Homework 4

Caroline Nelson

3/8/2017

Problem 1

It costs a company \$12 to purchase an hour of labor and \$15 to purchase an hour of capital. If L hours of labor and K units of capital are available, then $0.05L^{\frac{2}{3}}K^{\frac{1}{3}}$ machines can be produced. Suppose the company has \$100,000 to purchase labor and capital. What is the maximum number of machines it can produce?

```
library(quadprog)
machines<-function(L){
  K=(100000-12*L)/15
  machine = 0.05*L^(2/3)*K^(1/3)
  return(-machine)
}
s = optim(par=c(5000),fn=machines, method='L-BFGS-B')
```

```
## K=  2222.222 L=  5555.556
```

```
## They can produce  204.6684 machines.
```

Problem 2

Find a portfolio that achieves an expected monthly return of at least 1% and minimizes portfolio variance. What are the fractions invested in each stock? What are the portfolio's estimated mean, variance, and standard deviation?

```
cov = cor2cov(corr,stdev)
D = 2*cov
d = rep(0,27)
A = matrix(0,27,29)
#A needed to be transposed to meet requirement rows same as D
A[,1]=rep(1,27)
A[,2]=mean_ret
A[,3:29]=diag(27)
#greater than 0 requirement
b=c(1,0.01,rep(0,27))
s = solve.QP(D,d,A,b,meq=1)
```

```
## The following weights are the solution of the formulation:  0 0.05 0
0 0 0 0 0 0 0 0 0 0.02 0 0 0.02 0 0 0.14 0 0.27 0 0 0.13 0.06 0 0.31
```

```
## The standard deviation of the portfolio is  0.02973382
```

```
## The variance of the portfolio is  0.0008840998

## The mean return of the portfolio is  0.01
```

Problem 3

```
regdata = read.csv('variable_selection.csv')

model1 = lm(regdata$y~regdata$x2+regdata$x3)
model2 = lm(regdata$y~regdata$x1)
model3 = lm(regdata$y~regdata$x2)
model4 = lm(regdata$y~regdata$x3)
model5 = lm(regdata$y~regdata$x1+regdata$x2)
model6 = lm(regdata$y~regdata$x1+regdata$x3)
summary(model1)

##
## Call:
## lm(formula = regdata$y ~ regdata$x2 + regdata$x3)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.072 -2.111 -0.457  2.337  7.599
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.71994    1.34887   7.206 1.25e-10 ***
## regdata$x2   3.93181    0.13484  29.159  < 2e-16 ***
## regdata$x3  -0.02828    0.10516  -0.269    0.789
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.009 on 97 degrees of freedom
## Multiple R-squared:  0.8981, Adjusted R-squared:  0.896
## F-statistic: 427.7 on 2 and 97 DF,  p-value: < 2.2e-16

summary(model2)

##
## Call:
## lm(formula = regdata$y ~ regdata$x1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.6770  -5.0361  -0.2131   5.5414  20.8261
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   21.395      2.934   7.292 7.95e-11 ***
## regdata$x1     2.756      0.922   2.989  0.00354 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.979 on 98 degrees of freedom
## Multiple R-squared:  0.08355,    Adjusted R-squared:  0.0742
## F-statistic: 8.934 on 1 and 98 DF,  p-value: 0.003538
```

```
summary(model3)
```

```
##
## Call:
## lm(formula = regdata$y ~ regdata$x2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0853 -2.1931 -0.4309  2.3346  7.5392
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.4198     0.7537   12.50   <2e-16 ***
## regdata$x2    3.9343     0.1339   29.38   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.995 on 98 degrees of freedom
## Multiple R-squared:  0.8981, Adjusted R-squared:  0.897
## F-statistic: 863.4 on 1 and 98 DF,  p-value: < 2.2e-16
```

```
summary(model4)
```

```
##
## Call:
## lm(formula = regdata$y ~ regdata$x3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.9249  -5.9699  -0.2864   6.4416  20.9370
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  32.1473     3.4450   9.332 3.42e-15 ***
## regdata$x3   -0.2365     0.3262  -0.725     0.47
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.354 on 98 degrees of freedom
## Multiple R-squared:  0.005333,   Adjusted R-squared:  -0.004817
## F-statistic: 0.5255 on 1 and 98 DF,  p-value: 0.4703
```

```
summary(model5)
```

```
##
## Call:
## lm(formula = regdata$y ~ regdata$x1 + regdata$x2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4861 -0.2674  0.0260  0.3658  1.1301
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.15258    0.21048   0.725     0.47
## regdata$x1   2.99924    0.05337  56.195   <2e-16 ***
## regdata$x2   3.96917    0.02324 170.781   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5196 on 97 degrees of freedom
## Multiple R-squared:  0.997,  Adjusted R-squared:  0.9969
## F-statistic: 1.592e+04 on 2 and 97 DF,  p-value: < 2.2e-16
```

```
summary(model6)
```

```
##
## Call:
## lm(formula = regdata$y ~ regdata$x1 + regdata$x3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.2130  -5.4023   0.2336   5.6330  21.2464
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  23.6973     4.3676   5.426 4.24e-07 ***
## regdata$x1    2.7468     0.9244   2.972  0.00374 **
## regdata$x3   -0.2239     0.3139  -0.713  0.47747
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.002 on 97 degrees of freedom
## Multiple R-squared:  0.08833,   Adjusted R-squared:  0.06953
## F-statistic: 4.699 on 2 and 97 DF,  p-value: 0.01128
```

```
## [1] "The model with X1 and X2 is the best, with R-squared of 0.997,
and very significant coefficients."
```

Problem 4

In an electrical network, the power loss incurred when a current of I amperes flows through a resistance of R ohms is I^2R watts. In the figure below, 710 amperes of current must be sent from node 1 to node 4. The current flowing through each node must satisfy conservation of flow. For example, for node 1, 710 = flow through 1-ohm resistor + flow through 4-ohm resistor. Remarkably, nature determines the current flow through each resistor by minimizing the total power loss in the network.

```
A = matrix(0,10,5)
A[1,]=c(1,1,0,0,0)
A[2,]=c(0,0,0,1,1)
A[3,]=c(0,1,1,1,0)
A[4,]=c(0,1,1,0,-1)
A[5,]=c(-1,0,1,1,0)
A[6:10,]=diag(5)

b=c(710,710,710,rep(0,7))
d=c(rep(0,5))
D = matrix(c(1,0,0,0,0,0,4,0,0,0,0,0,6,0,0,0,0,0,12,0,0,0,0,0,3),5,5)

sol = solve.QP(D,d,t(A),b,factorized=FALSE)

## The optimal currents for each resistor are:  371.3846 338.6154
163.8462 207.5385 502.4615

## The minimal power loss is:  2031911 watts.
```

Problem 5

The goal is to determine a set of ratings for the 32 NFL teams that most accurately predicts the actual outcomes of the games played. Use NLP to find the ratings that best predict the actual point spreads observed. The model will estimate the home team advantage and the ratings. The objective is to minimize the sum of squared prediction errors.

```
nfl = read.csv('nflratings.csv')
mat = as.matrix(nfl)

err<-function(v){

  err_list = list()
  for(i in 1:256){
    ht = mat[i,2] #home team index
    at = mat[i,3] #away team index
    actual = mat[i,4] - mat[i,5]
    pred = v[ht] - v[at] + v[33]
    error = (actual - pred)^2
    err_list = c(err_list, error)

  }
  return(Reduce('+', err_list))
}
s = optim(par=c(rep(85,32),2),fn=err,method='BFGS')
```

```
## The 32 rankings are:  84.52 89.84 92.75 83.09 88.76 79.81 87.54 76.89 92.12 85.64 70.5 92.25 86.98 90.86 78.44 76.89 86.62 92.06 96.12 95.63 85.1 93.15 75.03 90.96 86.64 67.72 92.61 85.24 74.73 79.17 82.19 80.14

## The home team advantage is:  2.17
```