Exercie 1: Linear Decision Boundaries

Part A

Lets load and important the iris data set

```
data(iris)
summary(iris)
```

```
##   Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
##  Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
##  1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
##  Median :5.800   Median :3.000   Median :4.350   Median :1.300
##  Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
##  3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
##  Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
##        Species
##  setosa    :50
##  versicolor:50
##  virginica :50
##
##
##
```

```
# Lets look at the head of the iris data set
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```
# Number of rows of the iris dataset (number of individual observations)
nrow(iris)
```

```
## [1] 150
```

```
# All of the columns are numeric except for the species column
# They are continuous because they correspond to length and width (except for species) # The species ar
sapply(iris, class)
```

```
## Sepal.Length  Sepal.Width Petal.Length  Petal.Width      Species
##    "numeric"    "numeric"    "numeric"    "numeric"     "factor"
```

```
# The only categorical variable is species in this iris dataset
sapply(iris, class)
```

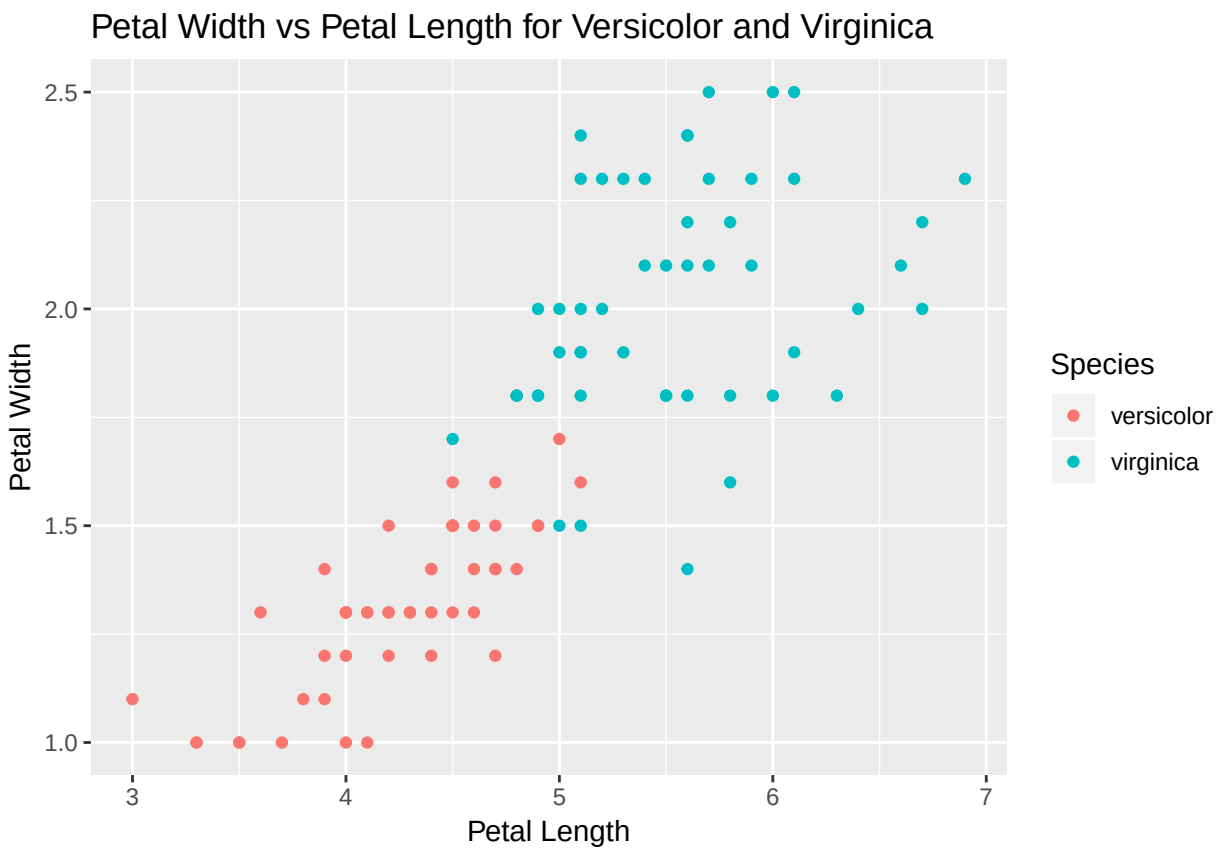
```
## Sepal.Length  Sepal.Width Petal.Length  Petal.Width      Species
##    "numeric"    "numeric"    "numeric"    "numeric"     "factor"
```

From our observations it looks like there are 50 samples of three species: setosa, versicolor and virginica. Each of the samples has measurements of sepal lenght, sepal width and petal length.

Let's plot the second and third iris classses:

```r
# Lets pick the 2nd and 3rd classes: versicolor and virginica and let's plot the subset
plot1 <- iris %>% filter(Species == 'versicolor' | Species == 'virginica') %>%
ggplot(aes(x = Petal.Length, y = Petal.Width, color = Species)) + geom_point() +
labs(color = "Species") + xlab("Petal Length") + ylab("Petal Width") +
ggtitle("Petal Width vs Petal Length for Versicolor and Virginica")

plot1
```



Part B

The following function will take in five inputs (c0, c1, c2) which are constants that will be estimated later in this exercise and (x1, x2) which correspond to petal width and petal length. The output will return a value between a probability value between 0 and 1 where all values above 0.5 will correspond to the 3rd iris class and values below 0.5 to the 2nd iris class.

```r
# inputs: c0, c1, c2, petal_width, petal_length
# outputs:

one_layer_neural_network <- function(c0, c1, c2, petal_width, petal_length) {

  # Linear function
  z <- c0 + c1*x1 + c2*x2

  #result
  result <- 1/(1 + exp(1)^(-z))
```

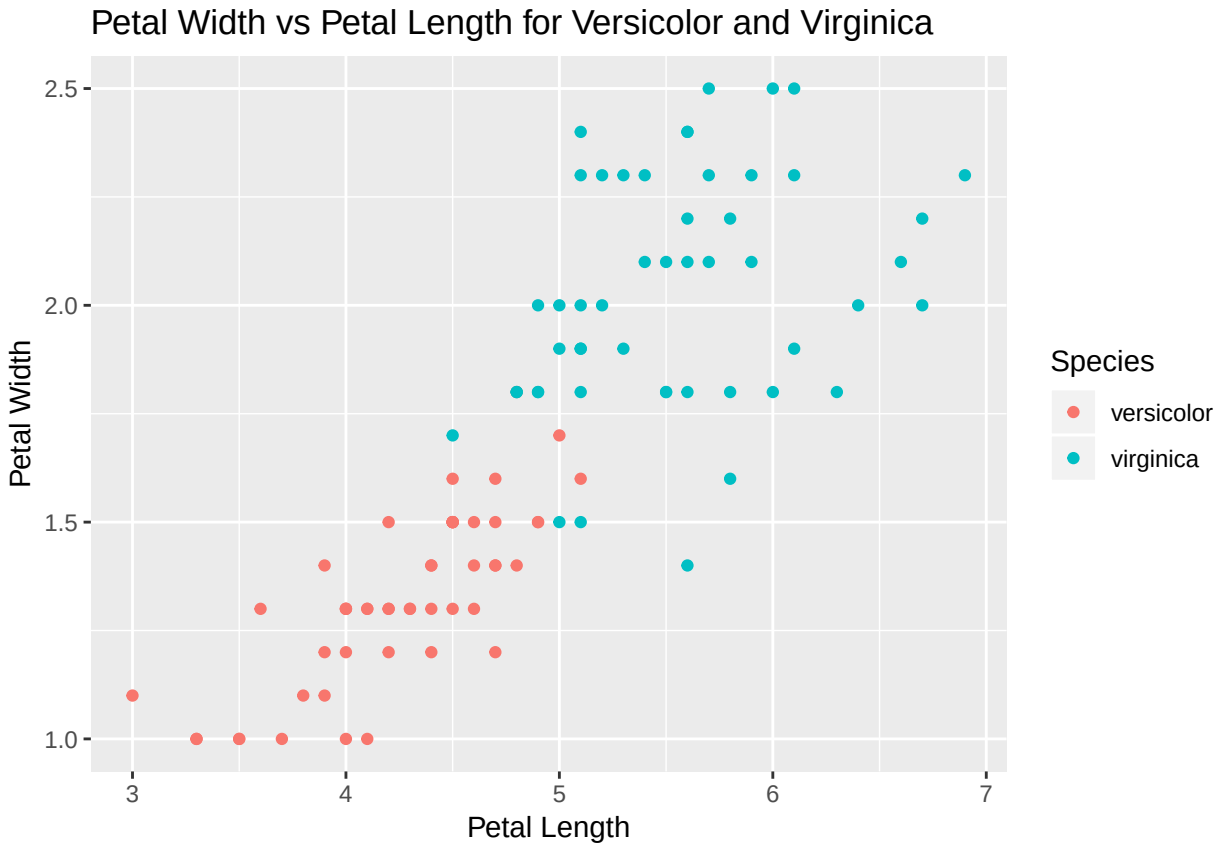
```
  return(result)
}
```

Part C

We are going to estimate the constant values such as...

```
# Lets pick the 2nd and 3rd classes: versicolor and virginica and let's plot the subset
plot2 <- iris %>% filter(Species == 'versicolor' | Species == 'virginica') %>%
ggplot(aes(x = Petal.Length, y = Petal.Width, color = Species)) + geom_point() +
labs(color = "Species") + xlab("Petal Length") + ylab("Petal Width") +
ggtitle("Petal Width vs Petal Length for Versicolor and Virginica")

plot2
```



Part D ### Part E