Homework 3

1

1

```
data("Weekly")

Weekly = Weekly %>%
  janitor::clean_names() %>%
  select(-today)

skimr::skim_without_charts(Weekly)
```

Table 1: Data summary

Name	Weekly
Number of rows	1089
Number of columns	8
Column type frequency:	
factor	1
numeric	7
Group variables	None

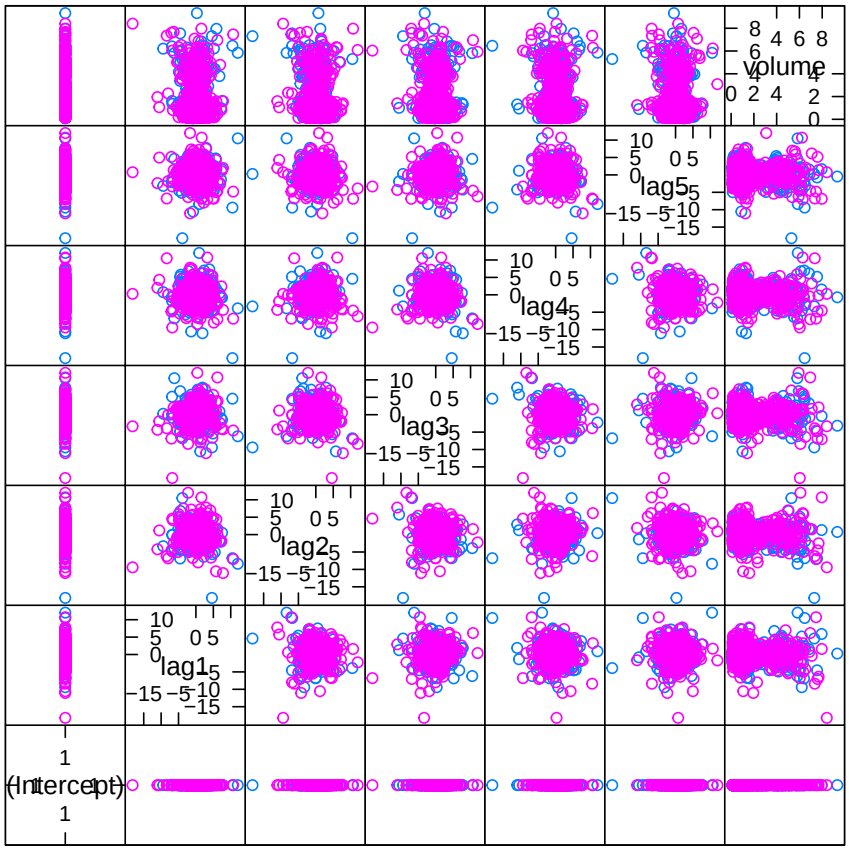
Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
direction	0	1	FALSE	2	Up: 605, Dow: 484

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
year	0	1	2000.05	6.03	1990.00	1995.00	2000.00	2005.00	2010.00
lag1	0	1	0.15	2.36	-18.20	-1.15	0.24	1.41	12.03
lag2	0	1	0.15	2.36	-18.20	-1.15	0.24	1.41	12.03
lag3	0	1	0.15	2.36	-18.20	-1.16	0.24	1.41	12.03
lag4	0	1	0.15	2.36	-18.20	-1.16	0.24	1.41	12.03
lag5	0	1	0.14	2.36	-18.20	-1.17	0.23	1.41	12.03
volume	0	1	1.57	1.69	0.09	0.33	1.00	2.05	9.33

```
caret::featurePlot(model.matrix(direction~.,Weekly %>% select(-year)),Weekly$direction,"pairs")
```



Scatter Plot Matrix

```
Weekly_Tr = Weekly %>%
  filter(year <=2008)

Weekly_Ts = Weekly %>%
  filter(year > 2008)
```

2

```
Weekly_logistic =
  train(
    X_tr,
    Y_tr,
    method = "glm",
    family = "binomial",
    trControl = TRC,
    metric = "ROC",
    preProcess = c("center", "scale")
  )
```

```
logistic_prediction =
  predict(Weekly_logistic,newdata = X_ts, type = "raw")

confusionMatrix(logistic_prediction,Y_ts)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction Down Up
##       Down   31 44
##       Up     12 17
##
##                Accuracy : 0.462
##                  95% CI : (0.363, 0.562)
##     No Information Rate : 0.587
##     P-Value [Acc > NIR] : 0.996
##
##                   Kappa : 0
##
##  Mcnemar's Test P-Value : 3.43e-05
##
##             Sensitivity : 0.721
##             Specificity : 0.279
##          Pos Pred Value : 0.413
##          Neg Pred Value : 0.586
##              Prevalence : 0.413
##          Detection Rate : 0.298
##    Detection Prevalence : 0.721
##       Balanced Accuracy : 0.500
##
##        'Positive' Class : Down
##
```

3

```
Weekly_logistic2 =
  train(model.matrix(direction~lag1+lag2,Weekly_Tr)[,-1],
        Y_tr,
        method = "glm",
        metric = "ROC",
        trControl = TRC,
        preProcess = c("center","scale"))
```

```
logistic_roc = pROC::roc(Y_ts,predict(Weekly_logistic2,newdata = X_ts,type = "prob")[,2])
```

4

LDA

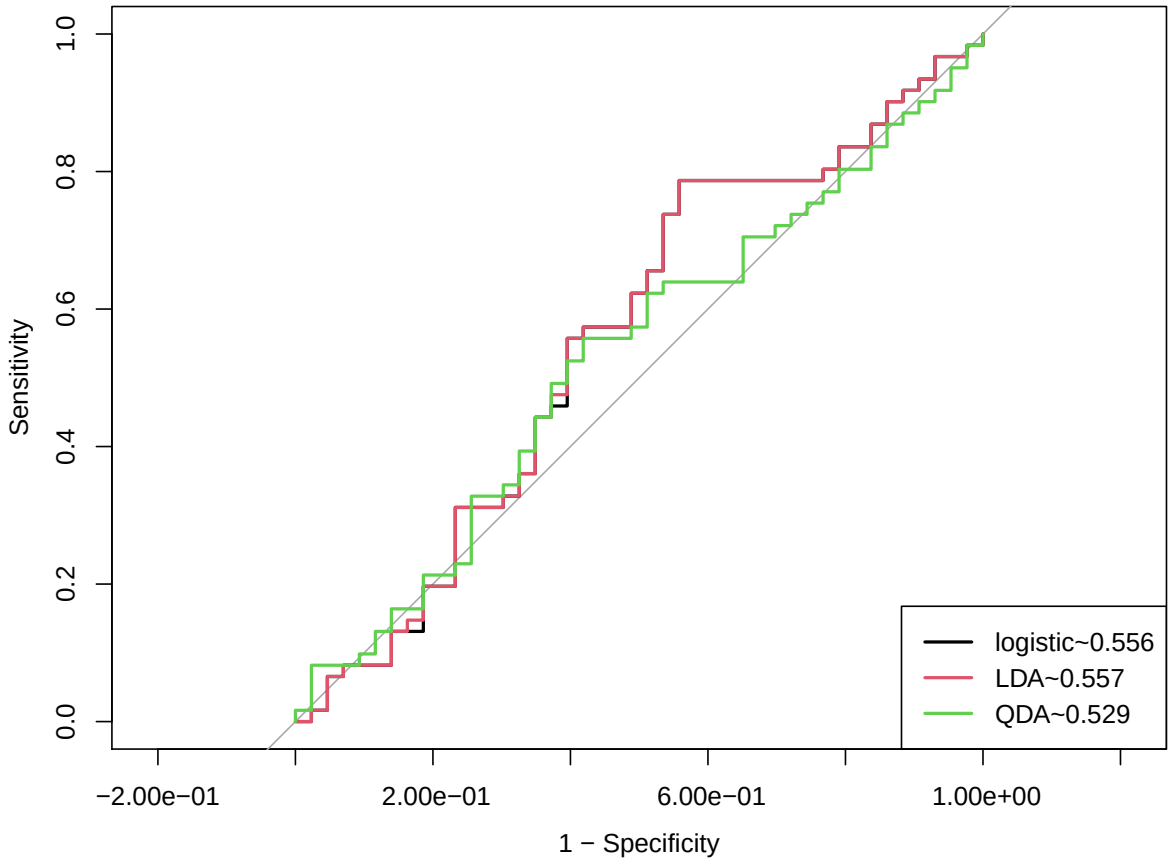
```
Weekly_lda =
  train(model.matrix(direction~lag1+lag2,Weekly_Tr)[,-1],
        Y_tr,
        method = "lda",
        metric = "ROC",
        trControl = TRC,
        preProcess = c("center","scale"))
```

```
lda_roc = pROC::roc(Y_ts,predict(Weekly_lda,newdata = X_ts,type = "prob")[,2])
```

QDA

```
Weekly_qda =
  train(model.matrix(direction~lag1+lag2,Weekly_Tr)[,-1],
        Y_tr,
        method = "qda",
        metric = "ROC",
        trControl = TRC,
        preProcess = c("center","scale"))
```

```
qda_roc = pROC::roc(Y_ts,predict(Weekly_qda,newdata = X_ts,type = "prob")[,2])

auc = c()

ROC = list(logistic_roc,lda_roc,qda_roc)

for (i in 1:3){
  auc = append(auc,ROC[[i]]$auc[1])
  plot(ROC[[i]],col = i, add = T * (i>1), legacy.axes = T * (i==1))
}

model_name =
  c("logistic","LDA","QDA")

legend("bottomright",
       legend = paste0(model_name,"~",round(auc,3)),col=1:3,lwd=2)
```


Sensitivity
1 − Specificity
0.0
0.2
0.4
0.6
0.8
1.0
−2.00e−01
2.00e−01
6.00e−01
1.00e+00
logistic~0.556
LDA~0.557
QDA~0.529