2. Run a LASSO version of logistic regression, using CV to estimate optimal shrinkage in the logistic regression parameters using minimum CV-error.

```
### 2. Run a LASSO version of logistic regression, using CV to estimate optimal shrinkage
# in the logistic regression parameters using minimum CV-error.

# (a) Report a list of the variables included/excluded in each logit. Does
# the pattern seem somewhat consistent with the ANOVA results from
# earlier? Explain in a sentence.
X.train.scale = as.matrix(data.train.scale[,-19])
Y.train = data.train.scale[,19]
X.valid.scale = as.matrix(data.valid.scale[,-19])
Y.valid = data.valid.scale[,19]


### Let's repeat our logistic regression analysis using the glmnet package.
#fit.log.glmnet = cv.glmnet(X.train.scale, Y.train, family = "multinomial")
all.LASSOs = cv.glmnet(X.train.scale, Y.train, family = "multinomial")

lambda.min = all.LASSOs$lambda.min
lambda.1se = all.LASSOs$lambda.1se
### Get predictions and investigate performance. The predict() function
### for glmnet() can give several different types of output. To get
### predicted values, set type="class". Remember to set s=0 for logistic
### regression.
# pred.log.glmnet = predict(fit.log.glmnet, X.valid.scale, type = "class",
#                           s = 0)
pred.log.glmnet.min = predict(fit.log.glmnet, X.valid.scale, type = "class",
                          s = lambda.min)
pred.log.glmnet.1se = predict(fit.log.glmnet, X.valid.scale, type = "class",
                          s = lambda.1se)


# table(Y.valid, pred.log.glmnet, dnn = c("Observed", "Predicted"))
table(Y.valid, pred.log.glmnet.min, dnn = c("Observed", "Predicted"))
table(Y.valid, pred.log.glmnet.1se, dnn = c("Observed", "Predicted"))

(misclass.log.glmnet = mean(Y.valid != pred.log.glmnet.min))
(misclass.log.glmnet = mean(Y.valid != pred.log.glmnet.1se))

lasso.min.coef = coef(all.LASSOs, s="lambda.min",exact=TRUE, family="multinomial")
lasso.1se.coef = coef(all.LASSOs, s="lambda.1se",exact=TRUE, family="multinomial")
```

(a) **Report a list of the variables included/excluded in each logit. Does the pattern seem somewhat consistent with the ANOVA results from earlier? Explain in a sentence.**

```
>lasso.min.coef
(Intercept)                    10.47789955
Compactness                    -6.43735489
Circularity                     4.33621387
Distance.Circularity           -0.28083026
Radius.Ratio                   18.69867552
Pr.Axis.Aspect.Ratio          -28.96879784
Max.Length.Aspect.Ratio        -4.05453435
Scatter.Ratio                   0.05192484
Elongatedness                   3.17196076
Pr.Axis.Rectangularity          .
Max.Length.Rectangularity      -7.57664804
Scaled.Variance.Along.Major.Axis -1.51693002
Scaled.Variance.Along.Minor.Axis 15.67578846
Scaled.Radius.of.Gyration      -3.10578781
Skewness.About.Major.Axis      -14.41921897
Skewness.About.Minor.Axis       0.04195768
```

Kurtosis.About.Minor.Axis	0.20849110
Kurtosis.About.Major.Axis	1.07739497
Hollows.Ratio	-5.46839934

$`4D`
19 x 1 sparse Matrix of class "dgCMatrix"

	1
(Intercept)	13.5155248
Compactness	4.5758051
Circularity	-12.3226861
Distance.Circularity	0.2808303
Radius.Ratio	21.2169082
Pr.Axis.Aspect.Ratio	-28.5843066
Max.Length.Aspect.Ratio	-10.0694529
Scatter.Ratio	.
Elongatedness	-3.1719608
Pr.Axis.Rectangularity	8.0309236
Max.Length.Rectangularity	-0.8662893
Scaled.Variance.Along.Major.Axis	.
Scaled.Variance.Along.Minor.Axis	-10.6377616
Scaled.Radius.of.Gyration	6.2865542
Skewness.About.Major.Axis	-10.7250672
Skewness.About.Minor.Axis	0.1011371
Kurtosis.About.Minor.Axis	-0.2084911
Kurtosis.About.Major.Axis	-12.7160699
Hollows.Ratio	5.4683993

$BUS
19 x 1 sparse Matrix of class "dgCMatrix"

	1
(Intercept)	0.1965590
Compactness	-4.5758051
Circularity	.
Distance.Circularity	-6.5642736
Radius.Ratio	-64.6005083
Pr.Axis.Aspect.Ratio	105.6089591
Max.Length.Aspect.Ratio	4.0545344
Scatter.Ratio	.
Elongatedness	-19.7803722
Pr.Axis.Rectangularity	.
Max.Length.Rectangularity	0.8662893
Scaled.Variance.Along.Major.Axis	46.2307520
Scaled.Variance.Along.Minor.Axis	.
Scaled.Radius.of.Gyration	3.1057878
Skewness.About.Major.Axis	10.7250672
Skewness.About.Minor.Axis	-5.1044260

Kurtosis.About.Minor.Axis	3.1220789
Kurtosis.About.Major.Axis	47.7261462
Hollows.Ratio	-25.2995979

$VAN
19 x 1 sparse Matrix of class "dgCMatrix"

	1
(Intercept)	-24.18998339
Compactness	12.94557435
Circularity	.
Distance.Circularity	21.91993544
Radius.Ratio	-18.69867552
Pr.Axis.Aspect.Ratio	28.58430664
Max.Length.Aspect.Ratio	29.33960573
Scatter.Ratio	-35.12848043
Elongatedness	9.29536204
Pr.Axis.Rectangularity	-19.40725058
Max.Length.Rectangularity	37.68734783
Scaled.Variance.Along.Major.Axis	.
Scaled.Variance.Along.Minor.Axis	.
Scaled.Radius.of.Gyration	-19.04715083
Skewness.About.Major.Axis	35.85145712
Skewness.About.Minor.Axis	-0.04195768
Kurtosis.About.Minor.Axis	-3.22501509
Kurtosis.About.Major.Axis	-1.07739497
Hollows.Ratio	13.67304898

Lasso.1se.coef

(Intercept)	7.49601576
Compactness	-6.19399153
Circularity	1.88339320
Distance.Circularity	-0.01332511
Radius.Ratio	11.93254099
Pr.Axis.Aspect.Ratio	-19.56533341
Max.Length.Aspect.Ratio	-0.06776796
Scatter.Ratio	.
Elongatedness	1.95032211
Pr.Axis.Rectangularity	.
Max.Length.Rectangularity	-4.05991526
Scaled.Variance.Along.Major.Axis	-0.77740303
Scaled.Variance.Along.Minor.Axis	12.21014816
Scaled.Radius.of.Gyration	-2.43221848
Skewness.About.Major.Axis	-11.87408238
Skewness.About.Minor.Axis	0.23303965
Kurtosis.About.Minor.Axis	0.09133776

Kurtosis.About.Major.Axis .
Hollows.Ratio -4.22927907

$`4D`
19 x 1 sparse Matrix of class "dgCMatrix"
1
(Intercept) 8.99540239
Compactness 3.44488105
Circularity -11.36835569
Distance.Circularity 0.01332511
Radius.Ratio 14.84922077
Pr.Axis.Aspect.Ratio -20.45341274
Max.Length.Aspect.Ratio -5.35577614
Scatter.Ratio .
Elongatedness -1.95032211
Pr.Axis.Rectangularity 4.10408335
Max.Length.Rectangularity -0.30553560
Scaled.Variance.Along.Major.Axis .
Scaled.Variance.Along.Minor.Axis -4.71025706
Scaled.Radius.of.Gyration 5.98296974
Skewness.About.Major.Axis -9.00831800
Skewness.About.Minor.Axis 0.33918807
Kurtosis.About.Minor.Axis -0.09133776
Kurtosis.About.Major.Axis -10.81017188
Hollows.Ratio 4.22927907

$BUS
19 x 1 sparse Matrix of class "dgCMatrix"
1
(Intercept) 0.73796804
Compactness -3.44488105
Circularity .
Distance.Circularity -3.91462628
Radius.Ratio -45.29986372
Pr.Axis.Aspect.Ratio 73.72073905
Max.Length.Aspect.Ratio 0.06776796
Scatter.Ratio .
Elongatedness -14.95283732
Pr.Axis.Rectangularity .
Max.Length.Rectangularity 0.30553560
Scaled.Variance.Along.Major.Axis 30.29959732
Scaled.Variance.Along.Minor.Axis .
Scaled.Radius.of.Gyration 2.43221848
Skewness.About.Major.Axis 9.00831800
Skewness.About.Minor.Axis -3.78573575
Kurtosis.About.Minor.Axis 1.76934625

Kurtosis.About.Major.Axis 32.16611869
Hollows.Ratio -17.62158854

$VAN
19 x 1 sparse Matrix of class "dgCMatrix"
1
(Intercept) -17.2293862
Compactness 9.3044324
Circularity .
Distance.Circularity 13.1945727
Radius.Ratio -11.9325410
Pr.Axis.Aspect.Ratio 19.5653334
Max.Length.Aspect.Ratio 18.3302287
Scatter.Ratio -23.2613542
Elongatedness 7.8143496
Pr.Axis.Rectangularity -8.5801997
Max.Length.Rectangularity 26.7657963
Scaled.Variance.Along.Major.Axis .
Scaled.Variance.Along.Minor.Axis .
Scaled.Radius.of.Gyration -13.5295229
Skewness.About.Major.Axis 21.9994724
Skewness.About.Minor.Axis -0.2330397
Kurtosis.About.Minor.Axis -2.3249012
Kurtosis.About.Major.Axis .
Hollows.Ratio 7.3440734

The variable with numerical values (not just dot) looks like variables in a list from ANOVA

(b) Compute and **report training and test error**. **How does test error compare to other methods?**

```
> (miss.lasso.min = mean(Y.valid != pred.lasso.min))
[1] 0.2169811
> (miss.lasso.1se = mean(Y.valid != pred.lasso.1se))
[1] 0.2122642
```

-> The error is lower than other methods.