HACETTEPE UNIVERSITY

DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING

# ELE489 Fundamentals Of Machine Learning

Hasan KARAŞALLI / 2210357107

Homework 1

Report

# Introduction:

In this homework, K-NN algorithm is implemented from scratch and it is tested with Wine dataset from the UCI Machine Learning Repository.

The  knn.py, analysis.ipynb and README.md files are uploaded to GitHub repository:

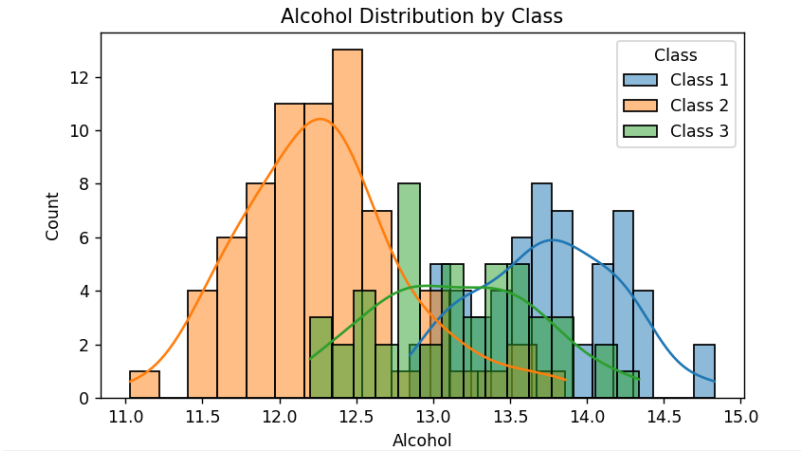
https://github.com/KarasalliHasan/ELE489_ML.git

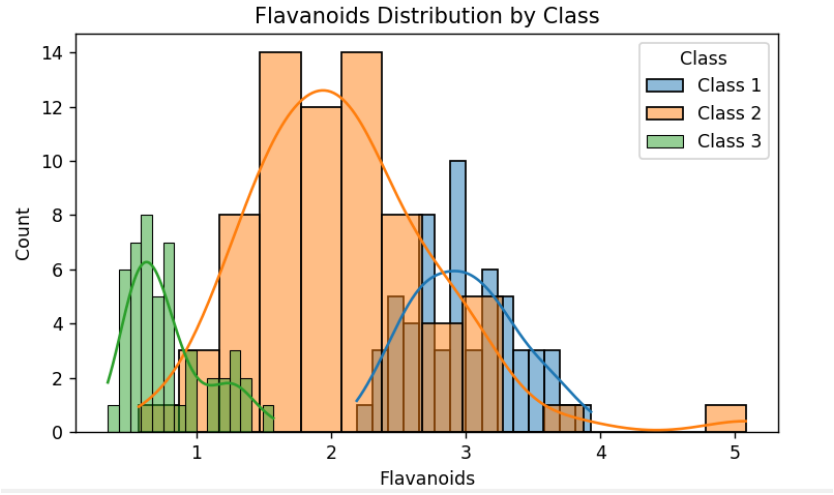
# Results&Comments:

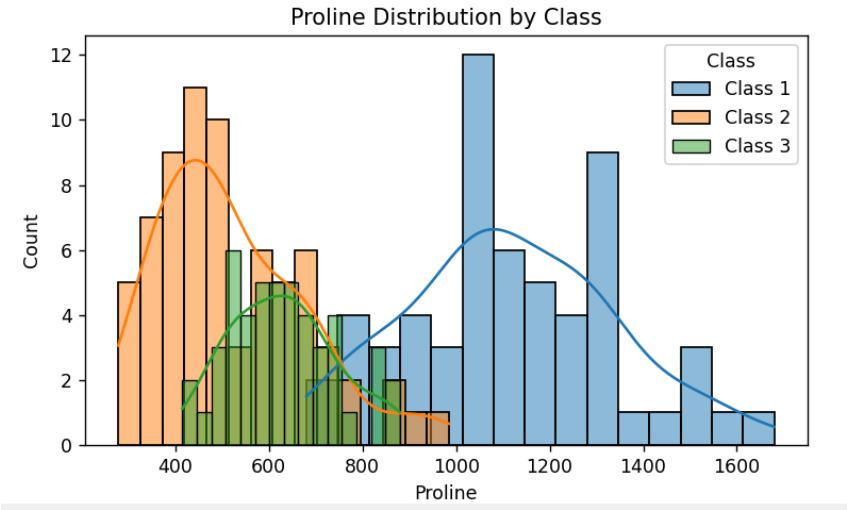
## 1)

After loading the dataset in the code,some of the features of dataset are visualized to see if the features overlap from different classes.

The Alcohol, Flavanoids and Proline features vs. classes are visualized to see the distributions of features over classes.

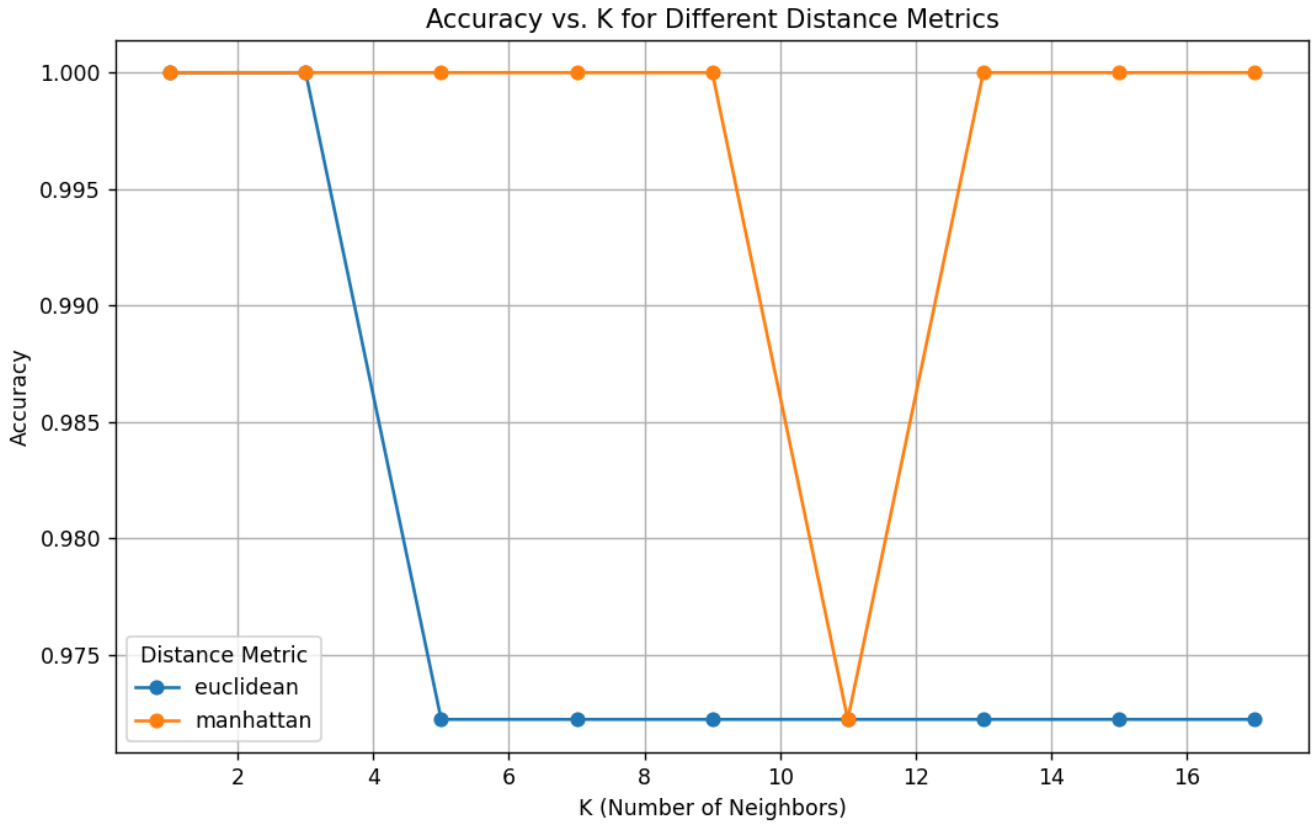





The plots show that the features are mostly well-separated, although there are some points where they overlap, indicating areas of intersection between the classes.

## 2)



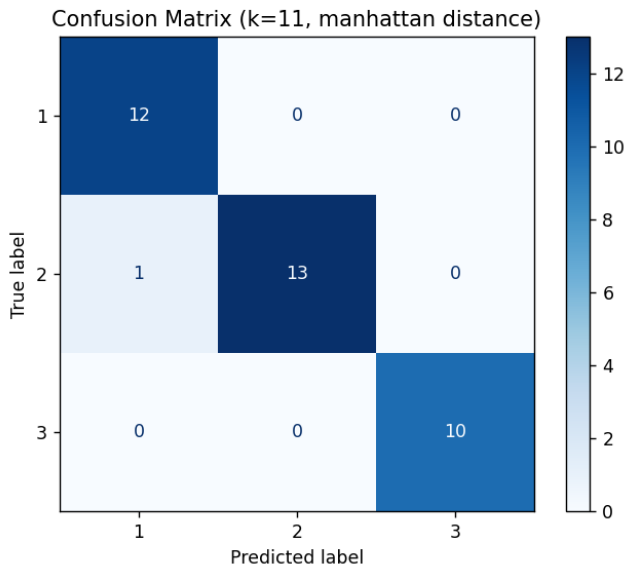

**Comment:** This plot shows the classification accuracy for each value of K using two different distance metrics: **Euclidean** and **Manhattan**.

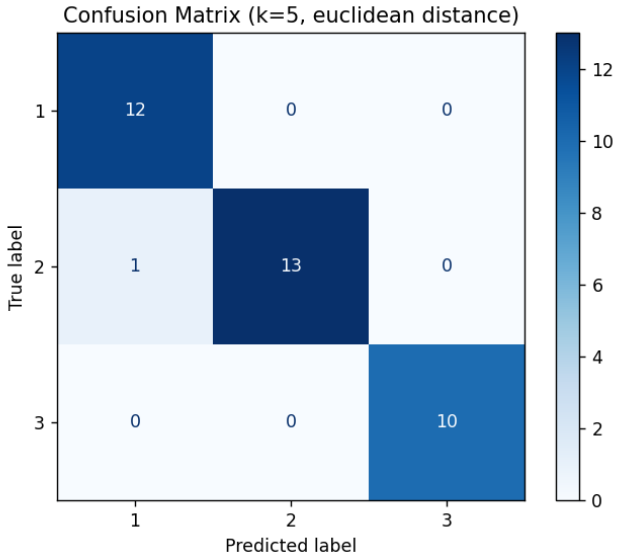
For Euclidean metric,the accuracy is starting with %100 and then the accuracy drops after some K value and remains constant.

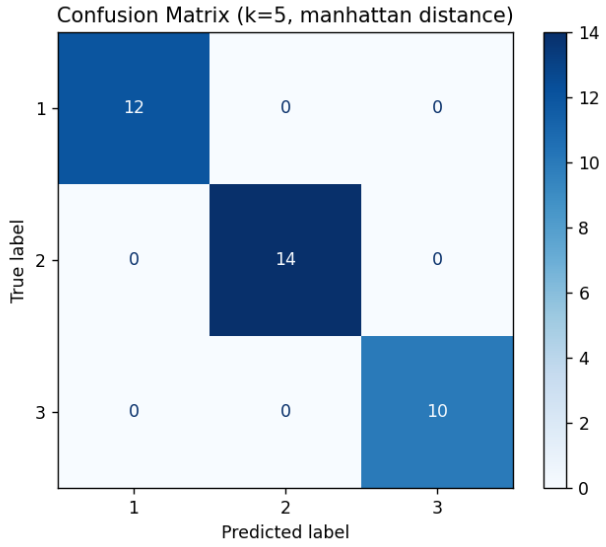
On the other hand, **Manhattan** distance remains constant longer for K values and it only drops at one point(K=11) and then it again rises back to %100 accuracy .

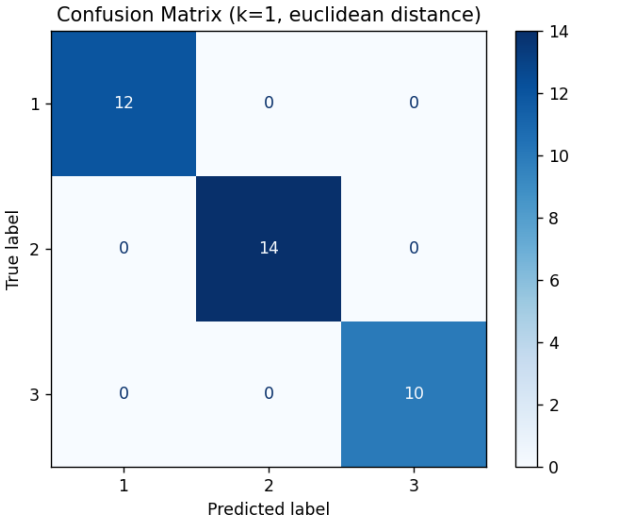
It shows that choosing the optimal K value and distance metric is essential for accuracy of K-NN algorithm.

**3)**




The confusion matrix makes it easier to visualize the performance of the classification model by displaying the number of correct and incorrect predictions for each class. It provides how well the model is good at distinguishing the different classes.It reveals which classes are confused with each other by the model.

**4)**

```
Classification Report:
              precision    recall  f1-score   support

           1       0.92      1.00      0.96        12
           2       1.00      0.93      0.96        14
           3       1.00      1.00      1.00        10

    accuracy                           0.97        36
   macro avg       0.97      0.98      0.97        36
weighted avg       0.97      0.97      0.97        36
```

*Classification report for parameters K=11 and metric="euclidean"*

**Class 1**: The model is 92% accurate when it predicts class 1. It finds **all** actual class 1 cases (100% recall).

**Class 2**: The model is perfect when it predicts class 2 (100% precision), but it misses **7%** of actual class 2 cases .

**Class 3**: The model predicts this class with no errors at all.

**Overall**: The **macro average** is 0.97 for both precision and F1-score, and 0.98 for recall. The **weighted average**, which considers the number of examples per class, is also high at 0.97. In sum,the model has high accuracy and precision according to the classification report.