**COMP421X Homework 05**

**Prepared By: Burak Özbek**

**Student Id: 0073811**

**HW05:Multiclass Kernel Machine**

**Question 2:** You are given a multivariate classification dataset that includes the actual forest cover types of 1500 samples. First 13 columns of the dataset show the geographic features of the samples, such as soil type and wilderness of the area. The last column of the dataset represents the cover type of the samples (i.e., 1, 2, or 3). There are 500 samples for each cover type.

**Answer:** Data is seperated as DataSet and Y vector

# read data into memory

data\_set <- read.csv("forest\_cover\_data.csv")

# get X and y values

X <- subset(data\_set, select = -c(Cover\_Type) )

y <- data\_set$Cover\_Type

**Question 3:** Divide the dataset into two parts by assigning the randomly selected 400 samples of each cover type to the training set and the remaining samples to the test set.

**Answer:**

# get train and test splits

train\_ratio <- 0.8

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)))

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

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

**Question 4:** Train SVM models for each class using the *one-vs-all* approach and the Gaussian kernel. Define kernel width parameter *s* as 5, and the regularization parameter *C* as 10.

**Answer:** Confusion matrixes for train data points which I calculated are shown as below:

[[1]]

-1 1

-1 798 21

1 2 379

[[2]]

-1 1

-1 756 79

1 44 321

[[3]]

-1 1

-1 742 18

1 58 382

**Question 5:** Calculate the confusion matrix for the predictions you made for the test data points using the parameters you learned from the trained SVM models.

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

y\_test

y\_test\_predicted 1 2 3

1 93 6 0

2 7 81 8

3 0 13 92