Handout # 22

```
############################################
#  R Code for canonical correlation analysis #
############################################

# Use the iris data set and consider the entire data set (all three species)
TX<- c(  5.1,  3.5,  1.4,  0.2,  1,
# more rows here
            5.9,  3.0,  5.1,  1.8,  3)

X<-matrix(data=TX,ncol=5,byrow=TRUE)
Iris <- data.frame(SepalL = X[,1],SepalW = X[,2],PetalL = X[,3], PetalW = X[,4], Group = X[,5])
sepal.meas <- Iris[,1:2]
petal.meas <- Iris[,3:4]

# CCA
# Find the blocks of the COVARIANCE matrix:
S11 <- cov(sepal.meas)
S22 <- cov(petal.meas)
S12 <- c(cov(sepal.meas[,1], petal.meas[,1]), cov(sepal.meas[,1], petal.meas[,2]),
       cov(sepal.meas[,2], petal.meas[,1]), cov(sepal.meas[,2], petal.meas[,2]))
S12 <- matrix(S12, ncol=ncol(S22), byrow=T) # S12 has q=2 columns, same as number of petal measurements
S21 <- t(S12)  # S21=transpose of S12

# Finding the E1 and E2 matrices:
E1 <- solve(S11) %*% S12 %*% solve(S22) %*% S21
E2 <- solve(S22) %*% S21 %*% solve(S11) %*% S12
eigen(E1)
eigen(E2)

# The canonical correlations are:
canon.corr <- sqrt(eigen(E1)$values)
canon.corr

# The canonical variates are based on the eigenvectors of E1 and E2:
# a1 = (0.78, -0.62)
# b1 = (0.77, -0.63)
# a2 = (0.26,  0.97)
# b2 = (-0.37, 0.93)

# Only the first canonical correlation is really substantial:
# u1 = 0.78*Sepal.Length - 0.62*Sepal.Width
# v1 = 0.77*Petal.Length - 0.63*Petal.Width

# Plotting the first set of canonical variables:
u1 <- as.matrix(Iris[,1:2]) %*% as.matrix(eigen(E1)$vectors[,1])
v1 <- as.matrix(Iris[,3:4]) %*% as.matrix(eigen(E2)$vectors[,1])
plot(u1,v1)
cor(u1,v1)

# Plotting the second set of canonical variables:
u2 <- as.matrix(Iris[,1:2]) %*% as.matrix(eigen(E1)$vectors[,2])
v2 <- as.matrix(Iris[,3:4]) %*% as.matrix(eigen(E2)$vectors[,2])
plot(u2,v2)
cor(u2,v2)
```

```
######################
> eigen(E1)
$values
[1] 0.88542265 0.01536035

$vectors
          [,1]      [,2]
[1,]  0.7809378 0.2607687
[2,] -0.6246088 0.9654013

> eigen(E2)
$values
[1] 0.88542265 0.01536035

$vectors
          [,1]       [,2]
[1,]  0.7743756 -0.3705408
[2,] -0.6327262  0.9288162

>
> # The canonical correlations are:
>
> canon.corr <- sqrt(eigen(E1)$values)
> canon.corr
[1] 0.9409690 0.1239369
>
> # The canonical variates are based on the eigenvectors of E1 and E2:
>
> # a1 = (0.78, -0.62)
> # b1 = (0.77, -0.63)
> # a2 = (0.26,  0.97)
> # b2 = (-0.37, 0.93)
>
> # Only the first canonical correlation is really substantial:
>
> # u1 = 0.78*Sepal.Length - 0.62*Sepal.Width
> # v1 = 0.77*Petal.Length - 0.63*Petal.Width
>
> # Plotting the first set of canonical variables:
>
> u1 <- as.matrix(Iris[,1:2]) %*% as.matrix(eigen(E1)$vectors[,1])
> v1 <- as.matrix(Iris[,3:4]) %*% as.matrix(eigen(E2)$vectors[,1])
> plot(u1,v1)
> cor(u1,v1)
         [,1]
[1,] 0.940969
>
> # Plotting the second set of canonical variables:
>
> u2 <- as.matrix(Iris[,1:2]) %*% as.matrix(eigen(E1)$vectors[,2])
> v2 <- as.matrix(Iris[,3:4]) %*% as.matrix(eigen(E2)$vectors[,2])
> plot(u2,v2)
> cor(u2,v2)
          [,1]
[1,] 0.1239369
```