Report on simulated data by using KNeighbourClassifier.

Machine learning algorithms like K-nearest neighbor (KNN) are mainly used for classification or regression. However, it is commonly used for classification prediction. Generally, it is based on instant learning and a nonparametric method that uses instances in a training dataset to learn how to classify new input using the scores previously measured.

There are certain steps on how the classification is done such as preparing data, training the model, predicting, and accuracy check. In Preparation for the data, we used the code that has been provided to generate simulated data where the number of samples is 150. Then we split the data into training and test set with a ratio of 80:20.

The next step is to define the classifier using KNeighbourClassifier class. Here, neighbors play a major role so selecting the right number of neighbors provides accurate results. We have given the neighbor parameter as 5. Then we fit the model on the train data, then we will check the accuracy after training the classifier.

The test data has been predicted by using the training model. After the prediction, we can check the accuracy by confusion matrix. We plot accuracy scores as a function of the number of neighbors.

Results:

From the data, I observed that the accuracy score is affected by increasing or decreasing the sample size, hence, I increased the sample size from 150 to 600 to check how the accuracy of the model is changing. The accuracy increases as the number of neighbors increases on training data, but the accuracy of the classifier on the testing data is high at 4 neighbors and decreases as the number of neighbors increases. Hence, the results show that the parameter tuning is so important because it can affect the overall accuracy.