HW 4

1) Consider the covariance matrix below:

A) Find the eigenvalues and eigenvectors of Σ.

(E-yI)v=0;Ev=Ivy
det(E): (1-y)(3-y)=4; 1-4y-y^2=0
y=-4+-(((4^2)-4(1)(1))^(1/2))/2(1)
=-4+-((16-4)^(1/2))/2
=-4+-(3.464/2)=-4-1.732,-4+1.732
=-5.732, -2.286

$$\Sigma = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$$

(1-y)o11=-o12 : (1+5.732)o11=-o12
(2+5.732)o11=-3o12
(1+2.286)o21=-o22: (2+2.286)o21=-3o22

(o11)^2+(6.732)^2x(o11^2)=1
o11=1/6.81:o12=(1-(1/6.81))^(1/2)=.978

B) Find the principal components of Σ.

((3.286)^2)+1)(o21^2)=1, o21=(1/11.8)^(1/2)
o21=.29, o22=(1-.29)^(1/2)=.843

PCV:.978x1+.116x2:.843x1+.29x2

2) Consider the Iris dataset in R.

R>attach(iris)

A) Form principal components variables based on the four variables Sepal.Length, Sepal.Width, Petal.Length, and Petal.Width. List out all possible principal components.

```
> x=cbind(Sepal.Length,Sepal.Width,Petal.Length,Petal.Width)
> cov(x)
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length    0.6856935  -0.0424340    1.2743154   0.5162707
Sepal.Width    -0.0424340   0.1899794   -0.3296564  -0.1216394
Petal.Length    1.2743154  -0.3296564    3.1162779   1.2956094
Petal.Width     0.5162707  -0.1216394    1.2956094   0.5810063
```

B) Report the proportion of variance and the cumulative proportion of variance explained by each principal component.

```
> prcomp(x)
Standard deviations (1, .., p=4):
[1] 2.0562689 0.4926162 0.2796596 0.1543862

Rotation (n x k) = (4 x 4):
                PC1        PC2        PC3        PC4
Sepal.Length  0.36138659 -0.65658877  0.58202985  0.3154872
Sepal.Width  -0.08452251 -0.73016143 -0.59791083 -0.3197231
Petal.Length  0.85667061  0.17337266 -0.07623608 -0.4798390
Petal.Width   0.35828920  0.07548102 -0.54583143  0.7536574
```

```
[1] 0.924618723 0.053066483 0.017102610 0.005212184

> summary(prcomp(x))
Importance of components%s:
                          PC1    PC2    PC3     PC4
Standard deviation     2.0563 0.49262 0.2797 0.15439
Proportion of Variance 0.9246 0.05307 0.0171 0.00521
Cumulative Proportion  0.9246 0.97769 0.9948 1.00000
```

C) Produce a scree plot of the principal components and explain how many principal components you would choose.

see attached files

I would choose only the first 2 principal components

since they have the largest difference.

3) Consider the Iris dataset in R.

A) Form linear discriminant variables based on the four variables Sepal.Length, Sepal.Width, Petal.Length, and Petal.Width. List out all possible linear discriminant variables.

```
> ldaout<-lda(Species~Sepal.Length+Sepal.Width+Petal.Length+Petal.Width,iris)
> ldaout
Call:
lda(Species ~ Sepal.Length + Sepal.Width + Petal.Length + Petal.Width,
    data = iris)

Prior probabilities of groups:
    setosa versicolor  virginica
 0.3333333  0.3333333  0.3333333

Proportion of trace:
   LD1    LD2
0.9912 0.0088

Group means:
           Sepal.Length Sepal.Width Petal.Length Petal.Width
setosa            5.006       3.428        1.462       0.246
versicolor        5.936       2.770        4.260       1.326
virginica         6.588       2.974        5.552       2.026

Coefficients of linear discriminants:
                    LD1         LD2
Sepal.Length  0.8293776  0.02410215
Sepal.Width   1.5344731  2.16452123
Petal.Length -2.2012117 -0.93192121
Petal.Width  -2.8104603  2.83918785
```

B) Predict the species of a flower with a sepal length of 5.8, sepal width of 3.1, petal length of 3.8, and petal width of 1.2.

```
versicolor
```

versicolor

C) Calculate the first linear discriminant score of a flower with a sepal length of 5.8, sepal width of 3.1, petal length of 3.8, and petal width of 1.2.

LD1=+1.5sw-2.2pl-2.8pw

D) Using histograms plot first linear discriminant scores by each species group (hint: one histogram for each group).

See second attached.

E) Using histograms plot second discriminant scores by each species group (hint: one histogram for each group).

See Second Attached

```
 hist(ldaout)
Error in hist.default(ldaout) : 'x' must be numeric
> hist(ldatable)
> ldapred<-predict(ldaout,iris)
> ldaclass<-ldapred$class
> ldatable<-table(ldaclass,iris$Species)
> hist(ldatable)
> ldatablesl<-table(ldaclass,iris$Species)
> hist(ldatablesl)
> ldaout<-lda(Species~Sepal.Width)
> ldapred<-predict(ldaout,iris)
> ldaclass<-ldapred$class
> ldatablesw<-table(ldaclass,iris$Species)
> hist(ldatablesw)
> ldaout<-lda(Species~Petal.Length)
> ldapred<-predict(ldaout,iris)
> ldaclass<-ldapred$class
> ldatablepl<-table(ldaclass,iris$Species)
> hist(ldatablepl)
> ldaout<-lda(Species~Petal.Width)
> ldapred<-predict(ldaout,iris)
> ldaclass<-ldapred$class
> ldatablepw<-table(ldaclass,iris$Species)
> hist(ldatablepw)
```

Thanks to: https://www.youtube.com/watch?v=gu3MMpakTlw