breast cancer prediction using support vector machines

INTRODUCTION:

The aim of this project was building a machine learning model using the SVM method. For this purpose, I used a dataset from the Wisconsin which contains much information of patients suffering from breast cancer.

GETTING STARTED:

Importing the libraries and data

Libraries such as pandas, seaborn... will be used to manipulate or have some visualization of the dataset

SOME INSIGHTS:

Id will not be relevant for the SVM kernel analysis, From the info function, we see that diagnosis is considered as an object, so M for malignant and B for benign, we will convert those values to be in binary later in the analysis

Statistic descriptive information:

Let's determine the number of patients suffering from malignant or benign tumour, From this result, we can confirm that we have a 212 malignant and 357 benign cases

we can observe the mean for each variables regarding malignant or benign tumour. We observe a strong difference of means regarding the area_mean, we are going to plot it

From the correlation matrix we obtain:

Strong correlation between the radius mean with area mean and perimeter mean this means that for our analysis we will just keep one of these parameters instead of keeping all of them.

From the texture_mean and symetry_mean we don't see any correlation with other parameters so we will considerate them as unique variables

Using feature engineering we reduced the total parameters columns to 22

Feature Scaling:

Used feature scaling to reduce the scale of the feature so that svm model works more accurately

Model Building:

After feature scaling, we now build the model directly using scikitlearn library

After building our model, we get an accuracy score of 0.96 which makes the model pretty accurate

Confusion Matrix:

From our confusion matrix we obtain:

105 patients considered as having a benign tumour (true negatives)

5 patients considered as having a malign tumour, but it is not (false positives)

4 patients considered as having a benign tumour, but it is not (false negatives)

49 patients considered as having a malign tumour (true positives)