**COMP421X Homework 04**

**Prepared By: Burak Özbek**

**Student Id: 0073811**

**HW04: Decision Tree**

**Question 2:** Using random seed 421, split the dataset into train and test sets by randomly assigning 50% of the data points in each class to the training dataset and the remaining data points to the test dataset.

**Answer:**

# get train and test splits

train\_ratio <- 0.5

set.seed(421)

train\_indices <- c(sample(which(y == 1), floor(sum(y == 1) \* train\_ratio)),

sample(which(y == 2), floor(sum(y == 2) \* train\_ratio)),

sample(which(y == 3), floor(sum(y == 3) \* train\_ratio)),

sample(which(y == 4), floor(sum(y == 4) \* train\_ratio)),

sample(which(y == 5), floor(sum(y == 5) \* train\_ratio)))

X\_train <- X[train\_indices,]

y\_train <- y[train\_indices]

X\_test <- X[-train\_indices,]

y\_test <- y[-train\_indices]

**Question 3:** Learn a decision tree classifier for the classification problem of identifying the cancer types using gene expressions.

**Answer:**

Ok.

**Question 4:** Report the rules extracted from the decision tree:

**Answer:** The extracted rule values is:

[1] "{x10 < 2.99507 AND x6 >= 2.63781} => [150-0-0-0-0]"

[1] "{x10 >= 2.99507 AND x79 >= 10.0297} => [0-0-72-0-0]"

[1] "{x10 < 2.99507 AND x6 < 2.63781 AND x1 < 0.757099} => [0-1-0-0-0]"

[1] "{x10 >= 2.99507 AND x79 < 10.0297 AND x169 >= 9.74823} => [0-0-0-0-68]"

[1] "{x10 < 2.99507 AND x6 < 2.63781 AND x1 >= 0.757099 AND x1 < 2.97659} => [0-0-0-1-0]"

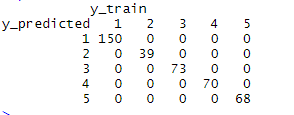
[1] "{x10 < 2.99507 AND x6 < 2.63781 AND x1 >= 0.757099 AND x1 >= 2.97659} => [0-0-1-0-0]"

[1] "{x10 >= 2.99507 AND x79 < 10.0297 AND x169 < 9.74823 AND x14 < 12.1903} => [0-0-0-69-0]"

[1] "{x10 >= 2.99507 AND x79 < 10.0297 AND x169 < 9.74823 AND x14 >= 12.1903} => [0-38-0-0-0]"

**Question 5:** Calculate the confusion matrix for the training set using the decision tree you will learn from the previous step.

**Answer:** Calculated confusion matrix for the train data points is shown as below:



**Question 6:** Calculate the confusion matrix for the data points in your test set using the discrimination rule you will develop using the trained multilayer perceptron.

**Answer** : Calculated confusion matrix for the test data points using trained decision tree values is shown as below

