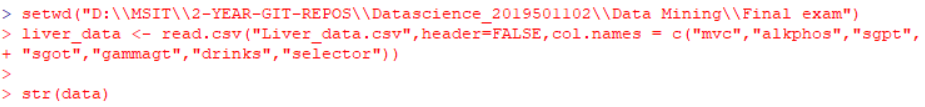
Q4. Fit 1, 2 and 3-nearest-neighbor classifiers to the Liver Disorders Data Set at

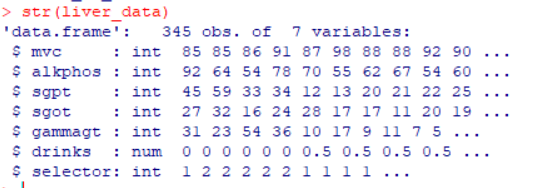
<http://archive.ics.uci.edu/ml/datasets/Liver+Disorders> for measures Euclidean and cosine.

Last but one column is a decision attribute. Replace decision values in to 4 classes (0<=c1<5, 5<=c2<10, 10<=c3<15, 15<=c4<=20). Last column is a data split column in to training and test sets. 1 means the object is used for training. 2 means the object is used for testing. Explain the input parameters you provided for the classifier. Compute the misclassification error on the training data and also on the test data. Annotate your program. (10M)

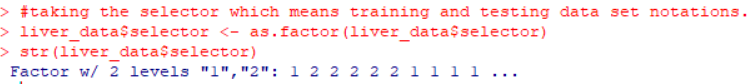
Answer:



* + Setting working directory
  + Reading csv and giving column names as given in the dataset documentation.



* Dataframe consisting of 345 observations of 7 variables.



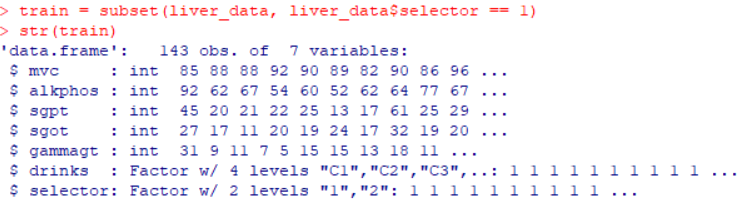
* Storing the selectors in separate variable.

liver\_data$drinks <- cut(liver\_data$drinks, breaks = c(0,5,10,15,20), labels = c('C1','C2','C3','C4'), right = FALSE)

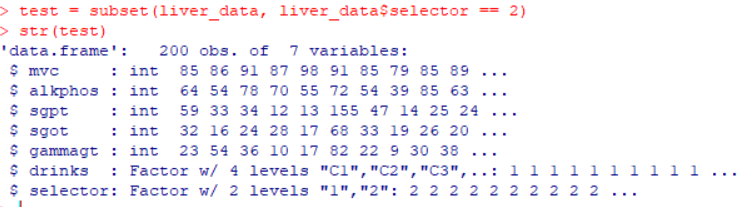
* This command will replace decision values into four columns, c1,c2,c3,c4.

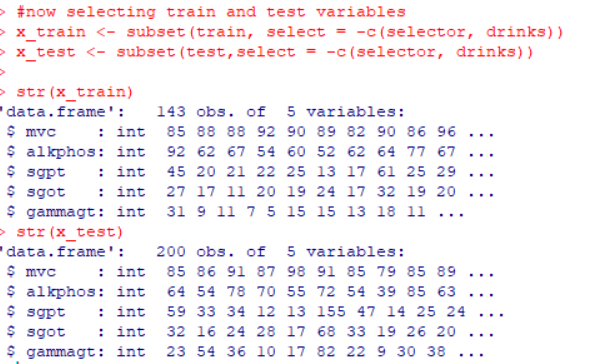


* This will remove all the missing values in the dataset.

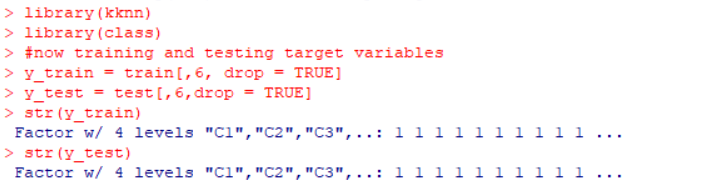


* Selecting the training and testing values, according to selectors as given in question. 1 means for training and 2 means for testing.

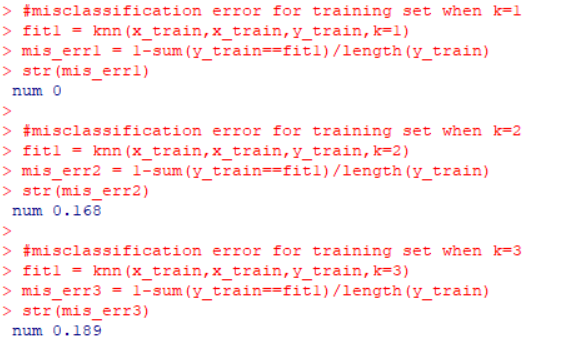




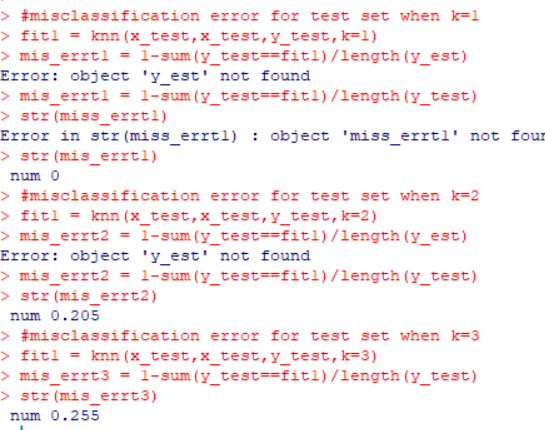
In the above, we have selected the training and testing variables.



Now, as mentioned in the question we have to compute misclassification for k=1,2,3 nearest classifiers models.



* Calculated misclassification error for the training sets. For k=1,2,3 classifiers.



* Calculated misclassification errors for testing sets for k=1,2,3 classifiers