Patrick Callahan, Arjun Chimni, Andrew Schulz

CSCI 184: Machine Learning

5.1.22

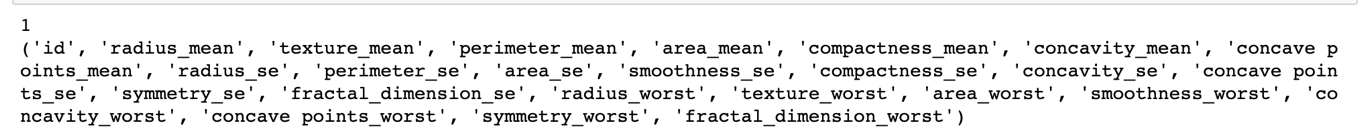
Naive Bayes Classifier Report

* The first thing we did was to download the document and open the cancer.csv file as a Dataframe in Python. Below is a picture of the cancer.csv dataframe:
  + Table

    Description automatically generated
* The next step was to print out the column names and column types. This can be seen below:
  + Table

    Description automatically generated
* Next, we plotted the “Radius Mean’ VS ‘Texture Mean’”. Looking at the data plotted, it appears that the data from these two columns is linearly separated.
  + Chart, scatter chart

    Description automatically generated
* After that, we encoded the target variable and made the M class be 1 and the B class be 0. We added this to the end of the dataframe
  + A picture containing text, jack, screenshot

    Description automatically generated
* After that, we used the wrapper method for feature selection to choose the subset of 25 features as the features for X and the target as Y.
  + 
  + These were the best performing features that we used for our model after performing the wrapper techniques.
* For the Naïve Bayes Classifier, we used the GaussianNB() classifier because the data was continuously distributed rather than being categorically distributed.
  + Chart, treemap chart

    Description automatically generated
  + Graphical user interface, text

    Description automatically generated