proceed to plot the results of the kernel density estimations for different values of "duration."

3. We use the `kernelDensity` function to compute the kernel density and plot the results.

4. Lastly, we perform Parzen window estimation for a subset of 250 data points and visualize the results.

References

[Alipour Fraydani, 2024] Alipour Fraydani, A. (2024). Homework on Machine Learning problem set 2, University of Tehran. *Unpublished Manuscript*, Department of Electrical Engineering, University of Tehran.