ESI 4606 Analytics I - Foundations of Data Science
Homework 6
Due: November 23st (11:00AM), 2022

Problem 1 (1.5 points)

Solution: (a) The ridge regression results:

```
(Intercept)          x1          x2          x3          x4          x5          x6
 0.96280890  0.31088821 -1.03773737  0.16597620 -0.61404585  0.06608268 -0.30525163
         x7          x8          x9
 0.02680982 -0.10025055  0.03171445
```

(b) The LASSO regression results:

```
(Intercept)          x1          x2          x3          x4          x5          x6
  0.9628089   0.3046287  -1.9228859   0.0000000   0.0000000   0.0000000   0.0000000
         x7          x8          x9
  0.0000000   0.0000000   0.0000000
```

Problem 2 (2 points)

Solution: (a) $z_{11} = \phi_{11}x_{11} + \phi_{21}x_{12} = -\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2} = -\sqrt{2}$

$z_{12} = \phi_{12}x_{11} + \phi_{22}x_{12} = \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2} = 0$

$z_{21} = \phi_{11}x_{21} + \phi_{21}x_{22} = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} = \sqrt{2}$

$z_{22} = \phi_{12}x_{21} + \phi_{22}x_{22} = -\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} = 0$

(a) Since $z_{31} = 1$, $z_{32} = 1$, and recall

$$\begin{cases} z_{31} = \phi_{11}x_{31} + \phi_{21}x_{32} = -\frac{\sqrt{2}}{2}x_{31} - \frac{\sqrt{2}}{2}x_{32} & (1) \\ z_{32} = \phi_{12}x_{31} + \phi_{22}x_{32} = \frac{\sqrt{2}}{2}x_{31} - \frac{\sqrt{2}}{2}x_{32} & (2) \end{cases}$$

Eq.(2)$-$Eq.(1), we get $\sqrt{2}x_{31} = z_{32} - z_{31} = 0 \Rightarrow x_{31} = 0$.

Eq.(1)+Eq.(2), we get $-\sqrt{2}x_{32} = z_{32} + z_{31} = 2 \Rightarrow x_{32} = -\sqrt{2}$.

Similarly, since $z_{41} = -1$, $z_{42} = -1$, and recall

$$\begin{cases} z_{41} = \phi_{11}x_{41} + \phi_{21}x_{42} = -\frac{\sqrt{2}}{2}x_{41} - \frac{\sqrt{2}}{2}x_{42} & (3) \\ z_{42} = \phi_{12}x_{41} + \phi_{22}x_{42} = \frac{\sqrt{2}}{2}x_{41} - \frac{\sqrt{2}}{2}x_{42} & (4) \end{cases}$$

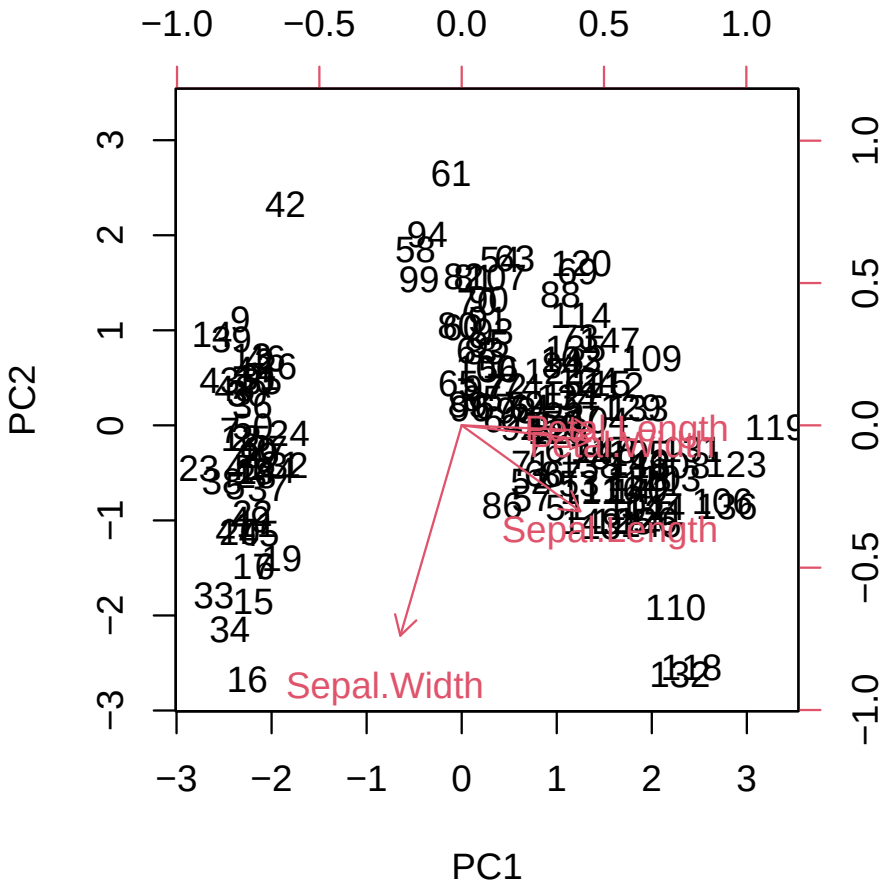
Eq.(4)$-$Eq.(3), we get $\sqrt{2}x_{41} = z_{42} - z_{41} = 0 \Rightarrow x_{41} = 0$.

Eq.(3)+Eq.(4), we get $-\sqrt{2}x_{42} = z_{42} + z_{41} = 2 \Rightarrow x_{42} = \sqrt{2}$.

Problem 3 (1.5 points)
Solution: (a) Loading vectors of PC1 and PC2:

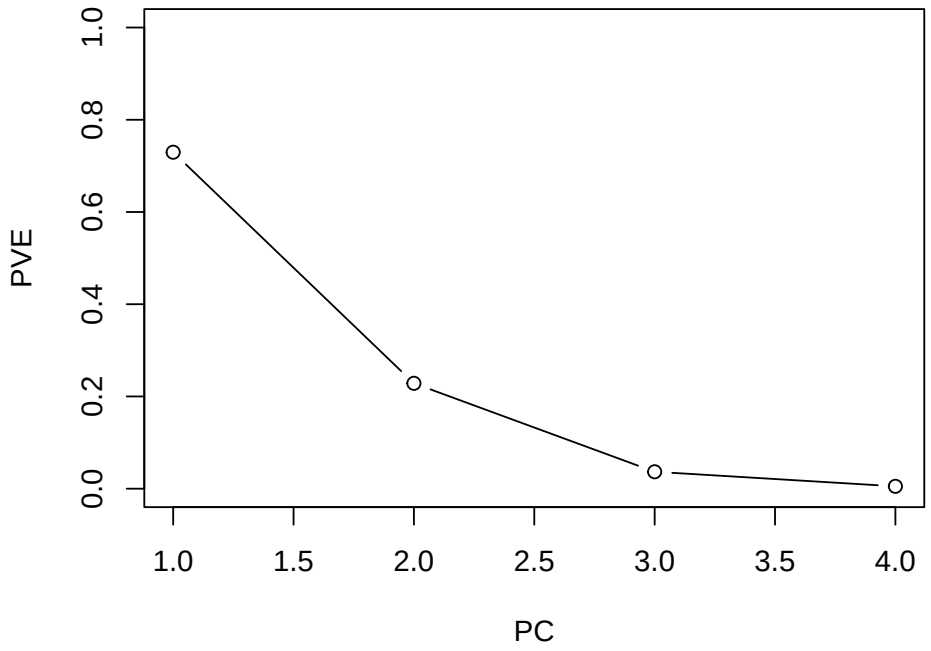
```
                    PC1         PC2
Sepal.Length  0.5210659 -0.37741762
Sepal.Width  -0.2693474 -0.92329566
Petal.Length  0.5804131 -0.02449161
Petal.Width   0.5648565 -0.06694199
```

(b) The biplot:



(c) The PVE: 0.729624454 0.228507618 0.036689219 0.005178709

scree plot:



Appendix A: R codes for Problem 1

```
library(glmnet)
data.mat=data.matrix(data)
#ridge regression
set.seed(2)
ridge.cv=cv.glmnet(data.matrix(data.mat[,2:11]),data.mat[,1],alpha=0)
lambda.best=ridge.cv$lambda.1se
lambda.best
ridge.model=glmnet(data.mat[,2:11],data.mat[,1],alpha=0,lambda=lambda.best,thresh=1e-12)
predict(ridge.model,s=lambda.best,type="coefficients")[1:10,]
#LASSO regression
set.seed(2)
lasso.cv=cv.glmnet(data.matrix(data.mat[,2:11]),data.mat[,1],alpha=1)
lambda.best=lasso.cv$lambda.1se
lambda.best
lasso.model=glmnet(data.mat[,2:11],data.mat[,1],alpha=1,lambda=lambda.best,thresh=1e-12)
predict(lasso.model,s=lambda.best,type="coefficients")[1:10,]
```

Appendix B: R codes for Problem 3

```
library(datasets)
data=iris[,1:4]
pca.obj=prcomp(data,scale=TRUE)
#output the first 2 PC loading vectors
pca.obj$rotation[,1:2]
# biplot
biplot(pca.obj,scale=0)
# PVE of each PC
pve=pca.obj$sdev^2/sum(pca.obj$sdev^2)
#scree plot
plot(pve,xlab="PC", ylab="PVE",ylim=c(0,1),type='b')
```