Week 5 Assignment: SVM Classifier

Scott Schirkofsky  
MSDS 680 Machine Learning

Regis University

SVM Classifier

**Introduction**

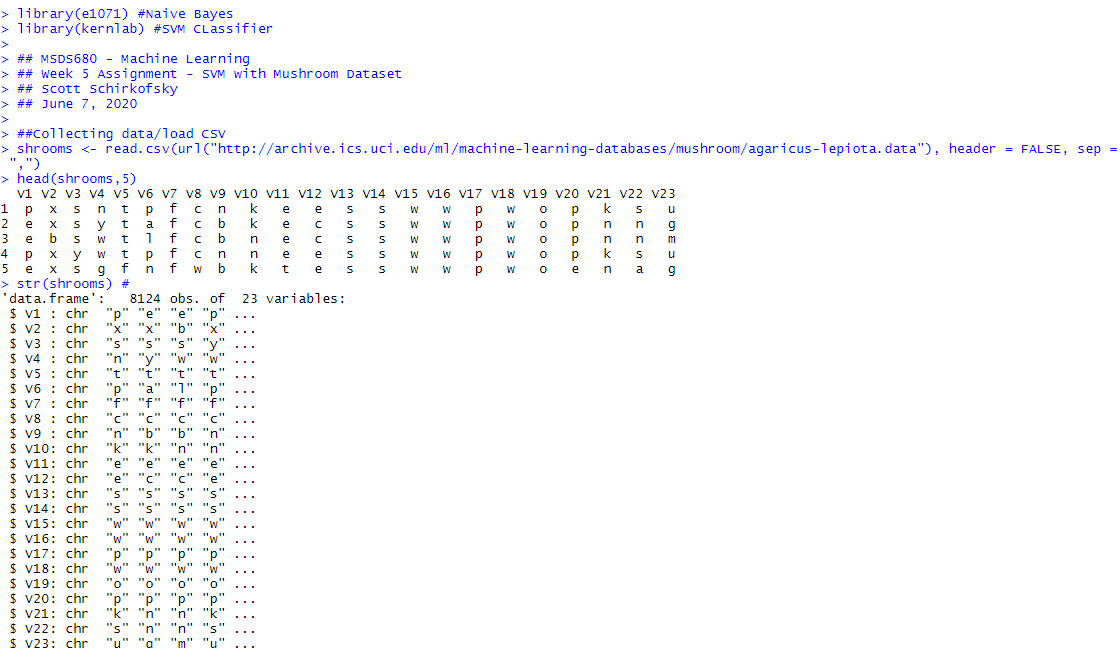
This assignment is SVM classifiers will utilize the mushroom dataset from UCI.

From the given data set: <http://archive.ics.uci.edu/ml/datasets/Mushroom>

This report is incomplete as I could not get the code to work and have given up on trying to make it work. I’ve included my R code file and a separate R file called “sandbox.r”. The sandbox file contains the three online examples that I was trying to use, none of them actually work. I’m lost and surrender on this assignment. It’s more than I can handle. I understand all the reading but there is a huge disconnect for me in applying the reading to the assignment.

**Collecting Data**

shrooms <- read.csv(url("http://archive.ics.uci.edu/ml/machine-learning-databases/mushroom/agaricus-lepiota.data"), header = FALSE, sep = ",")  
head(shrooms,5)  
str(shrooms)



**Transforming Data**

colnames(shrooms) <- c("edibility", "cap\_shape", "cap\_surface", "cap\_color", "bruises", "odor", "grill\_attachment", "grill\_spacing", "grill\_size","grill\_color", "stalk\_shape", "stalk\_root", "stalk\_surface\_above\_ring", "stalk\_surface\_below\_ring", "stalk\_color\_above\_ring", "stalk\_color\_below\_ring", "veil\_type", "veil\_color", "ring\_number", "ring\_type", "spore\_print\_color", "population", "habitat")  
sum(is.na(shrooms)) #check for NAs, none found  
shrooms <- subset(x = shrooms, select = -veil\_type) #remove veil\_type  
levels(shrooms$'stalk.root')[levels(shrooms$'stalk.root')== "?"] <- "u" #   
Table(shrooms$stalk.root) # Inspect Again levels = c("e","p")  
labels=c("edible","poisonous")  
shrooms <- as.data.frame(lapply(shrooms, as.numeric)) # Change all rows to numeric  
shrooms$class <- factor(shrooms$class) # Convert class back to factor  
str(shrooms)

**Build a classifier**

set.seed(77) # Get the same data each time  
idx <- sample(nrow(shrooms), round(nrow(shrooms)\*0.7)) # Create 2 samples with ratio 70:30  
shrooms\_train <- shroom[idx, ] # Split 70%  
shrooms\_test <- shroom[-idx, ] # Split 30%

##Train the model  
shrooms\_model <- svm(formula = class ~ ., data = shrooms\_train, kernel = "linear")  
summary(shrooms\_model)

##Test the Model  
shrooms\_pred <- predict(shrooms\_model, shrooms\_test)  
summary(shrooms\_pred)

##Evaluating Accuracy  
# Returns the percentage of correct predictions  
get.accuracy <- function(prediction, real) {

accuracy <- prediction == real

return (length(accuracy[accuracy == TRUE])/length(accuracy))

}  
get.accuracy(shrooms\_pred, shrooms\_test$class)

References

Whithell, R. (n.d.). Support Vector Machines. *RyanWhitell.com*. Retrieved from <http://ryanwhitell.com/mlr/mlr_svm>