Data-sets used

Digit Dataset

Iris Dataset

Logistic Regression

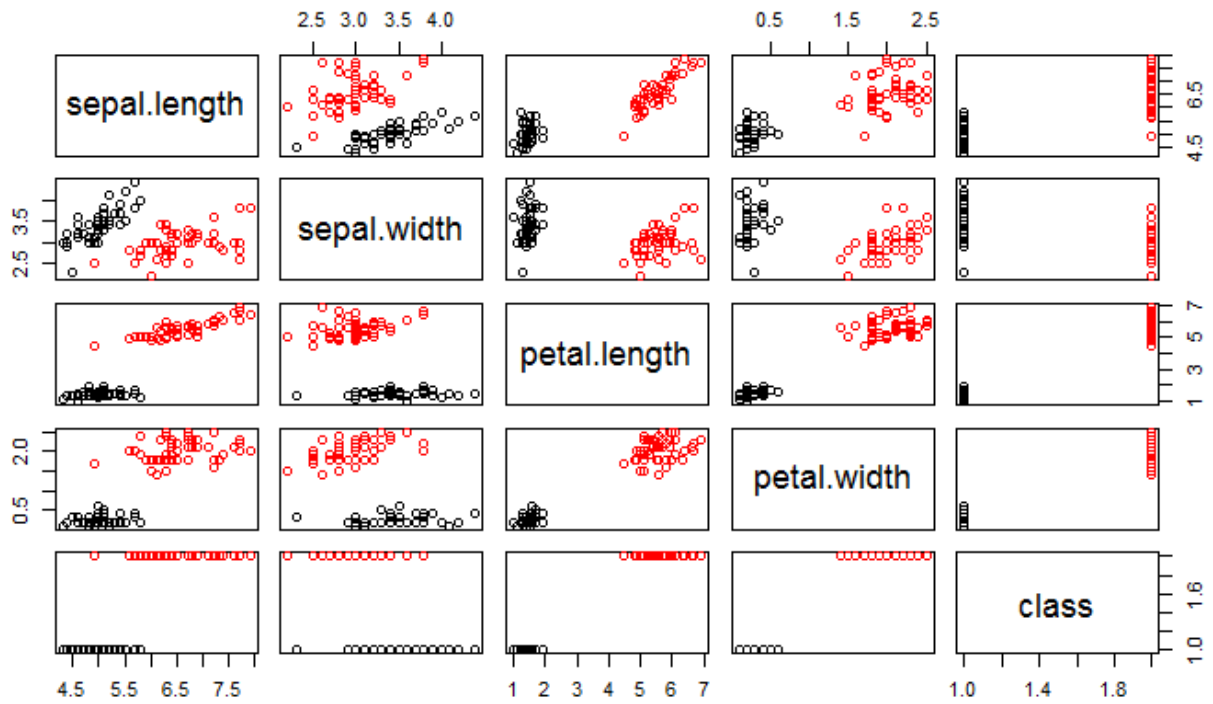
In this we perform logistic regression on our Datasets

Iris Dataset

1. Implement the logistic regression algorithm for two-class discrimination.

Data-set Iris

Plot Data



The two class seems to be pretty much separate. Let's implement logistic regression on it.

We are going to run the logistic regression for 500 steps and with learning rate 0.001

We stop if there is no much change in theta values (0.00001 difference)

Coefficients

```
$`Iris-setosa`
                      29
sepal.length  0.5648780713
sepal.width   1.3350987960
petal.length -2.0004793421
petal.width  -1.0101234148
theta_0       0.3182478349

$`Iris-virginica`
                      29
sepal.length -0.5656285414
sepal.width  -1.3344881979
petal.length  2.0003279285
petal.width   1.0113073135
theta_0      -0.3166929130
```

As we can see from above sepal width seems to be good predictor of Iris-setosa and petal length seems to be good predictor of Iris-Virginia

Evaluation

```
$confusion_matrix
                truelabel
prediction       Iris-setosa Iris-virginica
  Iris-setosa             13              0
  Iris-virginica           0             12
```

```
$accuarcy
[1] 1

$percision
   Iris-setosa Iris-virginica
             1              1

$recall
   Iris-setosa Iris-virginica
             1              1

$fmeasure
   Iris-setosa Iris-virginica
             1              1
```

As there is a clear separation of variables in our dataset we have 100% in all evaluations

5-fold Cross Validation

Let's do 5 fold cross validation to test our model further

```
$avgAccuracy
[1] 1

$avgPercision
   Iris-setosa Iris-virginica
             1              1

$avgRecall
   Iris-setosa Iris-virginica
             1              1

$avgFmeasure
   Iris-setosa Iris-virginica
             1              1
```

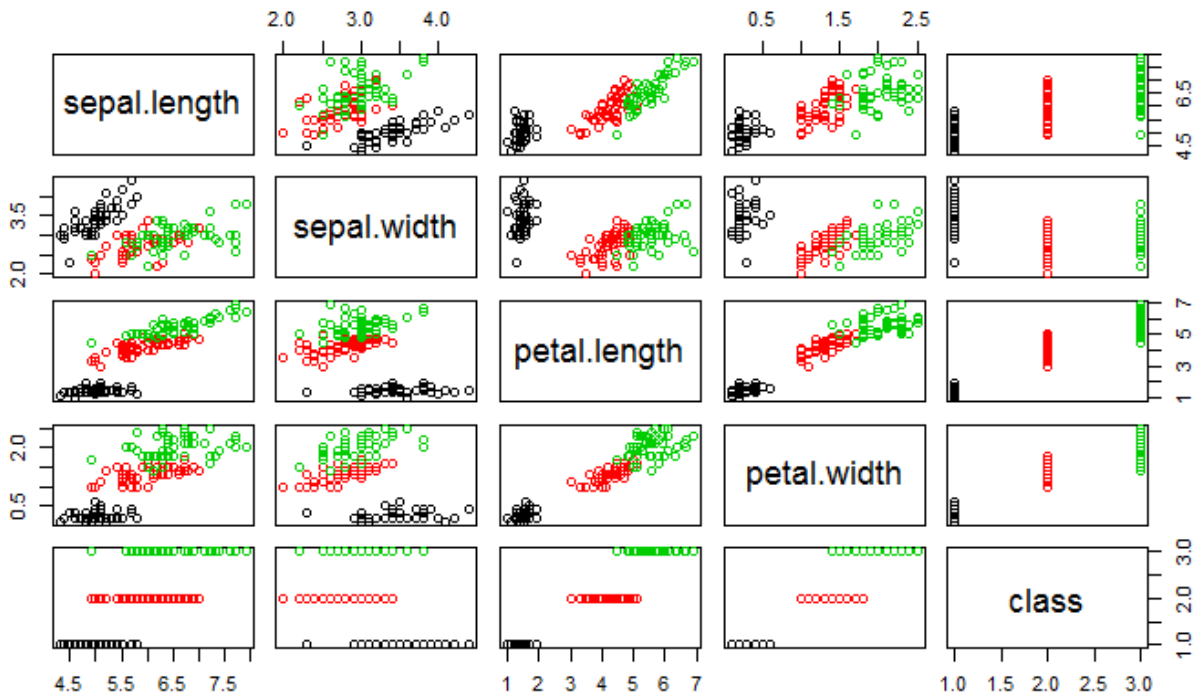
So in our cross validation also we get same result as before

3. Implement the logistic regression algorithm for K-class discrimination.

Data-set Iris

Number of classes 3

Plot Data-set



We are going to run the logistic regression for 500 steps and with learning rate 0.001

We stop if there is no much change in theta values (0.00001 difference)

Coefficients

```
$`Iris-setosa`
                          43
sepal.length  0.4488581428
sepal.width   1.5986167818
petal.length -2.4740212555
petal.width  -1.1305198608
theta_0       0.2841501169

$`Iris-versicolor`
                          43
sepal.length  0.4039267901
sepal.width  -1.3706302477
petal.length  0.4775172896
petal.width  -1.0507511957
theta_0       0.5108465848

$`Iris-virginica`
                          43
sepal.length -1.6163881221
sepal.width  -1.5176042741
petal.length  2.3840254426
petal.width   2.1787602488
theta_0      -0.8352562783
```

Evaluation

```
$confusion_matrix
                truelabel
prediction       Iris-setosa Iris-versicolor Iris-virginica
  Iris-setosa             13               0              0
  Iris-versicolor          0              10              0
  Iris-virginica           0               1             14

$accuarcy
[1] 0.9736842105

$percision
    Iris-setosa Iris-versicolor  Iris-virginica
   1.0000000000    0.9090909091   1.0000000000

$recall
    Iris-setosa Iris-versicolor  Iris-virginica
   1.0000000000    1.0000000000   0.9333333333


$fmeasure
    Iris-setosa Iris-versicolor  Iris-virginica
   1.0000000000    0.9523809524   0.9655172414
```

5-Fold Cross Validation

```
$avgAccuracy
[1] 0.9597701149

$avgPercision
    Iris-setosa Iris-versicolor  Iris-virginica
           1.00            0.88            1.00

$avgRecall
    Iris-setosa Iris-versicolor  Iris-virginica
    1.000000000     1.000000000     0.896969697

$avgFmeasure
    Iris-setosa Iris-versicolor  Iris-virginica
   1.0000000000    0.9345029240   0.9445887446
```