STA 4320 HW 4

Mary Robinson

12/4/2024

Question 1

Part a

Write a helper function

```
# "euclidean" is the default method, "manhattan" is used for other problems
two.dist = function(a, b, m = "euclidean"){
  X = rbind(a, b)
  return(as.numeric(dist(X,method = m)))
}
#as.numeric removes the row and column names
```

Part b

KNN regression function

```
# x is the data matrix without response
# y is the response vector
# k is the number of nearest neighbors in KNN
# x_new is a new data point that we want to predict

knn.reg = function(X, y, k, x_new){
  n = nrow(X) #extract the stats
  dist = numeric(n) # store pairwise distances between Xnew and each row of X
  for (i in 1:n){
    dist[i] = two.dist(X[i,], x_new, "euclidean")
    # apply the helper function to compute pairwise distances
  }
  k_near_loc = order(dist)[1:k]
  # order these pairwise distances and select the k smallest
  return(mean(y[k_near_loc]))
}
```

Part c

Test to see if our model works. We need it to equal 12.

```
X = matrix(1:5, nrow = 5, ncol = 1)
# 1:5 is the entries in the matrix and nrow and ncol is the matrix size mn
y = 11:15
x_new = 2.1
k = 3
knn.reg(X, y, k, x_new)
```

```
## [1] 12
```

```
# call the function name and the arguments
```

Our model works!

Part d

Let's see what our KNN model predicts for this sample dataset

```
#require the auto dataset ISLR2 package
require(ISLR2)

X = Auto[, 2:5] # data matrix of size n by 4
y = Auto[, 1] # response vector
k = 10
x_new = c(6, 310, 140, 3500) # new x to make prediction; the ordering matters
knn.reg(X, y, k, x_new) # call the function in the appropriate order
```

```
## [1] 18.16
```

Our KNN model predicts that the MPG when a car has 6 cylinders, 310 displacement of cylinders, 140 horsepower, and a weight of 3500 is 18.16.