**Final Part 6: Logit vs. KNN**

I have built classification models using logistic regression and KNN (K-nearest neighbors) algorithm to predict whether a patient is diabetic or not. Now, I am evaluating the performance of these models with certain evaluation techniques.

**Compare and evaluate the Accuracy Score:**

What is the difference?

In my analysis, I got accuracy score of 75.32% from logistic regression and 69.26% from KNN.

This is a smaller difference in the scores. This difference appeared because of underlying algorithm of the model. Each model in the machine learning developed with various parameters and techniques. In this context, logistic regression works on linearity principle and KNN algorithm works based on Euclidean distances from observations nearest k neighbors.

Which is best model?

In my opinion, we can choose logistic regression as best model among these two since it gives higher accuracy score. The reason behind higher accuracy can be because of linearity phenomena in the dataset. Hence logistic regression has performed well.

What is your comment?

To conclude, I did expect the results but surprised because of logistic regression performed well then KNN. That can be addressed by linearity patterns in the data.

**Compare and evaluate the Predictions:**

Using the above models, I have tested with 2 sample records in the code with the attributes of preg' , 'plas' , 'pres' ,'skin', 'test', 'mass', 'pedi', 'age'.

Sample Data:

sample1 = [5.5, 130.2, 82, 39, 175, 40, 0.57, 44]

sample2 = [2.5, 110, 62, 19, 135, 20, 0.37, 23]

same data has chosen for comparison of both models.

Results:

Logistic Regression:

predicted class for a patient with data - preg : 5.5 plas : 130.2 pres : 82 skin : 39 test : 175 mass : 40 pedi : 0.57 age : 44 is: 1

predicted value for a patient with the data - preg : 2.5 plas : 110 pres : 62 skin : 19 test : 135 mass : 20 pedi : 0.37 age : 23 is: 0

KNN:

predicted value for a patient with the data - preg : 5.5 plas : 130.2 pres : 82 skin : 39 test : 175 mass : 40 pedi : 0.57 age : 44 is: 1

predicted value for a patient with the data - preg : 2.5 plas : 110 pres : 62 skin : 19 test : 135 mass : 20 pedi : 0.37 age : 23 is: 0

**What is the difference?**

For the given input both the models have given exactly similar outputs. When we compare the results there is no major difference in the output. Since there is a minor difference in the accuracy, these models are giving similar outputs for same input data.

**What is the relationship?**

Both the models have given similar outputs hence we can see they are slightly relative.

**What is the comment?**

Yes, I did expect similar outputs from both the models as they are using same dataset and attributes moreover, accuracies of nearer to each other.

**Compare and evaluate the Cross Validations:**

Next, I have done 10-fold cross validation using same data set and attributes. This evaluation technique uses various sets of test data and produces best accuracy for the dataset provided.

Results

Logistic Regression

Highest accuracy : 0.818 Average accuracy : 0.772, Standard Deviation of the results obtained 0.034

KNN:

Highest accuracy : 0.792 Average accuracy : 0.704, Standard Deviation of the results obtained 0.059

**What is the difference?**

Above are the results obtained after cross validating the data using both models. I could see highest accuracy achieved for logistic regression model is 81.8% while mean accuracy is 77.2%, also standard deviation is 0.034. For KNN, we can see highest accuracy achieved is 79.2%, while mean accuracy is 70.4% and standard deviation is 0.059 which is more than logistic regression.

**What is the relationship?**

There is a difference in the output statistics. The accuracies are not differing at larger extent, hence these models are slightly relative. The best model can be chosen is logistic regression as it has highest accuracy then KNN with least standard deviation.

**What is your comment?**

Yes, I did expect above results. These results are convenient for our data chosen.