Homework 1 Solutions

Jacob Townson

January 18, 2018

1) Answer the following questions about the dataset "HeightData.txt".

Before we begin, we must read in the text file containing "HeightData.txt".

```
setwd("C:/Users/obewa_000/Documents/My real documents/University of Louisville/public_work/Classes/Homew
D=read.table("HeightData.txt",sep="\t",header=TRUE)
```

(a - 1pt) Find the average height of all children.

```
mean(D$childHeight)
```

```
## [1] 66.74593
```

(b - 1pt) Find the percentage of female children.

```
girls = D$gender=='female'
percent.girls = sum(girls)/length(girls)
percent.girls
```

```
## [1] 0.4850107
```

Thus 48.50107% of the children are female.

(c - 1pt) What is the height of the tallest child?

```
max(D$childHeight)
```

```
## [1] 79
```

Thus the tallest child is 79 inches tall.

(d - 1pt) How many families have exactly 4 children?

```
require(tidyr)
```

```
## Loading required package: tidyr
## Warning: package 'tidyr' was built under R version 3.4.3
```

```
require(dplyr)
```

```
## Loading required package: dplyr
```

```
## Warning: package 'dplyr' was built under R version 3.4.3

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
D.4 = select(D,family,children)
D.4 = distinct(D.4)
D.4 = filter(D.4, children == 4)
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.3
```

```
length(D.4$family)
```

```
## [1] 31
```

This is the number of families with exactly 4 children. It was helpful to reduce the data in this way as now we have a dataframe that contains the exact families with 4 children as well, which in normal cases could be useful for further analysis.

(e - 2pts) What are the heights of the children in the family with the tallest mother?

```
require(knitr)
```

```
## Loading required package: knitr

## Warning: package 'knitr' was built under R version 3.4.3
```

```
tallmom = max(D$mother)
Dmoms = filter(D, mother == tallmom)
kable(Dmoms)
```

family	father	mother	midparentHeight	children	childNum	gender	childHeight
128	68.7	70.5	72.42	2	1	male	71.0
128	68.7	70.5	72.42	2	2	female	61.7

This dataframe shows us all information about the family with the tallest mother.

(f - 2pts) What percentage of children are taller than both of their parents?

```
total = length(D$childHeight)
Dtall = filter(D, D$childHeight > D$father)
Dtall = filter(Dtall, Dtall$childHeight > Dtall$mother)
tallkids = length(Dtall$childHeight)
percent.tall = (tallkids/total)*100
percent.tall
```

```
## [1] 23.8758
```

This shows us that 23.8758% of children are taller than both of their parents.

(g - 2pts) What is the average height of females whose mother was at least 68 inches tall?

```
Dgals = filter(D, D$mother >= 68)
Dgals = filter(Dgals, Dgals$gender == 'female')
average = mean(Dgals$childHeight)
average
```

```
## [1] 66.46667
```

Thus we find that the avergae height for girls with mothers taller than 68 inches is 66.46667 inches.

2)

(a - 5pts) Suppose that $H_1^-, ..., H_m^-$ are generalized inverses of a matrix H and $\alpha_1, ..., \alpha_m$ are scalars. Find a condition on the scalars $\alpha_1, ..., \alpha_m$ that makes $\alpha_1 H_1^- + ... + \alpha_m H_m^-$ a generalized inverse of H.

Recall H^- is a generalized inverse of H given that $HH^-H = H$. We know that $H_1^-, ..., H_m^-$ are generalized inverses of a matrix H. Thus

$$H(\alpha_1 H_1^- + ... + \alpha_m H_m^-)H = H\alpha_1 H_1^- H + ... + H\alpha_m H_m^- H$$
$$= \alpha_1 H H_1^- H + ... + \alpha_m H H_m^- H = \alpha_1 H + ... + \alpha_m H = (\alpha_1 + ... + \alpha_m)H$$

Thus we can tell that $\alpha_1 H_1^- + ... + \alpha_m H_m^-$ is a generalized inverse of H if $\alpha_1 + ... + \alpha_m = 1$. QED

(b - 5pts) Let H be defined as in the homework. Find a generalized inverse of H which has no elements equal to 0.

```
require(MASS)
```

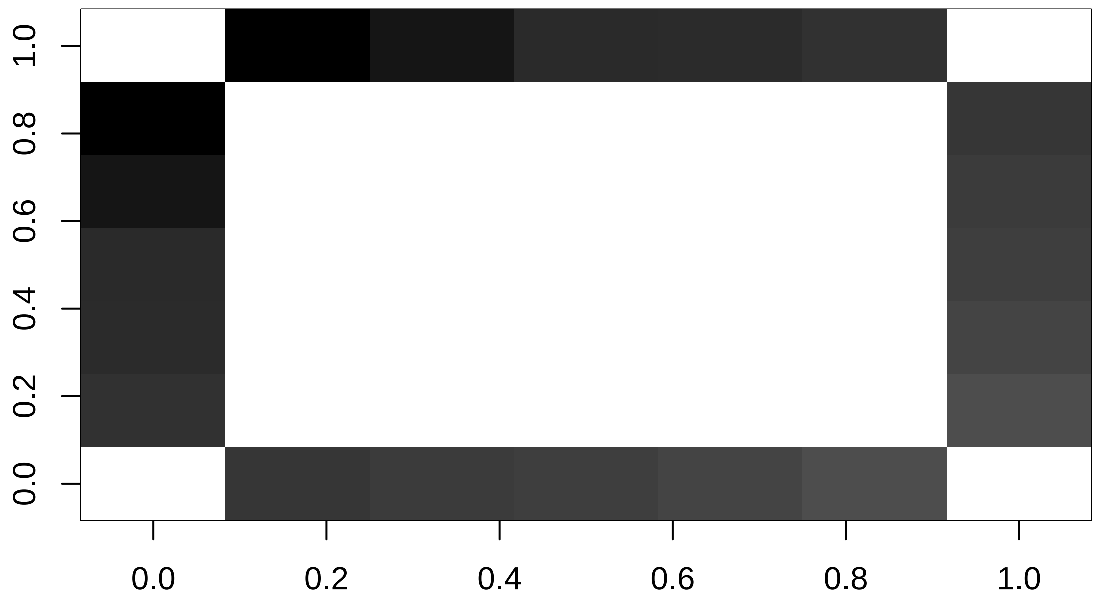
```
## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select
```

```
H = matrix(c(2,2,-1,-1,1,1,3,1,-2), 3,3)
ginv(H)
```

```
##             [,1]        [,2]        [,3]
## [1,]  0.03846154  0.32051282  0.05128205
## [2,] -0.11538462  0.37179487  0.17948718
## [3,]  0.15384615 -0.05128205 -0.12820513
```

3) Consider the large matrix labeled as O in the homework. Use computer software to find the two largest eigenvalues of O, denoted by λ_1 and λ_2. Letting u_1 be the eigenvector of O associated with λ_1 and u_2 be the eigenvector of O associated with λ_2, compute $A = \lambda_1 u_1 u_1' + \lambda_2 u_2 u_2'$.

```
O = matrix(c(255,54,71,87,88,93,255,54,255,255,255,255,255,97,71,255,255,255,255,255,101,87,255,255,255
image(t(O[7:1,]),col=gray((0:255)/255))
```



Using the eigen() function, we can find the max eigenvalue and corresponding eigenvectors and calculate A from there.

```
E = eigen(O)
E
```

```
## eigen() decomposition
## $values
## [1]  1.373685e+03  4.151384e+02  1.289019e+01  1.256982e-13  2.620790e-14
## [6] -5.587959e-01 -1.615523e+01
##
## $vectors
##            [,1]       [,2]        [,3]          [,4]          [,5]
## [1,] -0.2049720  0.6922375 -0.45765380  0.000000e+00  0.000000e+00
## [2,] -0.4186751 -0.1949149  0.53020254  3.521904e-01 -2.284790e-01
```

```
## [3,] -0.4219260 -0.1603606  0.09424771 -7.566207e-01  3.906608e-01
## [4,] -0.4246705 -0.1305774 -0.39000992  4.970539e-02 -4.997859e-01
## [5,] -0.4257125 -0.1211512 -0.21599687  5.223998e-01  6.629717e-01
## [6,] -0.4277085 -0.1019515 -0.10019312 -1.676749e-01 -3.253676e-01
## [7,] -0.2452864  0.6441876  0.54014312  1.249001e-14  6.938894e-16
##             [,6]       [,7]
## [1,] -0.02098782  0.5185617
## [2,]  0.09260833  0.5663812
## [3,]  0.20737922  0.1388647
## [4,]  0.54645723 -0.3156338
## [5,] -0.07064908 -0.2000308
## [6,] -0.79938550 -0.1537420
## [7,]  0.07316486 -0.4772307
```

```
E$values[order(E$values)[6:7]]
```

```
## [1]  415.1384 1373.6854
```

So we can see that the two largest eigenvalues are the first two, meaning that their corresponding eigenvectors are columns 1 and 2 of the matrix containing all of the eigenvectors.

```
lambda_1 = E$values[1]
lambda_2 = E$values[2]
u_1 = E$vectors[,1]
u_2 = E$vectors[,2]
lambda_1
```

```
## [1] 1373.685
```

```
u_1
```

```
## [1] -0.2049720 -0.4186751 -0.4219260 -0.4246705 -0.4257125 -0.4277085
## [7] -0.2452864
```

```
lambda_2
```

```
## [1] 415.1384
```

```
u_2
```

```
## [1]  0.6922375 -0.1949149 -0.1603606 -0.1305774 -0.1211512 -0.1019515
## [7]  0.6441876
```

These are our largest eigenvalues and their corresponding eigenvectors.

```
A = (lambda_1 * (u_1 %*% t(u_1))) + (lambda_2 * (u_2 %*% t(u_2)))
A
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## [1,] 256.64467  61.87156  72.71689  82.04861  85.05086  91.13034 254.18759
## [2,]  61.87156 256.56357 255.63722 254.80571 254.64227 254.23665  88.94556
## [3,]  72.71689 255.63722 255.22106 254.82905 254.80547 254.68418  99.28166
## [4,]  82.04861 254.80571 254.82905 254.81563 254.91253 255.03615 108.17127
## [5,]  85.05086 254.64227 254.80547 254.91253 255.04781 255.24942 111.04319
## [6,]  91.13034 254.23665 254.68418 255.03615 255.24942 255.60953 116.85023
## [7,] 254.18759  88.94556  99.28166 108.17127 111.04319 116.85023 254.92156
```