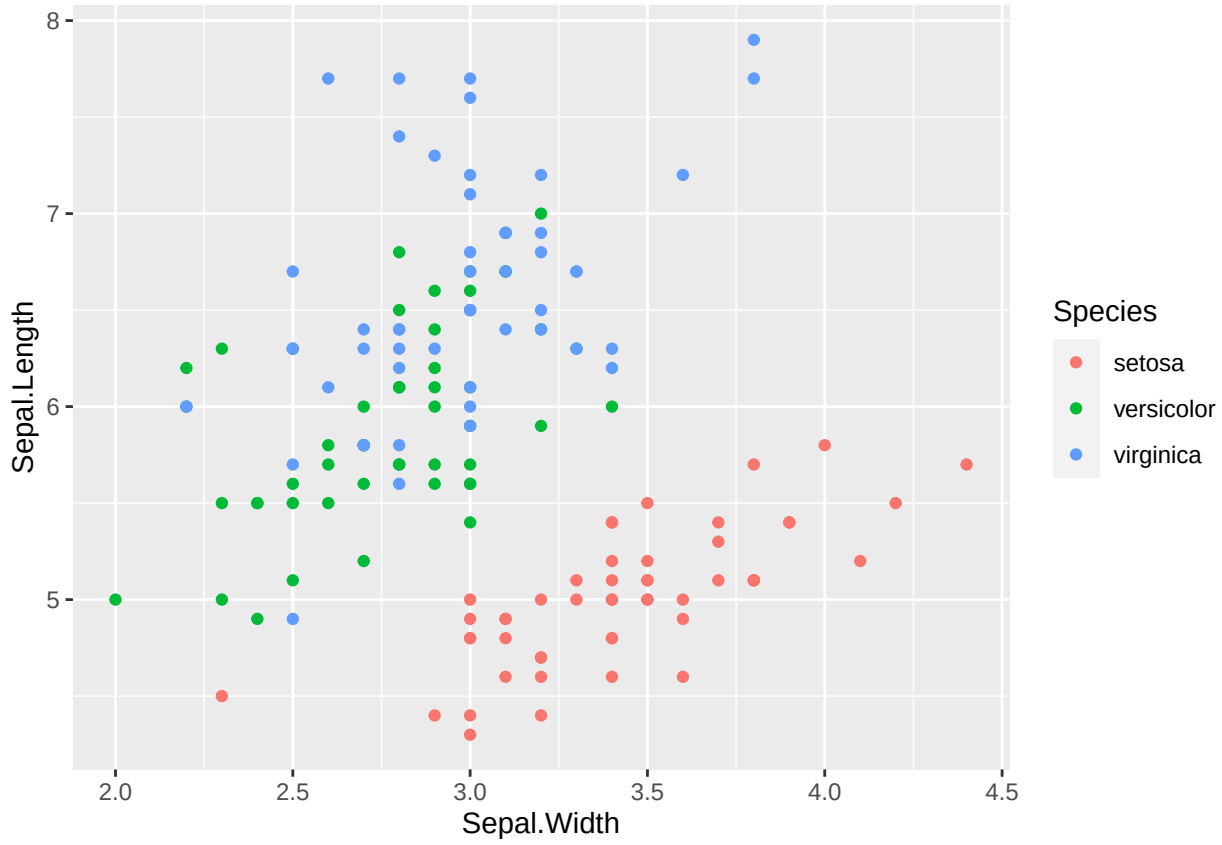
LDA

konda varshith

12/10/2020

Scatter plot of Sepal length vs Sepal Width



Mean

Covariance

```
##              Sepal.Length Sepal.Width
## Sepal.Length   0.12424898  0.09921633
## Sepal.Width    0.09921633  0.14368980
```

```
##              Sepal.Length Sepal.Width
## Sepal.Length   0.40434286  0.09376327
## Sepal.Width    0.09376327  0.10400408
```

```
##              Sepal.Length Sepal.Width
## Sepal.Length   0.26643265  0.08518367
## Sepal.Width    0.08518367  0.09846939
```

Prior of classes

```
## [1] 0.3333333
```

```
## [1] 0.3333333
```

```
## [1] 0.3333333
```

Pooled covariance

```
pooled_matrix <- (virginica_covariance*nrow(virginica)+(setosa_covariance*nrow(setosa))+versicolor_cova
pooled_matrix
```

```
##              Sepal.Length Sepal.Width
## Sepal.Length   0.26500816  0.09272109
## Sepal.Width    0.09272109  0.11538776
```

Underlying probabbilistic model

$$x|y = C_i, \mu_i, \sum N(\mu_i, \sum)$$
$$y|\pi = Multinomial(\pi_i.....\pi_k)$$

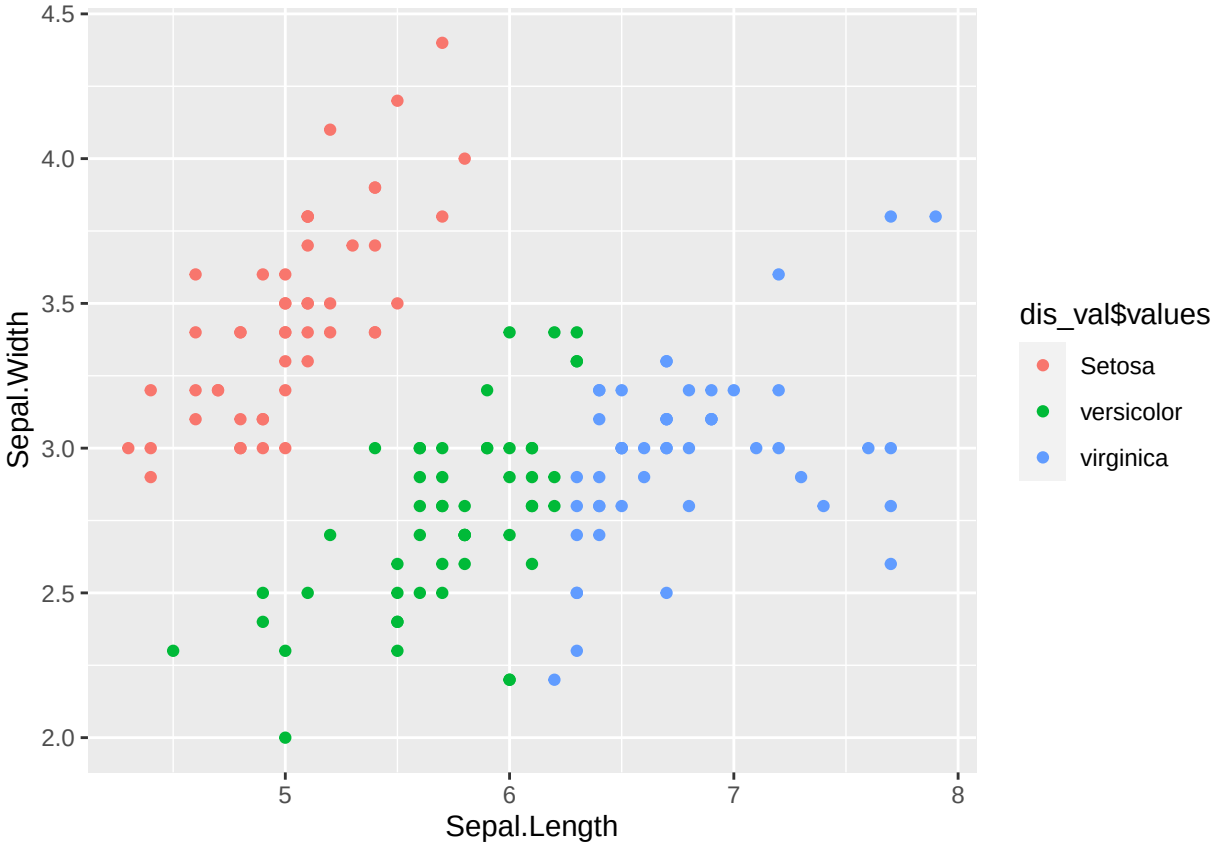
Discriminant function

```
disc_function <- function(class_data,class_mean,class_prior,class_covariance)
{
  (as.matrix(class_data[1:2]) %*% solve(class_covariance) %*% class_mean) -
  rep((0.5*(class_mean[1:2]%*%solve(class_covariance)%*%class_mean)),nrow(class_data))+rep(log(class_pr

}

setosa_disc <- disc_function(data,setosa_mean,setosa_prior,pooled_matrix)
versicolor_disc <-  disc_function(data,versicolor_mean,versicolor_prior,pooled_matrix)
virginica_disc <- disc_function(data,virginica_mean,virginica_prior,pooled_matrix)
```

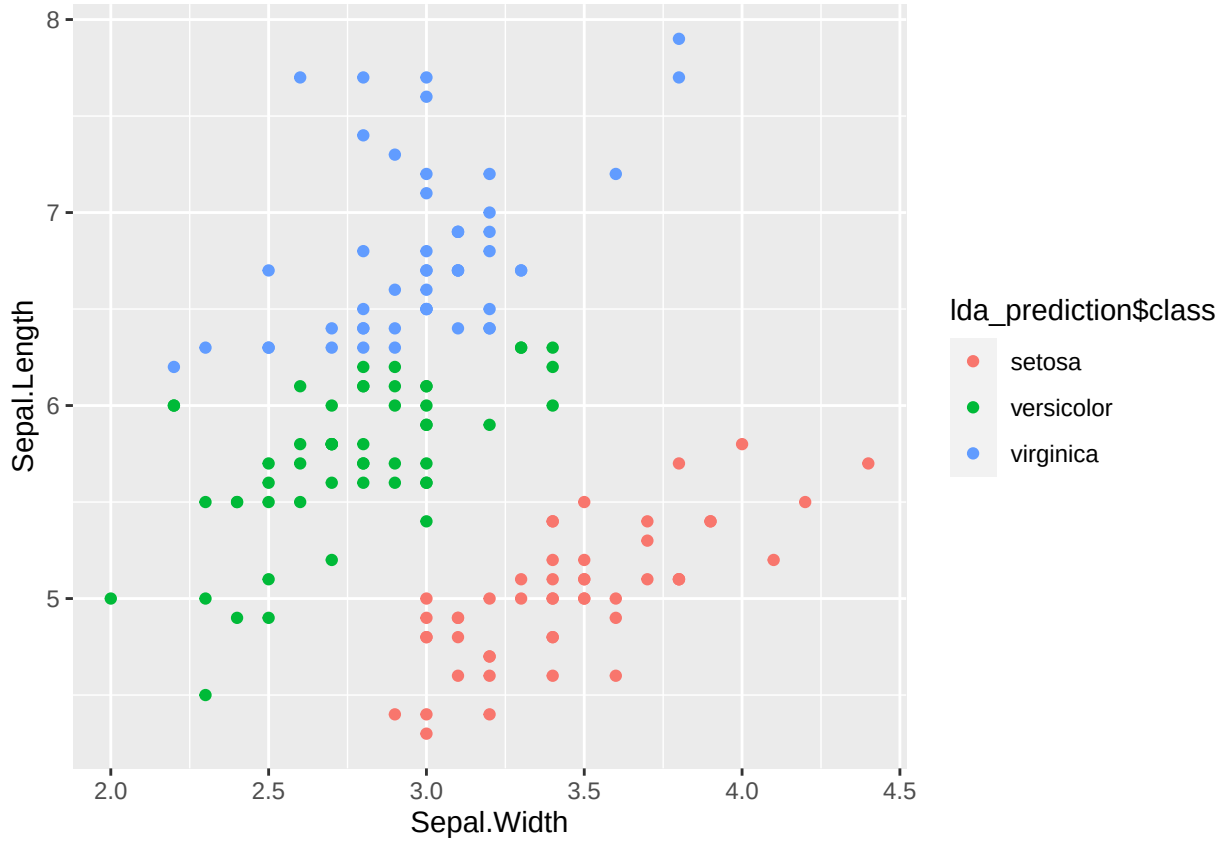
Decision boundaries

$$Log(\pi_l/\pi_k) - 1/2(\mu_k + \mu_l)^T \sum(\mu_k - \mu_l) + x^T \sum^{-1}(\mu_k - \mu_l)$$

Predicting using dicriminant function



Prediction using LDA



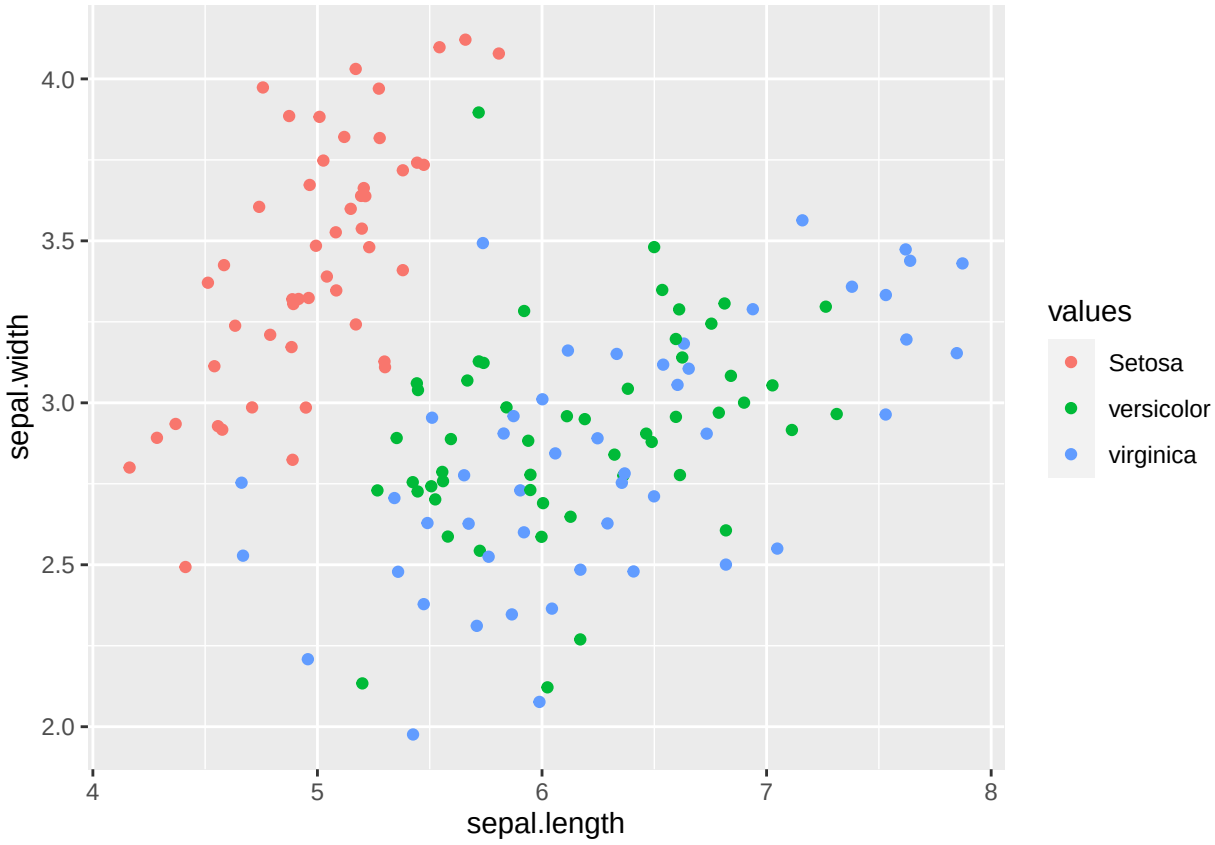
Confusion matrix using LDA function

```
##
##              setosa versicolor virginica
##   setosa         49          0         0
##   versicolor      1         36        15
##   virginica       0         14        35
```

Misclassification error using LDA function

```
## [1] 0.2
```

Generatin new data with discriminant function



Prediction using logistic regression

```
## # weights:  12 (6 variable)
## initial  value 164.791843
## iter  10 value 62.715967
## iter  20 value 59.808291
## iter  30 value 55.445984
## iter  40 value 55.375704
## iter  50 value 55.346472
## iter  60 value 55.301707
## iter  70 value 55.253532
## iter  80 value 55.243230
## iter  90 value 55.230241
## iter 100 value 55.212479
## final  value 55.212479
## stopped after 100 iterations
```

Confusion matrix

```
##
##              setosa versicolor virginica
##   setosa         50          0         0
```

```
##   versicolor      0         38        13
##   virginica       0         12        37
```

Misclassification error

```
## [1] 0.1666667
```