```
#### Lab 2 - Assignment 1 - LDA and logistic regression
### Setup
data.crabs = read.csv("australian-crabs.csv")
data.male = data.crabs[data.crabs$sex=="Male",]
data.female = data.crabs[data.crabs$sex=="Female",]
X = cbind(data.crabs$RW, data.crabs$CL)
X.male = cbind(data.male$RW, data.male$CL)
X.female = cbind(data.female$RW, data.female$CL)
n = dim(data.crabs)[1]
n.male = dim(data.male)[1]
n.female = dim(data.female)[1]
S.male = cov(X.male) # estimation
S.female = cov(X.female) # estimation
S = 1/n * (n.female*S.female + n.male*S.male) # estimation

### Functions
discriminant_function = function(X, S){
  µ = c(mean(X[,1]), mean(X[,2])) # estimation
  p = dim(X)[1]/n # proportional prior
  w0 = -(1/2) * t(µ)%*%solve(S)%*%µ + log(p)
  w = solve(S)%*%µ
  w1 = w[1]
  w2 = w[2]
  result = c(w0, w1, w2)
  names(result) = c("Intercept", "RW", "CL")
  return(result)
}

### Implementation
## Task 1 - Scatterplot of Carapace Length vs. Read Width with Sex
colorized
plot(data.male$CL, data.male$RW, col=4, pch=19, ylim=c(7,20),
     main="CL vs. RW", xlab="CL", ylab="RW")
points(data.female$CL, data.female$RW, col=2, pch=19)
legend("topleft", legend=c("Male", "Female"), pch=19, col=c("blue",
"red"))

## Task 2
# Get discriminant boundries for the two classes
disc.female = discriminant_function(X.female, S)
# Intercept          RW          CL
# -22.428769    8.248698   -2.161318
disc.male = discriminant_function(X.male, S)
# Intercept          RW          CL
# -12.5634175    2.5658514   -0.2138144
# Calculate boundry
# Boundry: RW = 1.73 + 0.34CL
boundry.lda = (disc.male[1]-disc.female[1] + (disc.male[3]-
disc.female[3])*X[,2]) / (disc.female[2]-disc.male[2])
lines(X[,2], boundry.lda)
# Intercept          RW          CL
# -9.865352   5.682847  -1.947504

# Make predictions
```

```
#disc = c(disc.male[1]-disc.female[1], disc.male[2]-disc.female[2],
disc.male[3]-disc.female[3])
disc = disc.male - disc.female
d = disc[1] + disc[2]*X[,1]+disc[3]*X[,2]
Yfit=(d>0)

# Plot predictions
plot(X[,2], X[,1], col=2*Yfit+2, pch=19,
     main="CL vs. RW", xlab="CL", ylab="RW")
lines(X[,2], boundry.lda)
legend("topleft", legend=c("Male", "Female"), col=c(4,2), pch=19)
#hist(d[where=d<0], col=rgb(1, 0, 0, 0.5), xlim=c(-15, 15))
#hist(d[where=d>0], col=rgb(0, 0, 1, 0.5), add=T)

## Task 3 - Plot Original observations with boundry
plot(data.male$CL, data.male$RW, col="blue", ylim=c(7,20),
     main="CL vs. RW", xlab="CL", ylab="RW")
points(data.female$CL, data.female$RW, col="red")
lines(X[,2], boundry.lda)
legend("topleft", legend=c("Male", "Female"), col=c(1, 2), pch=19)

## Task 4 - Logistic Regression
# Fit the model
glm = glm(sex~RW+CL, family=binomial(link="logit"), data=data.crabs)
# Make predictions
pred.female = glm$fitted.values<0.5
glm.male = X[!pred.female,]
glm.female = X[pred.female,]
# Plot the predictions
plot(glm.male[,2], glm.male[,1], col="blue", ylim=c(7,20),
     main="CL vs. RW", xlab="CL", ylab="RW")
points(glm.female[,2], glm.female[,1], col="red")
# Calculate boundry
boundry.glm = (-glm$coefficients[1] - glm$coefficients[3]*X[,2]) /
glm$coefficients[2]
lines(X[,2], boundry.glm)
```