Homework #4

Eric Pettengill

Using the `Default` dataset do the following.

(a) Fit a logistic regression model that uses `income` and `balance` to predict `default`

```
library(ISLR)
library(caret)

glmfit <- train(default ~ income + balance,
              data = Default,
              method = "glm",
              family = "binomial")
```

(b) Estimate the test error using 5-fold cross validation.

```
set.seed(1018)

fitControl <- trainControl(method = "repeatedcv",
                           number = 5)

glmfit <- train(default ~ income + balance,
              data = Default,
              method = "glm",
              family = "binomial",
              trControl = fitControl)

glmfit$resample[,c(3,1)]
```

```
##     Resample  Accuracy
## 1 Fold1.Rep1 0.9730000
## 2 Fold2.Rep1 0.9740000
## 3 Fold3.Rep1 0.9734867
## 4 Fold4.Rep1 0.9760120
## 5 Fold5.Rep1 0.9720000
```

```
## test error
1-mean(glmfit$resample[,1])
```

```
## [1] 0.02630025
```

(c) Add `student` to the model and estimate the test error using 5-fold cross validation. Comment.

```
set.seed(1018)

glmfit <- train(default ~ income + balance + student,
              data = Default,
              method = "glm",
              family = "binomial",
              trControl = fitControl)

glmfit$resample[,c(3,1)]
```

```
##     Resample  Accuracy
```

```
## 1 Fold1.Rep1 0.9730000
## 2 Fold2.Rep1 0.9735000
## 3 Fold3.Rep1 0.9729865
## 4 Fold4.Rep1 0.9760120
## 5 Fold5.Rep1 0.9710000
```

```
## test error
1-mean(glmfit$resample[,1])
```

```
## [1] 0.0267003
```

The model in (b) performs slightly better than the model in (c)

5.4 #6

(a) Determine the estimated standard errors for the coefficients(`income` and `balance`) of the model fit in part (a) above.

```
glmfit <- glm(default ~ income + balance, data = Default, family = "binomial")

summary(glmfit)$coefficients[-1,2]
```

```
##       income      balance
## 4.985167e-06 2.273731e-04
```

(b) Write a function, `boot.fn`, to use in the `boot()` function to estimate the standard errors.

```
boot.fn <- function(data, index) {
  coef(glm(default ~ income + balance, data = Default, family = "binomial", subset = index))[c(2,3)]
}
```

(c) Use the `boot()` function to estimate the standard errors of the model in (a).

```
set.seed(1018)

boot.obj <- boot::boot(Default, boot.fn, R = 1000)

apply(boot.obj$t, 2, sd)
```

```
## [1] 4.791449e-06 2.312668e-04
```

(d) Compare the estimates to the true values.

```
data.frame(Actual = summary(glmfit)$coefficients[-1,2],
           Bootstrap = apply(boot.obj$t, 2, sd))
```

```
##               Actual    Bootstrap
## income  4.985167e-06 4.791449e-06
## balance 2.273731e-04 2.312668e-04
```

The bootstrap estimate for `income` is slightly lower than the actual value given in the `glm` output while the bootstrap estimate for `balance` in slightly higher than the actual value.