

Homework 1

STAT6306; Due: 09/05/2017

Problem 0

R is a standard software interface for computing and graphics and Rstudio is an integrated development environment (IDE) for R. Install both on your computer.

- R: <http://lib.stat.cmu.edu/R/CRAN/>
- Rstudio: <https://www.rstudio.com/products/rstudio/#Desktop>

Problem 1

Suppose we have the following matrix:

```
set.seed(1)
A = matrix(rnorm(4*3),nrow=4,ncol=3)
```

We want to get the column mean for each column of the matrix A . Do this using each of the following techniques:

Part a

Hard coding (that is, write $(A[1,1] + A[2,1] + \dots)/4, \dots$)

```
#SOLUTION
c((A[1,1]+A[2,1]+A[3,1]+A[4,1])/4, (A[1,2]+A[2,2]+A[3,2]+A[4,2])/4, (A[1,3]+A[2,3]+A[3,3]+A[4,3])/4)

## [1] 0.07921043 0.18369829 0.54300434
```

Part b

For loop(s)

```
#SOLUTION
col_mean = rep(0,ncol(A))
for(j in 1:ncol(A)){
  col_mean[j] = mean(A[,j])
}
col_mean
```

```
## [1] 0.07921043 0.18369829 0.54300434
```

Part c

The apply (or related) function

```
#SOLUTION
apply_mean = apply(A, 2, mean)
apply_mean
```

```
## [1] 0.07921043 0.18369829 0.54300434
```

Problem 2

Many statistical methods can be computed/analyzed using the SVD¹. Let's look at solving least squares problems as they are fundamental to modern data analysis.

Part a

```
set.seed(10)
A = matrix(rnorm(24),nrow=6,ncol=4)
A[,1] = 1
```

Write $A = UDV^\top$ (that is, form `svd.out = svd(A)`).

Suppose we wish to solve for $\hat{x} = \arg \min_x \|Ax - b\|_2^2 = (A^\top A)^{-1} A^\top b$ for $b = (1, 2, 3, 4, 5, 6)^\top$. As an aside, to show this, note that²

$$\|Ax - b\|_2^2 = x^\top A^\top Ax + b^\top b - 2x^\top A^\top b \quad (1)$$

$$\Rightarrow \nabla_x = 2A^\top A\hat{x} - 2A^\top b \stackrel{\text{set}}{=} 0 \quad (2)$$

$$\Rightarrow \hat{x} = (A^\top A)^{-1} A^\top b \quad (3)$$

How can I solve this using the SVD? Here, let's follow the steps:

1. Form $U^\top b$
2. Solve $Dw = U^\top b$
3. Form $\hat{x} = Vw$

Produce this \hat{x} in R via this method. Note that in this particular case, all the singular values in D are nonzero and hence $\hat{x} = VD^{-1}U^\top b$.

```
#SOLUTION
svd.out = svd(A)
b = 1:6
Ut_b = t(svd.out$u)%*%b
w = solve(diag(svd.out$d), Ut_b)
x_hat = svd.out$v%*%w
x_hat
```

```
##           [,1]
## [1,]  3.1049466
## [2,]  0.8150028
## [3,] -0.1257175
## [4,] -0.8942109
```

¹For this question, I'm using common linear algebra notation of A , x , and b . This x is not to be confused with a feature

² ∇_x indicates gradient with respect to x

Part b

Suppose instead we have observations under the model $Y = \mathbb{X}\beta + \epsilon$, where $Y = b$ and $\mathbb{X} = A$. Using the R function `lm` and `predict`, what is the least squares solution $\hat{\beta}$ and the fitted values \hat{Y} for Y using the least squares solution?³

How does the produced coefficient vector $\hat{\beta}$ compare the \hat{x} ?

#SOLUTION

```
lm_out = lm(b~A-1)
beta_hat = lm_out$coefficients
beta_hat
```

```
##           A1           A2           A3           A4
##  3.1049466  0.8150028 -0.1257175 -0.8942109
```

```
Y_hat = predict(lm_out)
Y_hat
```

```
##           1           2           3           4           5           6
##  1.322700  2.252526  2.219226  4.838791  4.726426  5.640332
```

SOLUTION They are the same solution

Problem 3

Now, let's look at a new A

```
set.seed(100)
A = matrix(rnorm(4*3),ncol=4,nrow=3)
A[,1] = 1
```

and $b = (1, 2, 3)^\top$. This is an example of an *underdetermined* system.

Part a

What do(es) the corresponding \hat{x} look like using the SVD? What do(es) the $\hat{\beta}$ look like using `lm`?

#SOLUTION

```
svd.out = svd(A)
b = 1:3
Ut_b = t(svd.out$u)%*%b
w = solve(diag(svd.out$d), Ut_b)
x_hat = svd.out$v%*%w
x_hat
```

```
##           [,1]
## [1,]  2.6170341
## [2,] -1.5880142
## [3,] -0.8493085
## [4,]  1.9533232
```

```
lm_out = lm(b~A-1)
beta_hat = lm_out$coefficients
beta_hat
```

³Remember to not have R add an intercept as there is already a column of ones

```
##           A1           A2           A3           A4
##  3.110348 -3.069596 -1.051448           NA
```

Part b

What do(es) the corresponding $A\hat{x}$ look like using the SVD? What do(es) the $\hat{Y} = \mathbb{X}\hat{\beta}$ look like using predict?

#SOLUTION

```
A%*%x_hat
```

```
##           [,1]
## [1,]         1
## [2,]         2
## [3,]         3
```

```
predict(lm_out)
```

```
## 1 2 3
## 1 2 3
```

NOTE: it is worth considering why the two objects have different formatting.

Part c

Though this is just one simulated example and not a proof, your findings generalize to all situations when $p > n$. Summarize in words what these findings are.

SOLUTION: When $p > n$ the model is overfit and therefore the response values can be predicted exactly. However, this is not good practice because the model will have high variance and have poor predictions for new responses.

Problem 4

```
set.seed(1)
n = 2000
p = 500
X = matrix(rnorm(n*p),nrow=n,ncol=p)
X[,1] = 1
format(object.size(X),units='auto') #memory used by X
```

```
## [1] "7.6 Mb"
```

```
b = rep(0,p)
b[1:5] = 25
b_0 = 0
Xdf = data.frame(X)
Y = b_0 + X %*% b + rnorm(n)
hatBeta = coef(lm(Y~X-1)) #Here, the [-1] ignores the intercept
```

#Using out-of-core technique

```
write.table(X[1:500,],file='Xchunk1.txt',sep=',',row.names=F,col.names=names(Xdf))
write.table(X[501:1000,],file='Xchunk2.txt',sep=',',row.names=F,col.names=names(Xdf))
write.table(X[1001:1500,],file='Xchunk3.txt',sep=',',row.names=F,col.names=names(Xdf))
write.table(X[1501:2000,],file='Xchunk4.txt',sep=',',row.names=F,col.names=names(Xdf))
```

```
write.table(Y[1:500],file='Ychunk1.txt',sep=',',row.names=F,col.names=F)
write.table(Y[501:1000],file='Ychunk2.txt',sep=',',row.names=F,col.names=F)
write.table(Y[1001:1500],file='Ychunk3.txt',sep=',',row.names=F,col.names=F)
write.table(Y[1501:2000],file='Ychunk4.txt',sep=',',row.names=F,col.names=F)
```

Part a

Report the first 5 entries in $\hat{\beta}$ (that is, `hatBeta` in the above code) using `lm` on all the data simultaneously

#SOLUTION

```
print(hatBeta[1:5])
```

```
##          X1          X2          X3          X4          X5
## 24.99160 24.98724 24.96734 25.05268 25.04096
```

Part b

Alternatively, we can read in each chunk and update the solution using `biglm`. Here is the first part. Complete the procedure in the natural way on the remaining chunks. Compare the first 5 entries in $\hat{\beta}$ formed by this method with the entries in (a)

```
if(!require(biglm,quietly=TRUE)){
  install.packages('biglm',repos='http://cran.us.r-project.org');require(biglm)
}

# Chunk 1
Xchunk = read.table(file='Xchunk1.txt',sep=',',header=T)
Ychunk = scan(file='Ychunk1.txt',sep=',')
form = as.formula(paste('Ychunk ~ -1 + ',paste(names(Xchunk),collapse=' + '),collapse=''))
out.biglm = biglm(formula = form,data=Xchunk)
hatBeta[1:5]
```

```
##          X1          X2          X3          X4          X5
## 24.99160 24.98724 24.96734 25.05268 25.04096
```

```
coef(out.biglm)[1:5]
```

```
##          X1          X2          X3          X4          X5
## 25.08815 24.48229 26.09057 26.84305 24.64633
```

```
# Chunk 2
Xchunk = read.table(file='Xchunk2.txt',sep=',',header=T)
Ychunk = scan(file='Ychunk2.txt',sep=',')
out.biglm = update(out.biglm,moredata=Xchunk)
hatBeta[1:5]
```

```
##          X1          X2          X3          X4          X5
## 24.99160 24.98724 24.96734 25.05268 25.04096
```

```
coef(out.biglm)[1:5]
```

```
##          X1          X2          X3          X4          X5
## 24.96665 25.00382 24.97887 25.01876 25.08741
```

```
# Chunk 3
Xchunk = read.table(file='Xchunk3.txt',sep=',',header=T)
```

```
Ychunk = scan(file='Ychunk3.txt',sep=',')
out.biglm = update(out.biglm,moredata=Xchunk)
hatBeta[1:5]
```

```
##      X1      X2      X3      X4      X5
## 24.99160 24.98724 24.96734 25.05268 25.04096
```

```
coef(out.biglm)[1:5]
```

```
##      X1      X2      X3      X4      X5
## 24.98077 25.00492 24.97527 25.02329 25.08050
```

Can you figure out the final steps? Have we updated on all of the chunks?

#Solution

Chunk 4

```
Xchunk = read.table(file='Xchunk4.txt',sep=',',header=T)
Ychunk = scan(file='Ychunk4.txt',sep=',')
out.biglm = update(out.biglm,moredata=Xchunk)
```

```
print(hatBeta[1:5])
```

```
##      X1      X2      X3      X4      X5
## 24.99160 24.98724 24.96734 25.05268 25.04096
```

```
print(coef(out.biglm)[1:5])
```

```
##      X1      X2      X3      X4      X5
## 24.99160 24.98724 24.96734 25.05268 25.04096
```

Yes all chunks updated because we see the two solutions match

Problem 5

Forward selection.

Part a

Using the \mathbb{X} and Y generated in the previous problem, use forward selection and AIC to estimate b .

```
if(!require(leaps)){install.packages('leaps',repos='http://cran.us.r-project.org');require(leaps)}

## Loading required package: leaps

outFor      = regsubsets(x=X,y=Y,nvmax=p,method='forward',intercept=FALSE)
sumFor      = summary(outFor)
modelFor    = sumFor$which[which.min(sumFor$cp),]
leapsModel  = as.numeric(which(modelFor))
```

Part b (optional)

Save the \mathbb{X} generated in the previous problem to a .csv file. Using forward selection and AIC, estimate b without having \mathbb{X} stored in memory. Verify that your answer matches (a)

```

write.csv(x=X,file='featureMat.csv')
p          = ncol(X)
n          = nrow(X)
sigmaSq    = NULL#Try sigmaSq = out.biglm$qr$ss/(n-p)
gicType    = 'AIC'
outOfCore  = FALSE### To do forward selection out of core, will be slow

GICf = function(ind,gicType = 'AIC', sigmaSq = NULL,outOfCore = FALSE){
  if(outOfCore){
    grabVec = rep('NULL',p)
    grabVec[ind] = NA
    featureMat = read.csv('featureMat.csv', colClasses=grabVec)
    lm.out = lm(Y~.-1,data=featureMat)
  }else{
    lm.out = lm(Y~X[,ind]-1)
  }
  if(gicType == 'AIC'){
    scaleTerm = 2
  }else if(gicType == 'BIC'){
    scaleTerm = log(n)
  }else{stop('Only supports AIC or BIC')}

  if(!is.null(sigmaSq)){
    if(class(sigmaSq) != class(1) | sigmaSq < 0){stop('Invalid variance estimate')}
    return(sum(lm.out$residuals**2)/n + scaleTerm/n * length(ind)*sigmaSq )
  }else{
    return(n*log(sum(lm.out$residuals**2)/n) + scaleTerm * length(ind) )
  }
}

GIC          = Inf#initialize
indSelect    = c(1)#initialize
indSet       = 2:p#initialize
importantVar  = 0#initialize
addedNewVar  = FALSE#initialize

repeat{
  cat('We have selected thus far: ',indSelect,'\n')
  countFeatures = 0
  indSetSweep = 0#this gets the index in indSet of importantVar
  for(j in indSet){
    indSetSweep = indSetSweep + 1
    countFeatures = countFeatures + 1
    if(countFeatures %% round(length(indSet)/5) == 0){
      cat('We have looked at the first: ', countFeatures/length(indSet),' fraction of features \n')
    }
    indTmp = c(indSelect,j)
    GICnew = GICf(indTmp, gicType = gicType,
                  sigmaSq = sigmaSq,outOfCore = outOfCore)
    if(GICnew < GIC){
      GIC = GICnew
      importantVar      = j
    }
  }
}

```

```

        importantVarIndex = indSetSweep
        addedNewVar       = TRUE
    }

}
if(!addedNewVar){
    break
}else{
    indSet      = indSet[-importantVarIndex]
    indSelect = c(indSelect,importantVar)
}
print(GIC)
addedNewVar = FALSE
}

```

```

## We have selected thus far:  1
## We have looked at the first:  0.2004008  fraction of features
## We have looked at the first:  0.4008016  fraction of features
## We have looked at the first:  0.6012024  fraction of features
## We have looked at the first:  0.8016032  fraction of features
## [1] 14910.78
## We have selected thus far:  1 4
## We have looked at the first:  0.2008032  fraction of features
## We have looked at the first:  0.4016064  fraction of features
## We have looked at the first:  0.6024096  fraction of features
## We have looked at the first:  0.8032129  fraction of features
## [1] 14131.85
## We have selected thus far:  1 4 5
## We have looked at the first:  0.1991952  fraction of features
## We have looked at the first:  0.3983903  fraction of features
## We have looked at the first:  0.5975855  fraction of features
## We have looked at the first:  0.7967807  fraction of features
## We have looked at the first:  0.9959759  fraction of features
## [1] 12802.48
## We have selected thus far:  1 4 5 2
## We have looked at the first:  0.1995968  fraction of features
## We have looked at the first:  0.3991935  fraction of features
## We have looked at the first:  0.5987903  fraction of features
## We have looked at the first:  0.7983871  fraction of features
## We have looked at the first:  0.9979839  fraction of features
## [1] -50.2208
## We have selected thus far:  1 4 5 2 3
## We have looked at the first:  0.2  fraction of features
## We have looked at the first:  0.4  fraction of features
## We have looked at the first:  0.6  fraction of features
## We have looked at the first:  0.8  fraction of features
## We have looked at the first:  1  fraction of features
## [1] -56.59783
## We have selected thus far:  1 4 5 2 3 231
## We have looked at the first:  0.2004049  fraction of features
## We have looked at the first:  0.4008097  fraction of features
## We have looked at the first:  0.6012146  fraction of features
## We have looked at the first:  0.8016194  fraction of features
## [1] -62.0669

```



```

## We have selected thus far: 1 4 5 2 3 231 256
## We have looked at the first: 0.2008114 fraction of features
## We have looked at the first: 0.4016227 fraction of features
## We have looked at the first: 0.6024341 fraction of features
## We have looked at the first: 0.8032454 fraction of features
## [1] -67.40517
## We have selected thus far: 1 4 5 2 3 231 256 257
## We have looked at the first: 0.199187 fraction of features
## We have looked at the first: 0.398374 fraction of features
## We have looked at the first: 0.597561 fraction of features
## We have looked at the first: 0.796748 fraction of features
## We have looked at the first: 0.995935 fraction of features
## [1] -72.60968
## We have selected thus far: 1 4 5 2 3 231 256 257 278
## We have looked at the first: 0.1995927 fraction of features
## We have looked at the first: 0.3991853 fraction of features
## We have looked at the first: 0.598778 fraction of features
## We have looked at the first: 0.7983707 fraction of features
## We have looked at the first: 0.9979633 fraction of features
## [1] -77.69909
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152
## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features
## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -82.53152
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72
## We have looked at the first: 0.200409 fraction of features
## We have looked at the first: 0.400818 fraction of features
## We have looked at the first: 0.601227 fraction of features
## We have looked at the first: 0.801636 fraction of features
## [1] -87.65492
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88
## We have looked at the first: 0.2008197 fraction of features
## We have looked at the first: 0.4016393 fraction of features
## We have looked at the first: 0.602459 fraction of features
## We have looked at the first: 0.8032787 fraction of features
## [1] -91.77004
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390
## We have looked at the first: 0.1991786 fraction of features
## We have looked at the first: 0.3983573 fraction of features
## We have looked at the first: 0.5975359 fraction of features
## We have looked at the first: 0.7967146 fraction of features
## We have looked at the first: 0.9958932 fraction of features
## [1] -96.20694
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56
## We have looked at the first: 0.1995885 fraction of features
## We have looked at the first: 0.399177 fraction of features
## We have looked at the first: 0.5987654 fraction of features
## We have looked at the first: 0.7983539 fraction of features
## We have looked at the first: 0.9979424 fraction of features
## [1] -100.4436
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380

```

```

## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features
## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -104.5091
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370
## We have looked at the first: 0.2004132 fraction of features
## We have looked at the first: 0.4008264 fraction of features
## We have looked at the first: 0.6012397 fraction of features
## We have looked at the first: 0.8016529 fraction of features
## [1] -108.2332
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46
## We have looked at the first: 0.2008282 fraction of features
## We have looked at the first: 0.4016563 fraction of features
## We have looked at the first: 0.6024845 fraction of features
## We have looked at the first: 0.8033126 fraction of features
## [1] -112.087
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31
## We have looked at the first: 0.1991701 fraction of features
## We have looked at the first: 0.3983402 fraction of features
## We have looked at the first: 0.5975104 fraction of features
## We have looked at the first: 0.7966805 fraction of features
## We have looked at the first: 0.9958506 fraction of features
## [1] -115.6395
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314
## We have looked at the first: 0.1995842 fraction of features
## We have looked at the first: 0.3991684 fraction of features
## We have looked at the first: 0.5987526 fraction of features
## We have looked at the first: 0.7983368 fraction of features
## We have looked at the first: 0.997921 fraction of features
## [1] -119.4165
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451
## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features
## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -123.1261
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276
## We have looked at the first: 0.2004175 fraction of features
## We have looked at the first: 0.4008351 fraction of features
## We have looked at the first: 0.6012526 fraction of features
## We have looked at the first: 0.8016701 fraction of features
## [1] -126.2456
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2008368 fraction of features
## We have looked at the first: 0.4016736 fraction of features
## We have looked at the first: 0.6025105 fraction of features
## We have looked at the first: 0.8033473 fraction of features
## [1] -129.0992
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1991614 fraction of features
## We have looked at the first: 0.3983229 fraction of features

```

```

## We have looked at the first: 0.5974843 fraction of features
## We have looked at the first: 0.7966457 fraction of features
## We have looked at the first: 0.9958071 fraction of features
## [1] -132.062
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1995798 fraction of features
## We have looked at the first: 0.3991597 fraction of features
## We have looked at the first: 0.5987395 fraction of features
## We have looked at the first: 0.7983193 fraction of features
## We have looked at the first: 0.9978992 fraction of features
## [1] -134.9298
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features
## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -137.4954
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2004219 fraction of features
## We have looked at the first: 0.4008439 fraction of features
## We have looked at the first: 0.6012658 fraction of features
## We have looked at the first: 0.8016878 fraction of features
## [1] -139.6534
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2008457 fraction of features
## We have looked at the first: 0.4016913 fraction of features
## We have looked at the first: 0.602537 fraction of features
## We have looked at the first: 0.8033827 fraction of features
## [1] -141.8808
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1991525 fraction of features
## We have looked at the first: 0.3983051 fraction of features
## We have looked at the first: 0.5974576 fraction of features
## We have looked at the first: 0.7966102 fraction of features
## We have looked at the first: 0.9957627 fraction of features
## [1] -144.1929
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1995754 fraction of features
## We have looked at the first: 0.3991507 fraction of features
## We have looked at the first: 0.5987261 fraction of features
## We have looked at the first: 0.7983015 fraction of features
## We have looked at the first: 0.9978769 fraction of features
## [1] -146.5816
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features
## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -148.536
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2004264 fraction of features
## We have looked at the first: 0.4008529 fraction of features

```

```

## We have looked at the first: 0.6012793 fraction of features
## We have looked at the first: 0.8017058 fraction of features
## [1] -150.5055
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2008547 fraction of features
## We have looked at the first: 0.4017094 fraction of features
## We have looked at the first: 0.6025641 fraction of features
## We have looked at the first: 0.8034188 fraction of features
## [1] -152.5564
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1991435 fraction of features
## We have looked at the first: 0.3982869 fraction of features
## We have looked at the first: 0.5974304 fraction of features
## We have looked at the first: 0.7965739 fraction of features
## We have looked at the first: 0.9957173 fraction of features
## [1] -154.3012
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1995708 fraction of features
## We have looked at the first: 0.3991416 fraction of features
## We have looked at the first: 0.5987124 fraction of features
## We have looked at the first: 0.7982833 fraction of features
## We have looked at the first: 0.9978541 fraction of features
## [1] -155.8459
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features
## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -157.4925
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.200431 fraction of features
## We have looked at the first: 0.4008621 fraction of features
## We have looked at the first: 0.6012931 fraction of features
## We have looked at the first: 0.8017241 fraction of features
## [1] -158.9826
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2008639 fraction of features
## We have looked at the first: 0.4017279 fraction of features
## We have looked at the first: 0.6025918 fraction of features
## We have looked at the first: 0.8034557 fraction of features
## [1] -160.6064
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1991342 fraction of features
## We have looked at the first: 0.3982684 fraction of features
## We have looked at the first: 0.5974026 fraction of features
## We have looked at the first: 0.7965368 fraction of features
## We have looked at the first: 0.995671 fraction of features
## [1] -162.1046
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1995662 fraction of features
## We have looked at the first: 0.3991323 fraction of features
## We have looked at the first: 0.5986985 fraction of features
## We have looked at the first: 0.7982646 fraction of features

```

```

## We have looked at the first: 0.9978308 fraction of features
## [1] -163.5175
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features
## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -164.9286
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2004357 fraction of features
## We have looked at the first: 0.4008715 fraction of features
## We have looked at the first: 0.6013072 fraction of features
## We have looked at the first: 0.8017429 fraction of features
## [1] -166.3139
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2008734 fraction of features
## We have looked at the first: 0.4017467 fraction of features
## We have looked at the first: 0.6026201 fraction of features
## We have looked at the first: 0.8034934 fraction of features
## [1] -167.63
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1991247 fraction of features
## We have looked at the first: 0.3982495 fraction of features
## We have looked at the first: 0.5973742 fraction of features
## We have looked at the first: 0.7964989 fraction of features
## We have looked at the first: 0.9956236 fraction of features
## [1] -168.9448
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1995614 fraction of features
## We have looked at the first: 0.3991228 fraction of features
## We have looked at the first: 0.5986842 fraction of features
## We have looked at the first: 0.7982456 fraction of features
## We have looked at the first: 0.997807 fraction of features
## [1] -170.1037
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features
## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -171.2624
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2004405 fraction of features
## We have looked at the first: 0.4008811 fraction of features
## We have looked at the first: 0.6013216 fraction of features
## We have looked at the first: 0.8017621 fraction of features
## [1] -172.2749
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.200883 fraction of features
## We have looked at the first: 0.401766 fraction of features
## We have looked at the first: 0.602649 fraction of features
## We have looked at the first: 0.803532 fraction of features
## [1] -173.4132

```

```

## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.199115 fraction of features
## We have looked at the first: 0.3982301 fraction of features
## We have looked at the first: 0.5973451 fraction of features
## We have looked at the first: 0.7964602 fraction of features
## We have looked at the first: 0.9955752 fraction of features
## [1] -174.3076
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1995565 fraction of features
## We have looked at the first: 0.3991131 fraction of features
## We have looked at the first: 0.5986696 fraction of features
## We have looked at the first: 0.7982262 fraction of features
## We have looked at the first: 0.9977827 fraction of features
## [1] -175.2412
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features
## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -176.2701
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2004454 fraction of features
## We have looked at the first: 0.4008909 fraction of features
## We have looked at the first: 0.6013363 fraction of features
## We have looked at the first: 0.8017817 fraction of features
## [1] -177.3429
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2008929 fraction of features
## We have looked at the first: 0.4017857 fraction of features
## We have looked at the first: 0.6026786 fraction of features
## We have looked at the first: 0.8035714 fraction of features
## [1] -178.4328
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1991051 fraction of features
## We have looked at the first: 0.3982103 fraction of features
## We have looked at the first: 0.5973154 fraction of features
## We have looked at the first: 0.7964206 fraction of features
## We have looked at the first: 0.9955257 fraction of features
## [1] -179.4216
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1995516 fraction of features
## We have looked at the first: 0.3991031 fraction of features
## We have looked at the first: 0.5986547 fraction of features
## We have looked at the first: 0.7982063 fraction of features
## We have looked at the first: 0.9977578 fraction of features
## [1] -180.2969
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features
## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -181.2114

```

```

## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2004505 fraction of features
## We have looked at the first: 0.4009009 fraction of features
## We have looked at the first: 0.6013514 fraction of features
## We have looked at the first: 0.8018018 fraction of features
## [1] -182.0113
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2009029 fraction of features
## We have looked at the first: 0.4018059 fraction of features
## We have looked at the first: 0.6027088 fraction of features
## We have looked at the first: 0.8036117 fraction of features
## [1] -182.8712
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.199095 fraction of features
## We have looked at the first: 0.39819 fraction of features
## We have looked at the first: 0.5972851 fraction of features
## We have looked at the first: 0.7963801 fraction of features
## We have looked at the first: 0.9954751 fraction of features
## [1] -183.5727
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1995465 fraction of features
## We have looked at the first: 0.399093 fraction of features
## We have looked at the first: 0.5986395 fraction of features
## We have looked at the first: 0.7981859 fraction of features
## We have looked at the first: 0.9977324 fraction of features
## [1] -184.1086
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features
## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -184.5495
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2004556 fraction of features
## We have looked at the first: 0.4009112 fraction of features
## We have looked at the first: 0.6013667 fraction of features
## We have looked at the first: 0.8018223 fraction of features
## [1] -184.9314
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2009132 fraction of features
## We have looked at the first: 0.4018265 fraction of features
## We have looked at the first: 0.6027397 fraction of features
## We have looked at the first: 0.803653 fraction of features
## [1] -185.3189
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1990847 fraction of features
## We have looked at the first: 0.3981693 fraction of features
## We have looked at the first: 0.597254 fraction of features
## We have looked at the first: 0.7963387 fraction of features
## We have looked at the first: 0.9954233 fraction of features
## [1] -185.933
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1995413 fraction of features

```

```

## We have looked at the first: 0.3990826 fraction of features
## We have looked at the first: 0.5986239 fraction of features
## We have looked at the first: 0.7981651 fraction of features
## We have looked at the first: 0.9977064 fraction of features
## [1] -186.5797
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features
## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -187.0832
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2004608 fraction of features
## We have looked at the first: 0.4009217 fraction of features
## We have looked at the first: 0.6013825 fraction of features
## We have looked at the first: 0.8018433 fraction of features
## [1] -187.5153
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2009238 fraction of features
## We have looked at the first: 0.4018476 fraction of features
## We have looked at the first: 0.6027714 fraction of features
## We have looked at the first: 0.8036952 fraction of features
## [1] -188.0571
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1990741 fraction of features
## We have looked at the first: 0.3981481 fraction of features
## We have looked at the first: 0.5972222 fraction of features
## We have looked at the first: 0.7962963 fraction of features
## We have looked at the first: 0.9953704 fraction of features
## [1] -188.5812
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.199536 fraction of features
## We have looked at the first: 0.3990719 fraction of features
## We have looked at the first: 0.5986079 fraction of features
## We have looked at the first: 0.7981439 fraction of features
## We have looked at the first: 0.9976798 fraction of features
## [1] -189.1875
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features
## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -189.8068
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2004662 fraction of features
## We have looked at the first: 0.4009324 fraction of features
## We have looked at the first: 0.6013986 fraction of features
## We have looked at the first: 0.8018648 fraction of features
## [1] -190.3245
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2009346 fraction of features
## We have looked at the first: 0.4018692 fraction of features

```



```

## We have looked at the first: 0.6028037 fraction of features
## We have looked at the first: 0.8037383 fraction of features
## [1] -190.875
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1990632 fraction of features
## We have looked at the first: 0.3981265 fraction of features
## We have looked at the first: 0.5971897 fraction of features
## We have looked at the first: 0.7962529 fraction of features
## We have looked at the first: 0.9953162 fraction of features
## [1] -191.2206
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1995305 fraction of features
## We have looked at the first: 0.399061 fraction of features
## We have looked at the first: 0.5985915 fraction of features
## We have looked at the first: 0.7981221 fraction of features
## We have looked at the first: 0.9976526 fraction of features
## [1] -191.4326
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features
## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -191.6843
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2004717 fraction of features
## We have looked at the first: 0.4009434 fraction of features
## We have looked at the first: 0.6014151 fraction of features
## We have looked at the first: 0.8018868 fraction of features
## [1] -191.9425
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2009456 fraction of features
## We have looked at the first: 0.4018913 fraction of features
## We have looked at the first: 0.6028369 fraction of features
## We have looked at the first: 0.8037825 fraction of features
## [1] -192.1816
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1990521 fraction of features
## We have looked at the first: 0.3981043 fraction of features
## We have looked at the first: 0.5971564 fraction of features
## We have looked at the first: 0.7962085 fraction of features
## We have looked at the first: 0.9952607 fraction of features
## [1] -192.4373
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1995249 fraction of features
## We have looked at the first: 0.3990499 fraction of features
## We have looked at the first: 0.5985748 fraction of features
## We have looked at the first: 0.7980998 fraction of features
## We have looked at the first: 0.9976247 fraction of features
## [1] -192.6734
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features

```

```

## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -192.8536
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2004773 fraction of features
## We have looked at the first: 0.4009547 fraction of features
## We have looked at the first: 0.601432 fraction of features
## We have looked at the first: 0.8019093 fraction of features
## [1] -193.2287
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2009569 fraction of features
## We have looked at the first: 0.4019139 fraction of features
## We have looked at the first: 0.6028708 fraction of features
## We have looked at the first: 0.8038278 fraction of features
## [1] -193.4272
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1990408 fraction of features
## We have looked at the first: 0.3980815 fraction of features
## We have looked at the first: 0.5971223 fraction of features
## We have looked at the first: 0.7961631 fraction of features
## We have looked at the first: 0.9952038 fraction of features
## [1] -193.4895
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1995192 fraction of features
## We have looked at the first: 0.3990385 fraction of features
## We have looked at the first: 0.5985577 fraction of features
## We have looked at the first: 0.7980769 fraction of features
## We have looked at the first: 0.9975962 fraction of features
## [1] -193.6785
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2 fraction of features
## We have looked at the first: 0.4 fraction of features
## We have looked at the first: 0.6 fraction of features
## We have looked at the first: 0.8 fraction of features
## We have looked at the first: 1 fraction of features
## [1] -193.9464
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2004831 fraction of features
## We have looked at the first: 0.4009662 fraction of features
## We have looked at the first: 0.6014493 fraction of features
## We have looked at the first: 0.8019324 fraction of features
## [1] -193.9991
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.2009685 fraction of features
## We have looked at the first: 0.401937 fraction of features
## We have looked at the first: 0.6029056 fraction of features
## We have looked at the first: 0.8038741 fraction of features
## [1] -194.0603
## We have selected thus far: 1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first: 0.1990291 fraction of features
## We have looked at the first: 0.3980583 fraction of features
## We have looked at the first: 0.5970874 fraction of features
## We have looked at the first: 0.7961165 fraction of features
## We have looked at the first: 0.9951456 fraction of features

```

```
## [1] -194.2166
## We have selected thus far:  1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first:  0.1995134  fraction of features
## We have looked at the first:  0.3990268  fraction of features
## We have looked at the first:  0.5985401  fraction of features
## We have looked at the first:  0.7980535  fraction of features
## We have looked at the first:  0.9975669  fraction of features
## [1] -194.3396
## We have selected thus far:  1 4 5 2 3 231 256 257 278 152 72 88 390 56 380 370 46 31 314 451 276 301
## We have looked at the first:  0.2  fraction of features
## We have looked at the first:  0.4  fraction of features
## We have looked at the first:  0.6  fraction of features
## We have looked at the first:  0.8  fraction of features
## We have looked at the first:  1  fraction of features
setdiff(indSelect,leapsModel)

## [1] 112  64 318  58 491 366 111 460 350 135 173 206  54 118 480 376
setdiff(leapsModel,indSelect)

## numeric(0)
```

Problem 6 (optional)

On the first set of lecture notes, we covered an example for predicting punctuation given a male user has entered the phrase “thank you”. We computed the loss for two different procedures \hat{f}_1 and \hat{f}_2 . Now, we want to compute the risk, which is the expected value of the loss.

As a review, suppose a random variable Z takes a value 1 with probability π and 0 with probability $1 - \pi$, where $0 \leq \pi \leq 1$. Then

$$\mathbb{E}Z = 1 * \pi + 0 * (1 - \pi) = \pi.$$

Compute the following risks.

Part a

$$\begin{aligned} R(\hat{f}_1) &= \mathbb{E}\ell(\hat{f}_1(X), Y) = \dots \\ &= \ell(“.”, “.”)\mathbb{P}(Y = “.”|X) + \ell(“.”, “!”)\mathbb{P}(Y = “!”|X) \\ &= 0 \cdot 0.43 + 1 \cdot 0.57 \end{aligned}$$

Part b

$$\begin{aligned} R(\hat{f}_2) &= \mathbb{E}\ell(\hat{f}_2(X), Y) = \dots \\ &= \ell(“!”, “.”)\mathbb{P}(Y = “.”|X) + \ell(“!”, “!”)\mathbb{P}(Y = “!”|X) \\ &= 1 \cdot 0.43 + 0 \cdot 0.57 \end{aligned}$$