Centre of Excellence in Artificial Intelligence

AI42001:Machine Learning Foundations and Applications

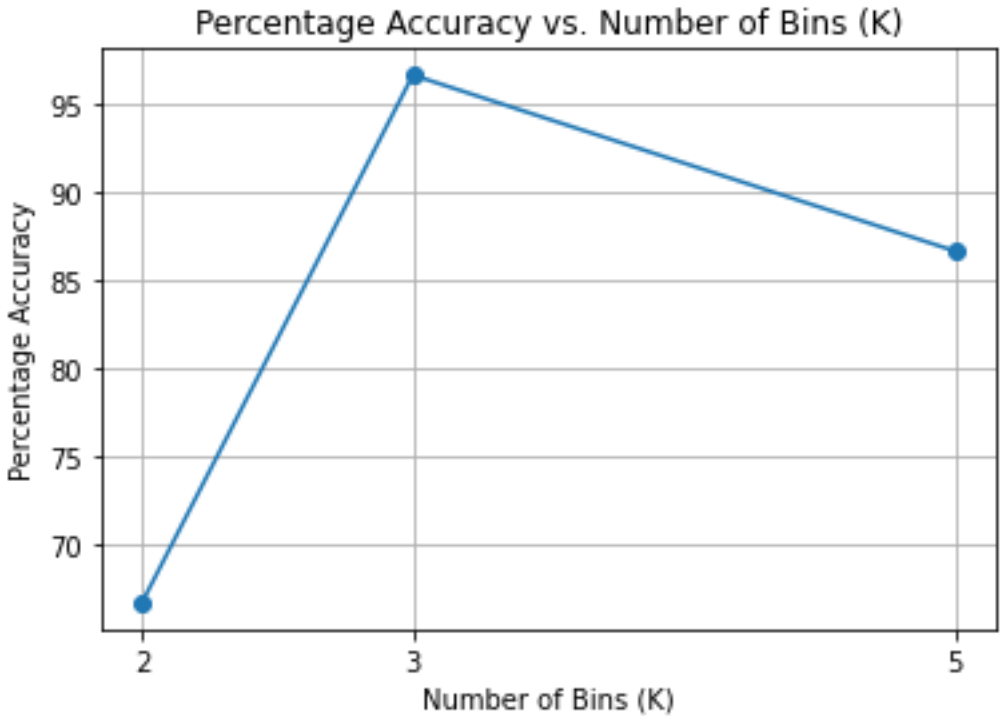
Name-Vaibhav Gupta **Roll Number**-20IE10041

Date:15-02-2024 **Assignment -**4

1. Experiment 1: Report the effect of varying the number of bins K in [NB_CLS] on Test data. Choose K values from [2, 3, 5]. Plot Percentage Accuracy vs K. Find the best value of the hyperparameter K

When shuffling data using random_state= 0, and splitting the data using random_state =42, we get

```
Percentage Accuracy on K=2 is: 66.66666666666666
Percentage Accuracy on K=3 is: 96.66666666666667
Percentage Accuracy on K=5 is: 86.66666666666667
```




```
Best value of K: 3
Best Percentage Accuracy: 96.66666666666667
```

2. Experiment 2: Add noise to only a fraction of the training data: consider separately 10%, 40%, 80%, 90% of the training data for noise addition. Choose a normal distribution with zero mean and standard deviation 2.0. Next, design a Naive Bayes using the optimal K found in the earlier experiment. How does the performance vary as compared to that of the noiseless case (Experiment 1)?

```
Accuracy for noiseless case (Experiment 1): 96.66666666666667
Accuracy with noise fraction 10%: 96.66666666666667
Accuracy with noise fraction 40%: 76.66666666666667
Accuracy with noise fraction 80%: 70.0
Accuracy with noise fraction 90%: 96.66666666666667
```

The observations show that the classifier's accuracy decreases significantly with even a small amount of noise, and while the accuracy improves with higher noise fractions, it remains lower than the noiseless case. Noise has a disruptive effect on the classifier's ability to correctly classify instances, and higher noise fractions can sometimes lead to unexpected behavior in classifier performance.

While the classifier demonstrates a degree of robustness to low levels of noise, its performance is noticeably affected by moderate and high levels of noise. The unexpected behavior observed with higher noise fractions underscores the complex and nuanced nature of noise's impact on classifier performance, emphasizing the need for thorough analysis and mitigation strategies when dealing with noisy data.