HW\_2.2

Arjun Goyal

8/28/2019

## K-Nearest Neighbors

#KKNN Function

We write the function *kknn\_accuracy* to evaluate the k-nearest neighbors model using different values of k.

kknn\_accuracy = function(k\_val){  
 pred\_responses <- rep(0, nrow(cred\_card\_data\_headers)) #vector of 0s to insert the predicted response value of the model.  
 for (i in 1:nrow(cred\_card\_data\_headers)){  
 #using [-i] to exclude the ith data point from the nearest neighbor calculation  
 kknn.model = kknn(R1 ~., train = cred\_card\_data\_headers[-i,], test = cred\_card\_data\_headers[i,], k = k\_val, kernel = "rectangular",scale = TRUE)  
 #fitted(kknn.model) gives the value of the fraction of nearest neighbors to the ith data point that are 1.  
 fit <- fitted(kknn.model)  
 binary\_response <- as.integer(fit + 0.5) #maps the fitted value to 0 or 1  
 pred\_responses[i] <- binary\_response #adds the model's fitted value to the pred\_responses vector  
 }  
 acc = sum(pred\_responses == cred\_card\_data\_headers[,11]) / nrow(cred\_card\_data\_headers)  
 return(acc)  
}

We run a for loop iterating through [0,20] for values of k.

k\_acc <- rep(0, 20)  
for (i in 1:20){ #values of k  
 k\_acc[i] = kknn\_accuracy(i)  
}

To determine the best accuracy delivered by the model, we run:

max(k\_acc) #best k accuracy

## [1] 0.851682

To determine the value of k for which this best accuracy occurs, we run the following code to find the index of the maximum accuracy:

max\_k <- which(k\_acc %in% max(k\_acc))  
max\_k

## [1] 6

#Plotting Accuracies

We can create a plot of the values of k vs. the accuracies: 