BREAST CANCER DETECTION

CH Snigdha Setty

Siddharth Pampana

Ealuri Rohith

A.Yamini

M. Jahnavi

INTRODUCTION

- This projet is based on analyzing and predicting the breast cancer , using most popular supervised machine learning algorithm K-Nearest Neighbor with different distance measures And Classification rules using Machine Learning.

- KNN works by finding the distances between a query and all the examples in the data, selecting the specified number examples (K) closest to the query, then votes for the most frequent label (in the case of classification) or averages the labels (in the case of regression).
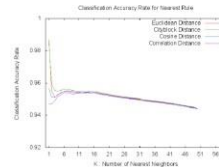
WORK FLOW

- Breast Cancer Data: The data is used to train the machine learning algorithm. We used an inbuilt data from sklearn named "load_breast_cancer". This breast cancer data includes 569 examples of cancer biopsies, each with 30 features.

- Data Analysis: We took a preprocessed data from sklearn and the next step is analyzing the data. There are 212 – Malignant and 357 – Benign.

- Train and Test Split : The original dataset needs to be splitted into training data and test data.Training data is used to train the machine learning model and test data is used to evaluate the model.

- K-Nearest Neighbour : Knn classifier can be used for both classifications and regression tasks. It provides higher accuracy through cross validation.

- Evaluation : Once the model is trained , the model will be evaluated based on the test data which got splitted.

K-NN GRAPHS BASED ON CLASSES
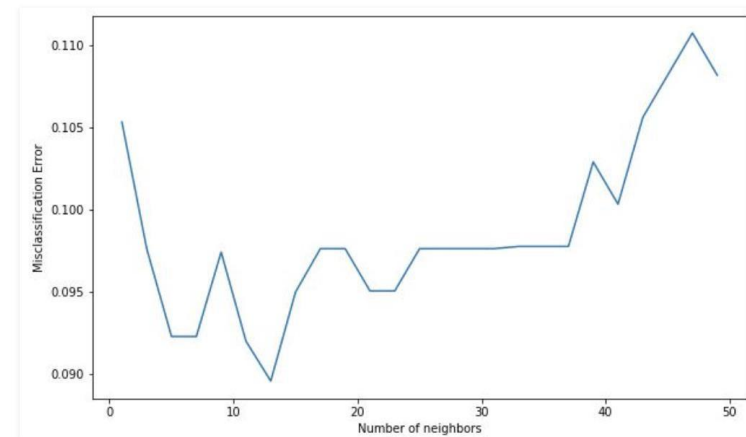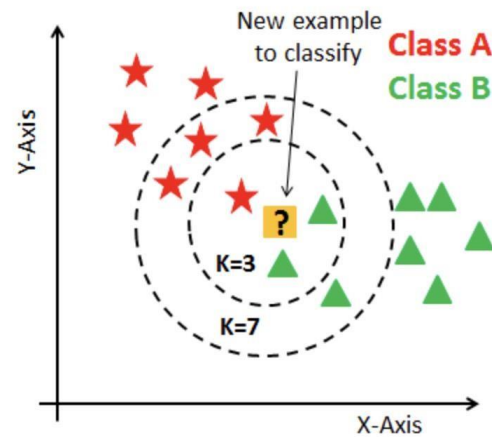
- Taking any two features a graph will be created. Based on those graph values, points will be plotted and two classes will be defined.

- As we have a labeled data the classess will be labeled according to the data set.

- Now by taking values in the train a graph will be plotted and caluculates the distance between its nearest plot. By doing this it will predict the class.

- Based on this th accuracy will be caluculated.

- Initially it goes from low to some extent by increasing the neighbours and again from where you increase the neighbours it drops down because like feeding more and more data the machine may get confused.

CONCLUSION

- It is observed that KNN yields a best accuracy outcome of 95.8%, for larger datasets, the running time of the KNN Algorithm can be high.

- It can be concluded that for a smaller dataset without images and for fast and easy prediction of Breast Cancer using KNN classsifier is the best fit according to the results which we acquired.