

Stock Market-Exercises

December 5, 2018

0.1 Exercise

Using the Smarket dataset (all data, not only training) perform 10 fold-cv with logistic regression.

```
In [1]: library(ISLR)
        require(caret)
```

```
Loading required package: caret
Loading required package: lattice
Loading required package: ggplot2
```

```
In [2]: glmFit <- train(Smarket[, -8:-9],
                        y = Smarket[, 9],
                        method = "glm",
                        preProcess = c("center", "scale"),
                        tuneLength = 10,
                        control=glm.control(maxit=500),
                        trControl = trainControl(method = "cv"))

                        glmFit
```

Generalized Linear Model

```
1250 samples
  7 predictor
  2 classes: 'Down', 'Up'
```

```
Pre-processing: centered (7), scaled (7)
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 1126, 1125, 1124, 1125, 1125, 1126, ...
Resampling results:
```

Accuracy	Kappa
0.5136197	0.01439017

Hemos utilizado la regresión logística para el conjunto entero de datos de Smarket (no solo el training) pero hemos prescindido de la columna today, ya que la variable categórica "Direction" se obtiene a partir de esta.

Podemos observar que hemos obtenido una mejor precisión y valor de Kappa aunque la mejora es bastante baja. A continuación prediciremos el modelo.

```
In [3]: glmPred = predict(glmFit, type="raw")  
        table(Smarket$Direction, glmPred)  
        mean(glmPred==Smarket$Direction)
```

```
      glmPred  
      Down  Up  
Down    221 381  
Up      190 458
```

0.5432

Como consecuente podríamos deducir que este modelo no sería el mejor para realizar una clasificación de este data set debido a que la tasa media de acierto es baja.