Part3.

Code:

```
spam.dat<-read.table('spam.data.txt',sep=' ')
spam.dat$V58<-as.factor(spam.dat$V58)
#Part a
tree.model1<-tree(V58~.,data=spam.dat,mindev=0.0008)
cv.model <- cv.tree(tree.model1)
plot(cv.model)
size=110
size110tree<-prune.tree(tree.model1, best=size)
plot(size110tree)
plot(size110tree)
text(size110tree,cex=0.5)
summary(size110tree)
# Misclassification error rate: 0.03478 = 160 / 4600

#Part b
tree.model2<-tree(V58~.,data=spam.dat,mindev=0.005)
cv.model <- cv.tree(tree.model2)
plot(cv.model)
size=20
size20tree<-prune.tree(tree.model2, best=size)
plot(size20tree)
plot(size20tree)
text(size20tree,cex=0.5)
summary(size20tree)
# Misclassification error rate: 0.07826 = 360 / 4600

#Part c
tree.model3<-tree(V58~.,data=spam.dat,mindev=0.006)
cv.model <- cv.tree(tree.model3)
plot(cv.model)
best.size <- cv.model$size[which(cv.model$dev==min(cv.model$dev))]
best.size #[1] 18 17 16 15 14 13
bestsizetree <- prune.tree(tree.model3, best=best.size)
plot(bestsizetree)
text(bestsizetree,cex=0.8)
summary(bestsizetree)#number of terminal nodes is 13.
# Misclassification error rate: 0.08261 = 380 / 4600
```

Part 4.

Code:

```
library(boot)
spam.dat<-read.table('spam.data.txt',sep=' ')
spam.dat$V58<-as.factor(spam.dat$V58)
LR <- glm(V58~., family='binomial', data=spam.dat)
summary(LR)
cvLR<-cv.glm(spam.dat,LR,K=10)

#The cross validation error is:
cvLR[[3]]
#[1] 0.05848635 0.05832672
```

Result of Logistic Regression:

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.127	-0.203	0.000	0.114	5.364

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-1.569e+00	1.420e-01	-11.044	< 2e-16	***
V1	-3.895e-01	2.315e-01	-1.683	0.092388	.
V2	-1.458e-01	6.928e-02	-2.104	0.035362	*
V3	1.141e-01	1.103e-01	1.035	0.300759	
V4	2.252e+00	1.507e+00	1.494	0.135168	
V5	5.624e-01	1.018e-01	5.524	3.31e-08	***
V6	8.830e-01	2.498e-01	3.534	0.000409	***
V7	2.279e+00	3.328e-01	6.846	7.57e-12	***
V8	5.696e-01	1.682e-01	3.387	0.000707	***
V9	7.343e-01	2.849e-01	2.577	0.009958	**
V10	1.275e-01	7.262e-02	1.755	0.079230	.
V11	-2.557e-01	2.979e-01	-0.858	0.390655	
V12	-1.383e-01	7.405e-02	-1.868	0.061773	.
V13	-7.961e-02	2.303e-01	-0.346	0.729557	
V14	1.447e-01	1.364e-01	1.061	0.288855	
V15	1.236e+00	7.254e-01	1.704	0.088370	.
V16	1.039e+00	1.457e-01	7.128	1.01e-12	***
V17	9.599e-01	2.251e-01	4.264	2.01e-05	***
V18	1.203e-01	1.172e-01	1.027	0.304533	
V19	8.131e-02	3.505e-02	2.320	0.020334	*
V20	1.047e+00	5.383e-01	1.946	0.051675	.
V21	2.419e-01	5.243e-02	4.615	3.94e-06	***
V22	2.013e-01	1.627e-01	1.238	0.215838	
V23	2.245e+00	4.714e-01	4.762	1.91e-06	***
V24	4.264e-01	1.621e-01	2.630	0.008535	**
V25	-1.920e+00	3.128e-01	-6.139	8.31e-10	***
V26	-1.040e+00	4.396e-01	-2.366	0.017966	*
V27	-1.177e+01	2.113e+00	-5.569	2.57e-08	***
V28	4.454e-01	1.991e-01	2.237	0.025255	*
V29	-2.486e+00	1.502e+00	-1.656	0.097744	.
V30	-3.299e-01	3.137e-01	-1.052	0.292972	
V31	-1.702e-01	4.815e-01	-0.353	0.723742	
V32	2.549e+00	3.283e+00	0.776	0.437566	
V33	-7.383e-01	3.117e-01	-2.369	0.017842	*
V34	6.679e-01	1.601e+00	0.417	0.676490	
V35	-2.055e+00	7.883e-01	-2.607	0.009124	**
V36	9.237e-01	3.091e-01	2.989	0.002803	**
V37	4.651e-02	1.754e-01	0.265	0.790819	
V38	-5.968e-01	4.232e-01	-1.410	0.158473	
V39	-8.650e-01	3.828e-01	-2.260	0.023844	*
V40	-3.046e-01	3.636e-01	-0.838	0.402215	
V41	-4.505e+01	2.660e+01	-1.694	0.090333	.
V42	-2.689e+00	8.384e-01	-3.207	0.001342	**
V43	-1.247e+00	8.064e-01	-1.547	0.121978	
V44	-1.573e+00	5.292e-01	-2.973	0.002953	**
V45	-7.923e-01	1.556e-01	-5.091	3.56e-07	***

```
V46        -1.459e+00  2.686e-01  -5.434 5.52e-08 ***
V47        -2.326e+00  1.659e+00  -1.402 0.160958
V48        -4.016e+00  1.611e+00  -2.493 0.012672 *
V49        -1.291e+00  4.422e-01  -2.920 0.003503 **
V50        -1.881e-01  2.494e-01  -0.754 0.450663
V51        -6.574e-01  8.383e-01  -0.784 0.432914
V52         3.472e-01  8.926e-02   3.890 0.000100 ***
V53         5.336e+00  7.064e-01   7.553 4.24e-14 ***
V54         2.403e+00  1.113e+00   2.159 0.030883 *
V55         1.199e-02  1.884e-02   0.636 0.524509
V56         9.118e-03  2.521e-03   3.618 0.000297 ***
V57         8.437e-04  2.251e-04   3.747 0.000179 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6170.2  on 4600  degrees of freedom
Residual deviance: 1815.8  on 4543  degrees of freedom
AIC: 1931.8

Number of Fisher Scoring iterations: 13
```

Part5.

Code:

```
library(glmnet)
spam.dat<-read.table('spam.data.txt',sep=' ')
y<-as.factor(spam.dat$V58)
X<-model.matrix(~.,spam.dat[,1:57])
cv.myNewlogit <- cv.glmnet(X,y,alpha=1,nfolds=10,family='binomial')
plot(cv.myNewlogit,xvar="lambda")
cv.myNewlogit$lambda.min #=0.0004034505
```