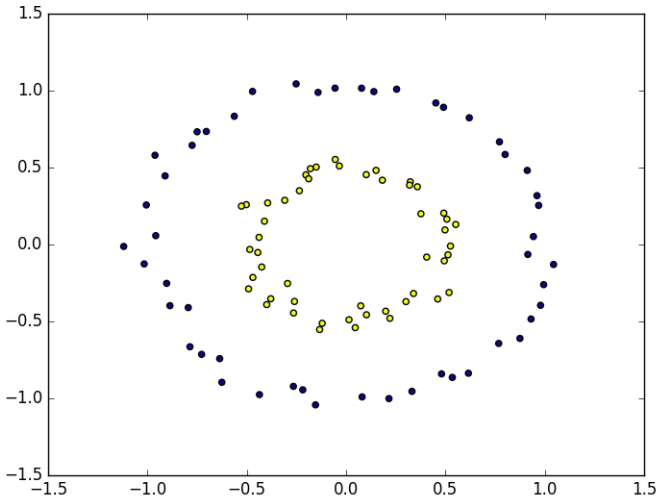
ML Assignment 2

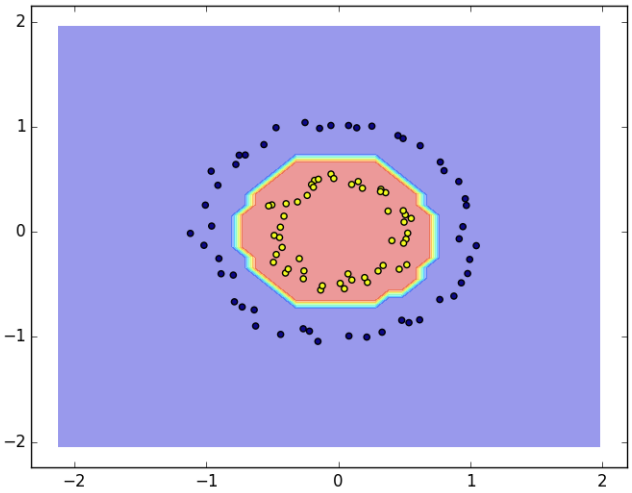
Kushagra Arora

September 2017

1 Visualisation

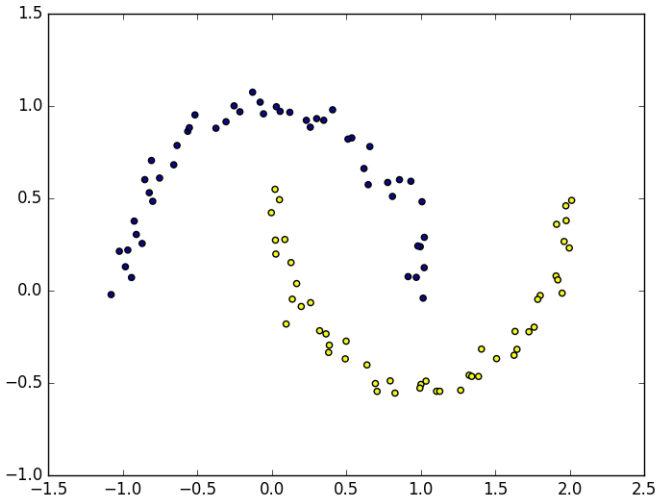
1. **Date_1.h5**

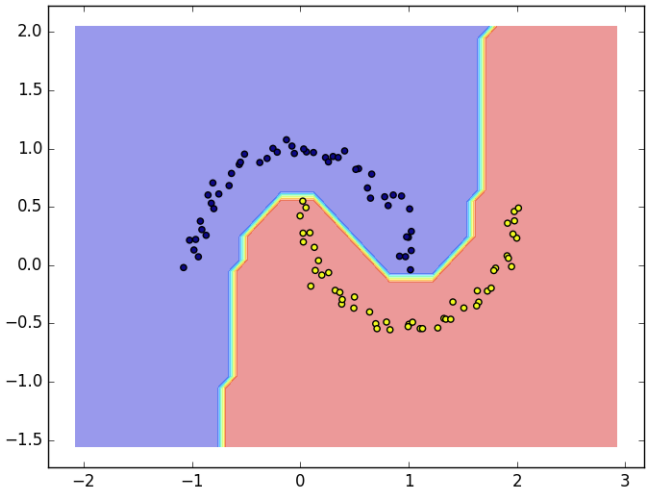




The above data are simply two concentric circles. The decision boundary shows that there is not much noise there is not much noise. The kernel used here is rbf since a circle with plotted in a large number of dimensions.

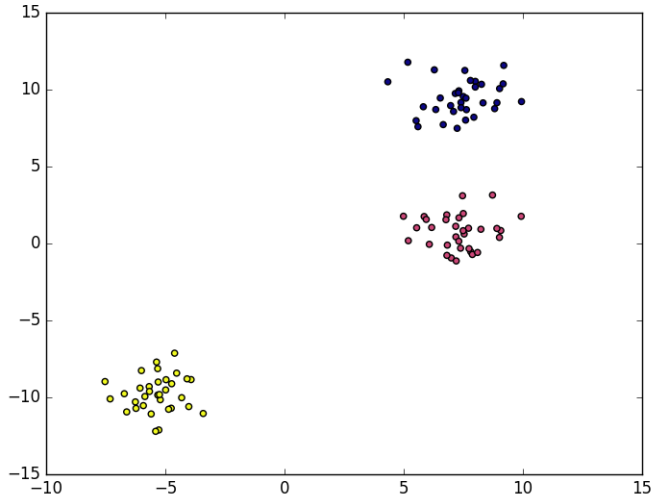
2. **Data_2.h5**

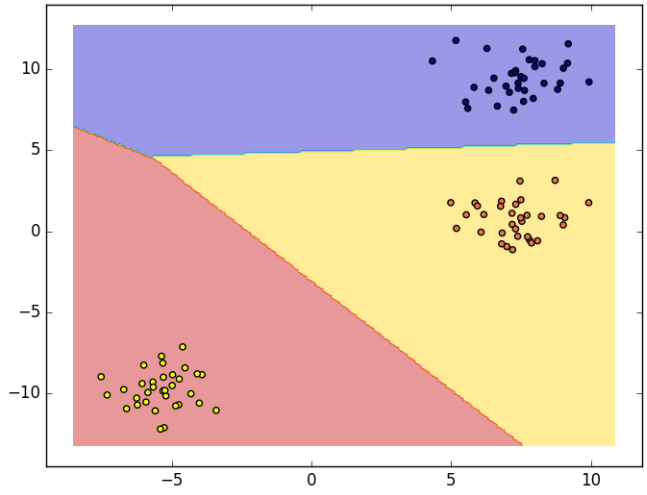




The kernel used is rbf. This can also be done using a polynomial kernel of a dimension greater than equal to 3. The data is clean and noiseless and there are no outliers. $\sigma = 0.8$ in rbf kernel.

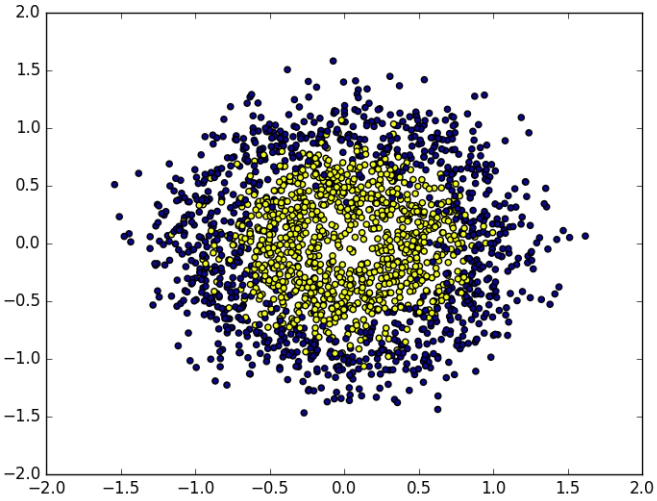
3. **Data_3.h5**

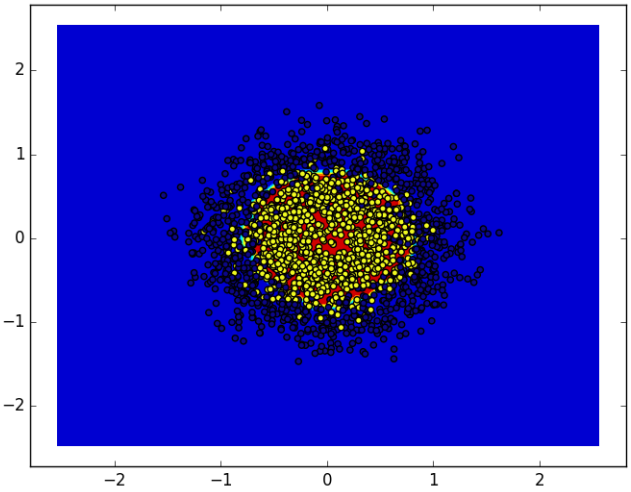




The kernel used is linear. The data is clean and there are no outliers.

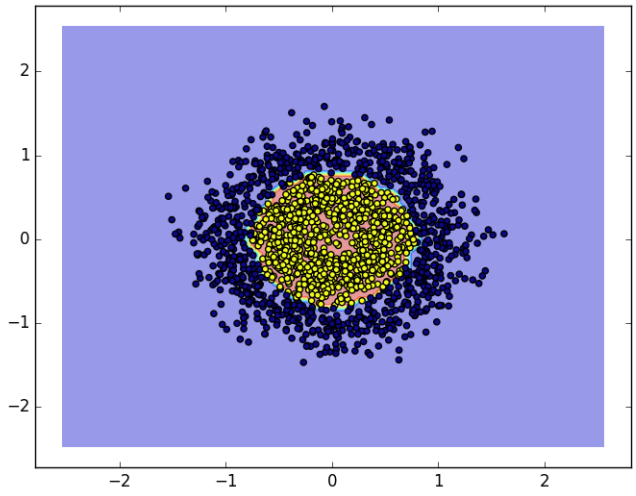
4. **Data_4.h5**



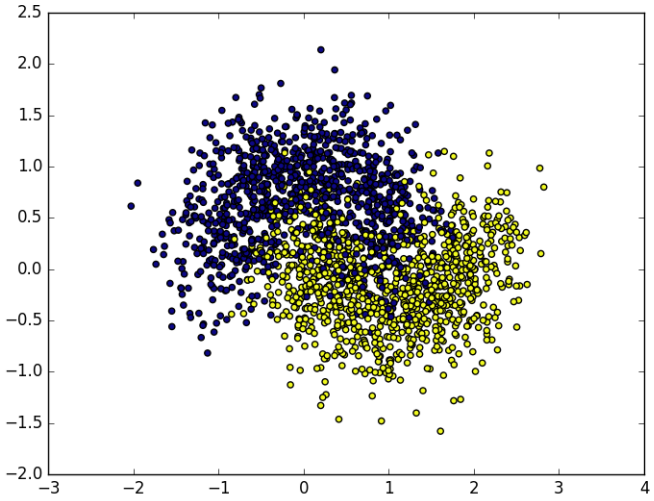


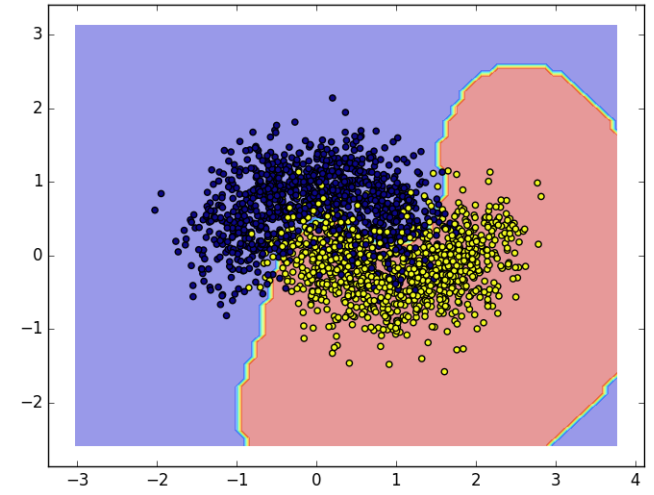
The kernel used in rbf. The data has outliers.

After outlier removal



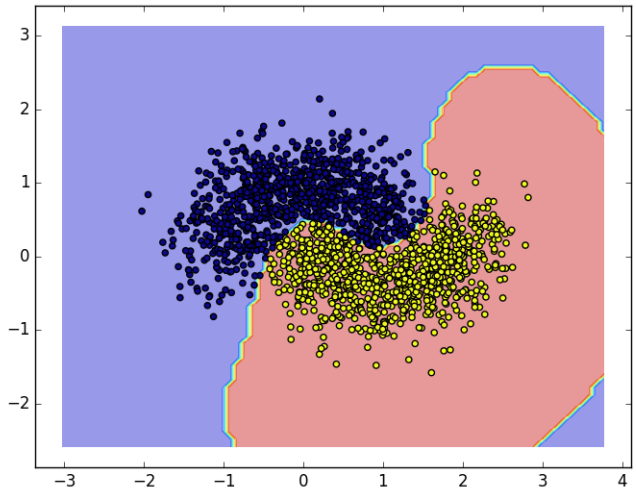
5. **Data_5.h5**





The kernel used in rbf. The data has outliers.

After outlier removal



2 Analysing Data Sets

- **Hyperparamter tuning**

  The hyperparameters are searched using grid search on C and *gamma*.

  Evaluation metric : Acccuracy score.

  Classifer used : One Vs Rest

  1. Data_1
     - Kernel = linear
       Best parameter = C = 0.5
       Accuracy = 42.42
     - Kernel = rbf
       Best paramter = C = 0.3, gamma = 1.0
       Accuracy = 48.4848
  2. Data_2
     - Kernel = linear
       Best parameter = C = 0.3
       Accuracy = 81.81
     - Kernel = rbf
       Best paramter = C = 0.3, gamma = 1.0
       Accuracy = 69.7
  3. Data_3
     - Kernel = linear
       Best parameter = C = 0.1
       Accuracy = 100.00
     - Kernel = rbf
       Best paramter = C = 0.1, gamma = 0.5
       Accuracy = 75.75
  4. Data_4
     - Kernel = linear
       Best parameter = C = 0.1
       Accuracy = 45.75
     - Kernel = rbf
       Best paramter = C = 0.1, gamma = 0.5
       Accuracy = 48.4
  5. Data_5
     - Kernel = linear
       Best parameter = C = 0.5
       Accuracy = 83.93
     - Kernel = rbf
       Best paramter = C = 0.1, gamma = 1.0
       Accuracy = 77.57
  6. Part_A_train
     - Kernel = linear
       Best parameter = C = 0.3
       Accuracy = 81.81

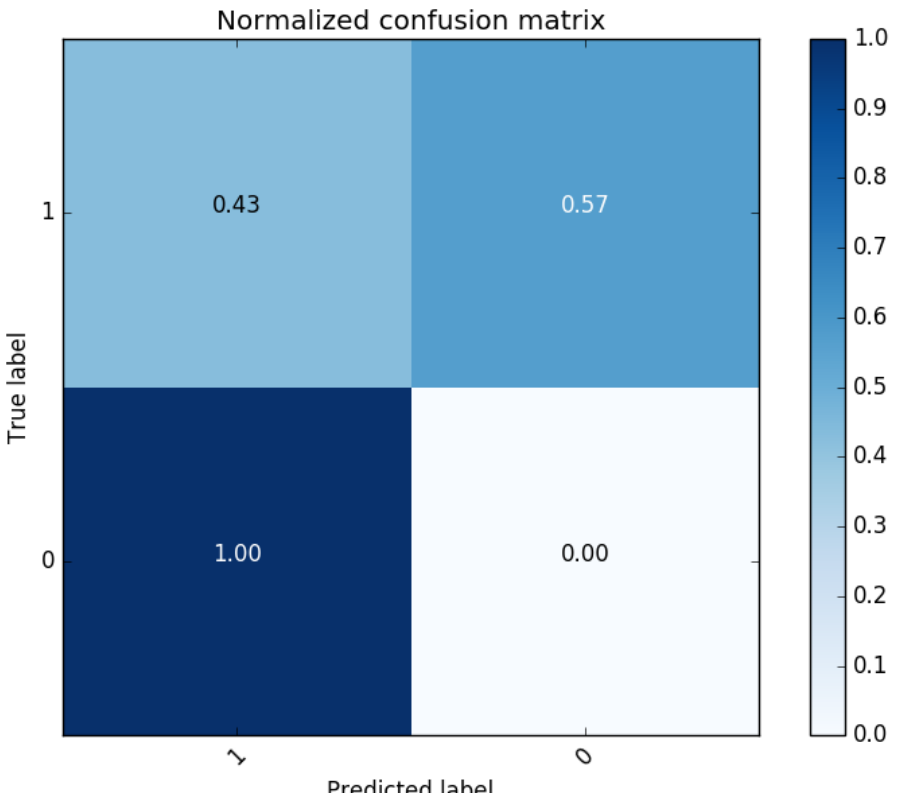
- Kernel = rbf
  Best paramter = C = 0.3, gamma = 1.0
  Accuracy = 69.7

Classifier used : One Vs One

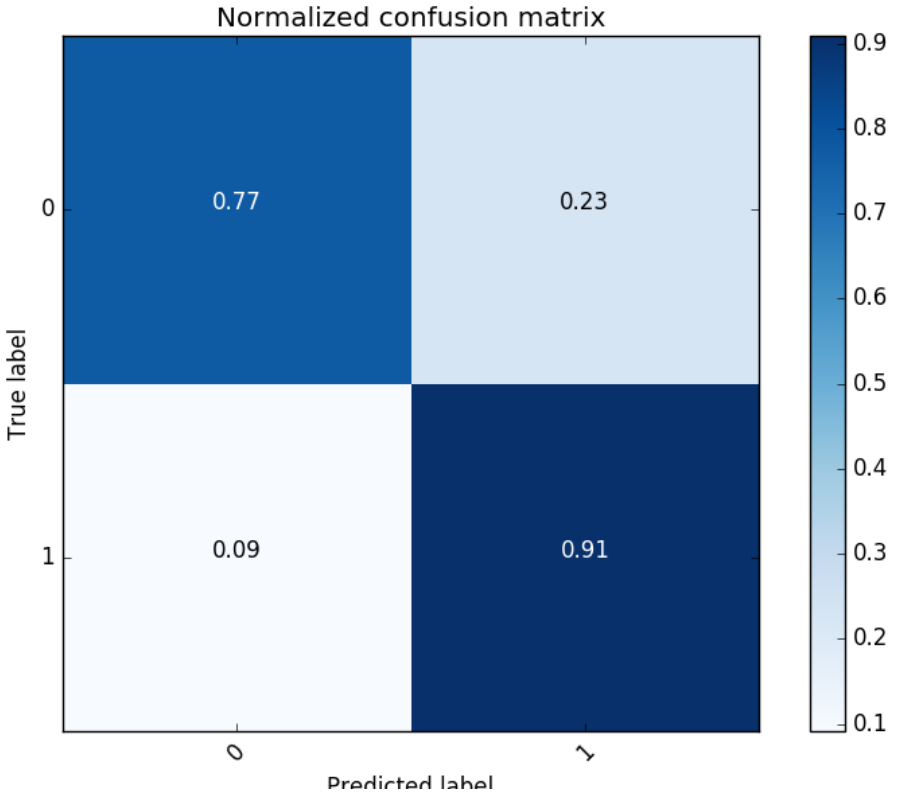
1. Data_1
   - Kernel = linear
     Best parameter = C = 0.5
     Accuracy = 42.75
   - Kernel = rbf
     Best paramter = C = 0.3, gamma = 1.0
     Accuracy = 48.4848
2. Data_2
   - Kernel = linear
     Best parameter = C = 0.3
     Accuracy = 81.81
   - Kernel = rbf
     Best paramter = C = 0.3, gamma = 1.0
     Accuracy = 69.7
3. Data_3
   - Kernel = linear
     Best parameter = C = 0.1
     Accuracy = 100.00
   - Kernel = rbf
     Best paramter = C = 0.1, gamma = 0.5
     Accuracy = 75.75
4. Data_4
   - Kernel = linear
     Best parameter = C = 0.1
     Accuracy = 45.75
   - Kernel = rbf
     Best paramter = C = 0.1, gamma = 0.5
     Accuracy = 48.4
5. Data_5
   - Kernel = linear
     Best parameter = C = 0.5
     Accuracy = 83.93
   - Kernel = rbf
     Best paramter = C = 0.1, gamma = 1.0
     Accuracy = 77.57
6. Part_A_train
   - Kernel = linear
     Best parameter = C = 0.3
     Accuracy = 81.81
   - Kernel = rbf
     Best paramter = C = 0.3, gamma = 1.0
     Accuracy = 69.7

- **Confusion Matrix**

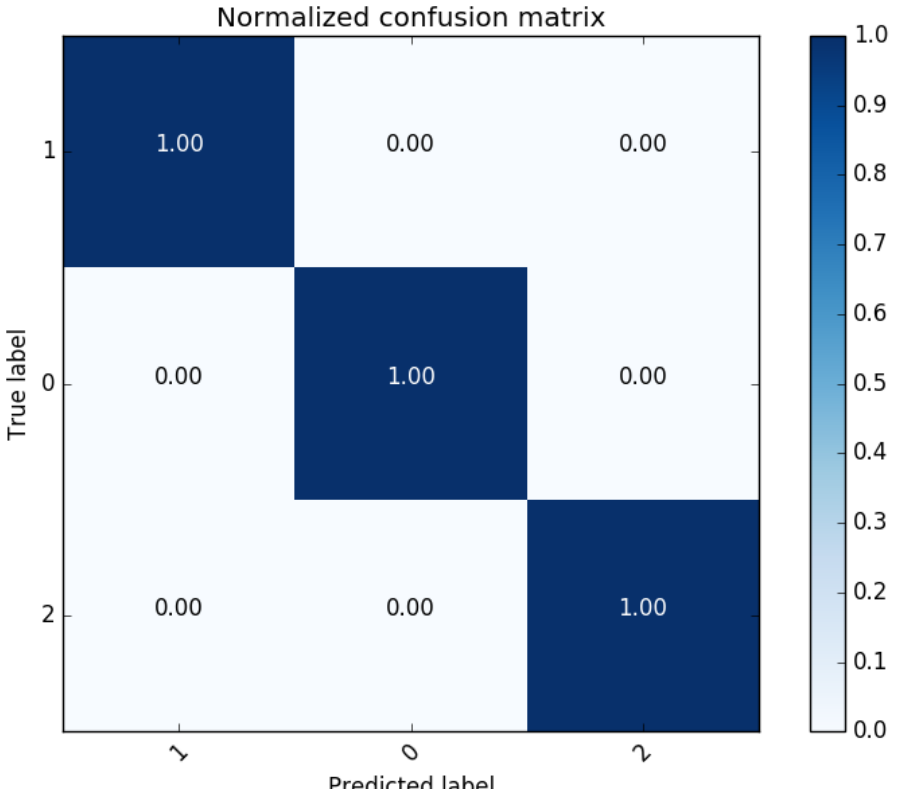
1. Data_1



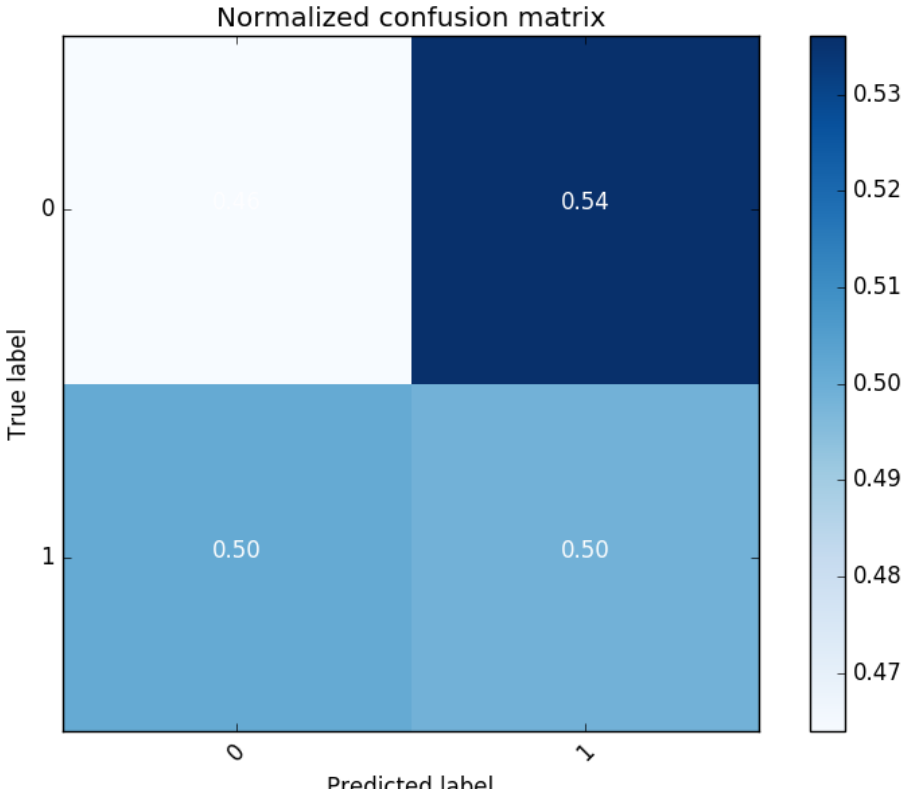
2. Data_2


Normalized confusion matrix
0.77
0.23
0.09
0.91
0
1
True label
0
1
Predicted label
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

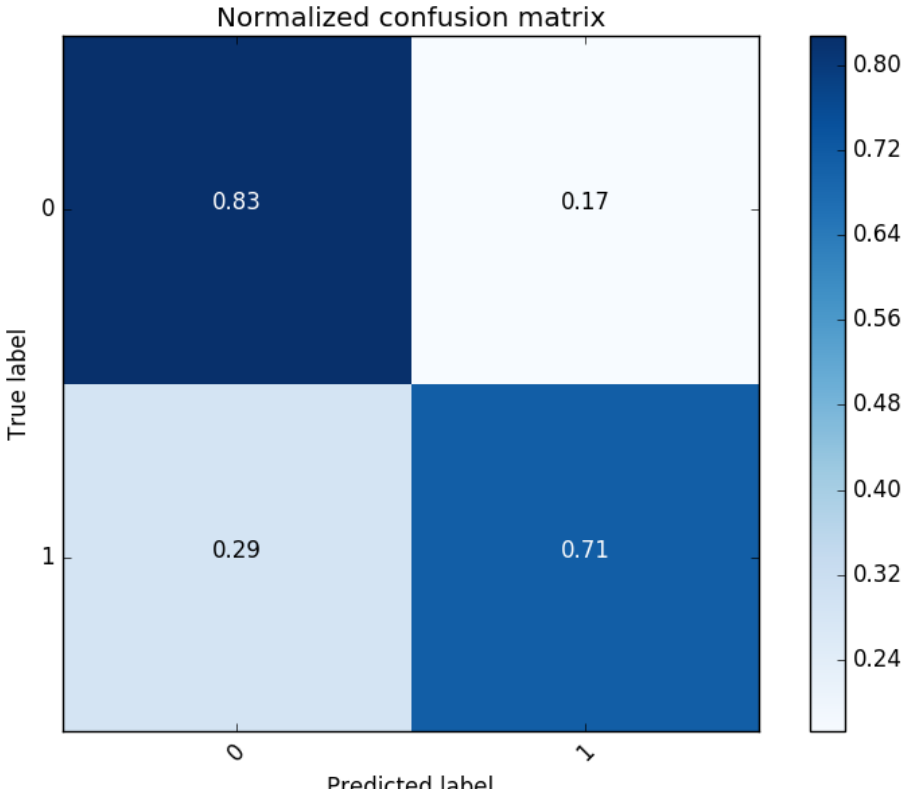
3. Data_3


Normalized confusion matrix
1.00
0.00
0.00
0.00
1.00
0.00
0.00
0.00
1.00
True label
1
0
2
1
0
2
Predicted label
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

4. Data_4


Normalized confusion matrix
0.46
0.54
0.50
0.50
True label
0
1
0
1
Predicted label
0.53
0.52
0.51
0.50
0.49
0.48
0.47

5. Data_5


Normalized confusion matrix
0.83
0.17
0.29
0.71
0
1
True label
0
1
Predicted label
0.80
0.72
0.64
0.56
0.48
0.40
0.32
0.24