Math 445 HW 5

Erick Castillo

3/12/2021

Problem 1: Use a model based clustering method to cluster the data in the Titanic dataset. Compare the results with the methods used in the prior HW.

```
BIC = mclustBIC(pred.cols)
mod1 = Mclust(pred.cols, x=BIC, G=2)

try1 = ifelse(mod1$classification == 1, 1, 0)
try2 = ifelse(mod1$classification == 1, 0, 1)

sum(abs(boat$Survived-try1))
```

```
## [1] 311
```

```
sum(abs(boat$Survived-try2))
```

```
## [1] 401
```

Notice that this method of clustering scored worse than the methods used on the prior HW. Here, the correct number of values that got clustered was 401. This is smaller than kmeans, divisive, and agglomerative which got scores of 446, 427, and 427 respectively.

Problem 2: Use Ward's hierarchical clustering method. Compare to the other methods.

```
wardclust = hclust(dist(pred.cols), method = "ward.D")
ward = cutree(wardclust, k = 2)

try1 = ifelse(ward == 1, 1, 0)
try2 = ifelse(ward == 1, 0, 1)

sum(abs(boat$Survived-try1))
```

```
## [1] 471
```

```
sum(abs(boat$Survived-try2))
```

```
## [1] 241
```

This Ward's hierarchical clustering got the best clustering score of all prior 5 methods. This method was able to correctly cluster 471 values. All other methods scored values less than this.