**Lab 9: Supervised Machine Learning**

**Amar Fazlic**

**1.4.1 Based on the previous result, explain the content of the variables “x\_train”, “x\_test”, “y\_train”, “y\_test”**

X\_train - The training part of the first sequence (x)

X\_test - The test part of the first sequence (x)

Y\_test - The test part of the second sequence (y)

Y\_train -The training part of the second sequence (y)

**1.4.2 Explain what it means to split a set in a stratified way**

When splitting the stratified way it means we aim to split data so that each split is similar with respect to something, most of the time we aim to achieve that the train and test set have approximately the same percentage of samples.

**2.1.1 What contains the GaussianNB\_predictions variable?**

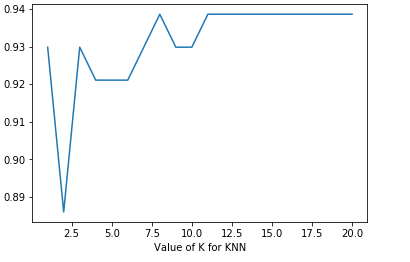
GaussianNB\_predictions = GaussianNB\_model.predict(x\_test)

The GaussianNB\_predictions variable contains the data from the predict. This variable is usually marked as history.

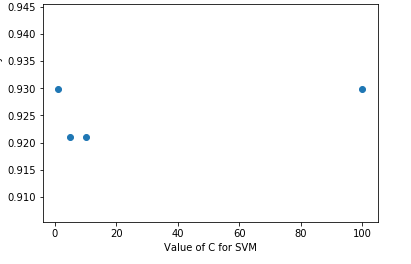
**2.2.1 Describe the relationship between the parameter K and the obtained classification accuracy (different accuracy values depending on the value of K).**

**2.2.2 Based on the results, indicate the optimal K value to classify Breast Cancer data.**

From k=7.5 the accuracy is the highest and then it drops, so there is no need to train further.



**2.3.1 Describe the relationship between the parameter C and the obtained classification accuracy. Include the plot obtained and discuss it.**



**2.3.2 Based on the results, indicate the optimal C value to classify Breast Cancer data.**

The optimal value is at 100.