

Final Report

Samuel Burge, Chad Austgen, Skylar Liu

April 19, 2022

Supervised Learning Task

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
## 1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##   Mean  :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
##   Max.  :25.0    Max.    :120.00
```

Methodology

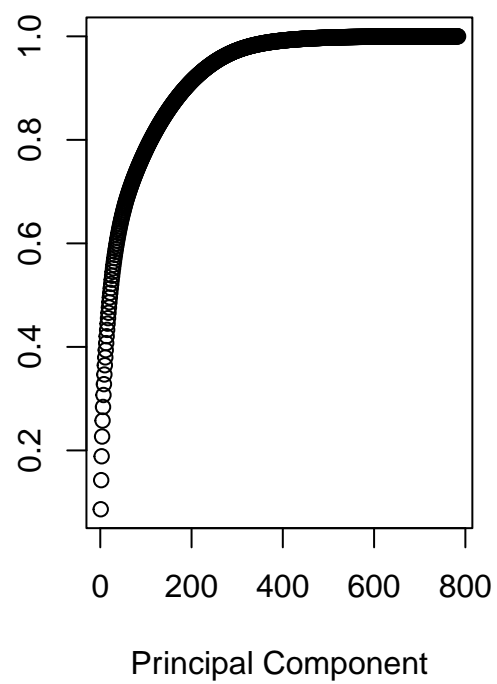
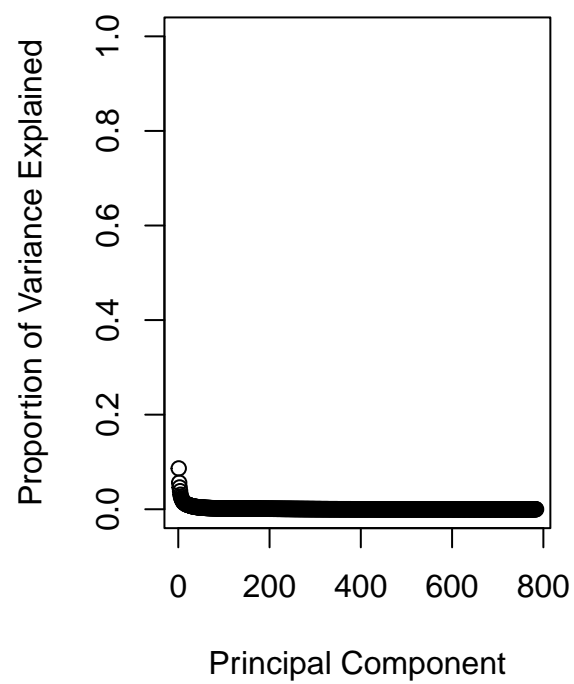
...

Results

We can put a description of the results, like a table of the average training and cross-validation errors, in this section. I was also thinking we could include ROC curves.

Unsupervised Learning Task

For the unsupervised learning task, we initially looked at performing principal components analysis (PCA) to reduce the number of dimensions in the data set. The results produced principal components that did not significantly reduce the feature space while simultaneously retaining a large portion of the information (or rather the variance) from the original data. The two scree plots below show that the number of principal components we would have to use to capture at least 90% of the variation in the data set was about 187, which did not seem beneficial enough to consider for the analysis. Therefore, we decided to retain all the original features in the data set.



References