SML Assignment - 2

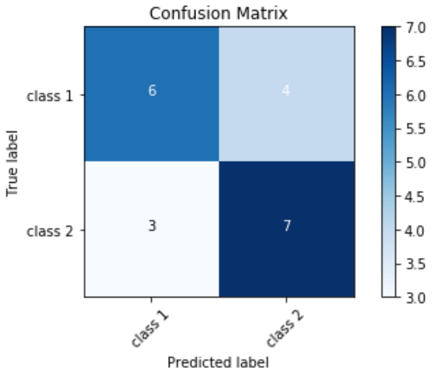
Kaustav Vats (2016048)

Question 1 (Book Questions)

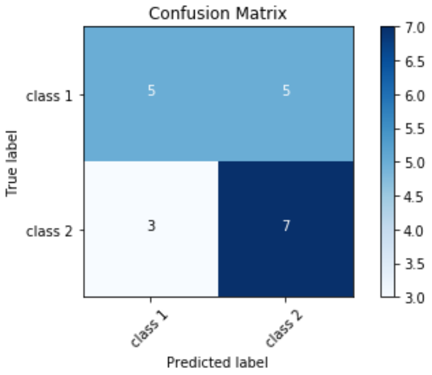
Classes- W1, W2

Confusion Matrix

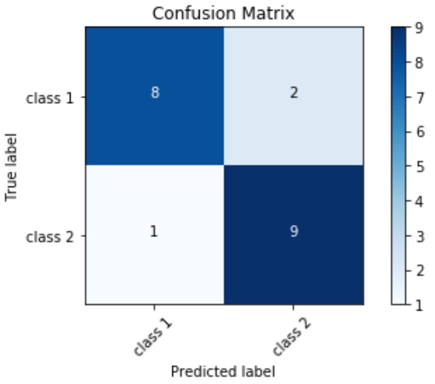
For feature X1 only



For feature X1 and X2



For feature X1, X2 and X3



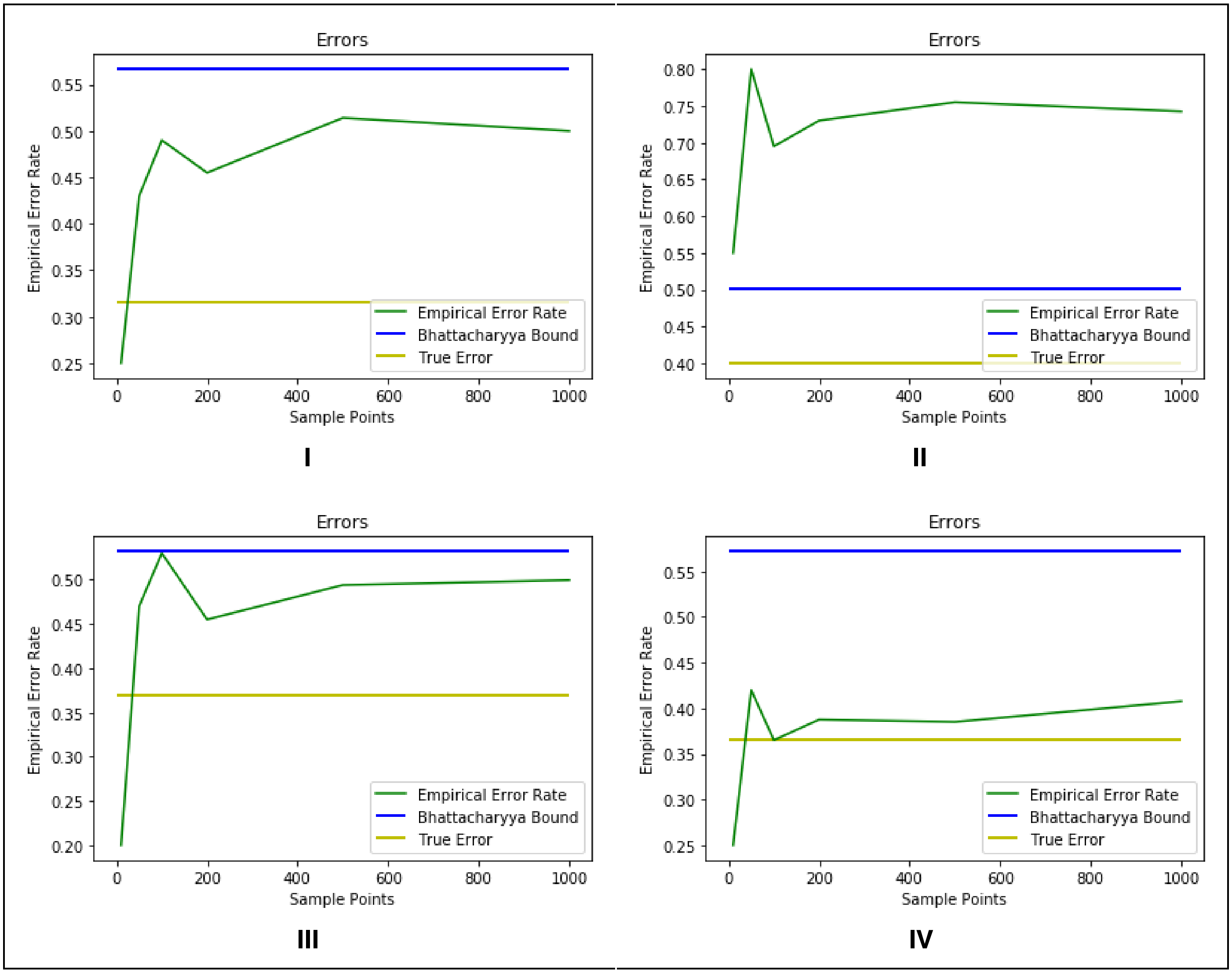
Features\Evaluation Metric	Training Accuracy	Empirical Error	Bhattacharyya Bound
X1	65.0	35.0	0.473996
X1 and X2	60.0	40.0	0.459847
X1, X2 and X3	85.0	15.0	0.411357

In particular, is it ever possible for a finite set of data that the empirical error might be larger for more data dimensions?
Not necessarily, Increasing the sample size will affect the distribution, It might change in such a way that distribution is much better separated than before.

Question 2 & 3 Combined (Book Questions)

Normal Distribution\ Evaluation Metric	Avg Training Accuracy	Empirical Error	Bhattacharyya Bound	True Error
W1 N(-0.5, 1) | W2 N(0.5, 1) P(w1) == P(w2)	56.01	43.99	0.566574	0.315975
W1 N(-0.5, 2) | W2 N(0.5, 2) P(w1) = 2/3& P(w2) = 1/3	28.79	71.21	0.501807	0.400435
W1 N(-0.5, 2) | W2 N(0.5, 2) P(w1) == P(w2)	55.85	44.15	0.532247	0.369875
W1 N(-0.5, 3) | W2 N(0.5, 1) P(w1) == P(w2)	63.08	36.92	0.571936	0.365817

Error Curves



I

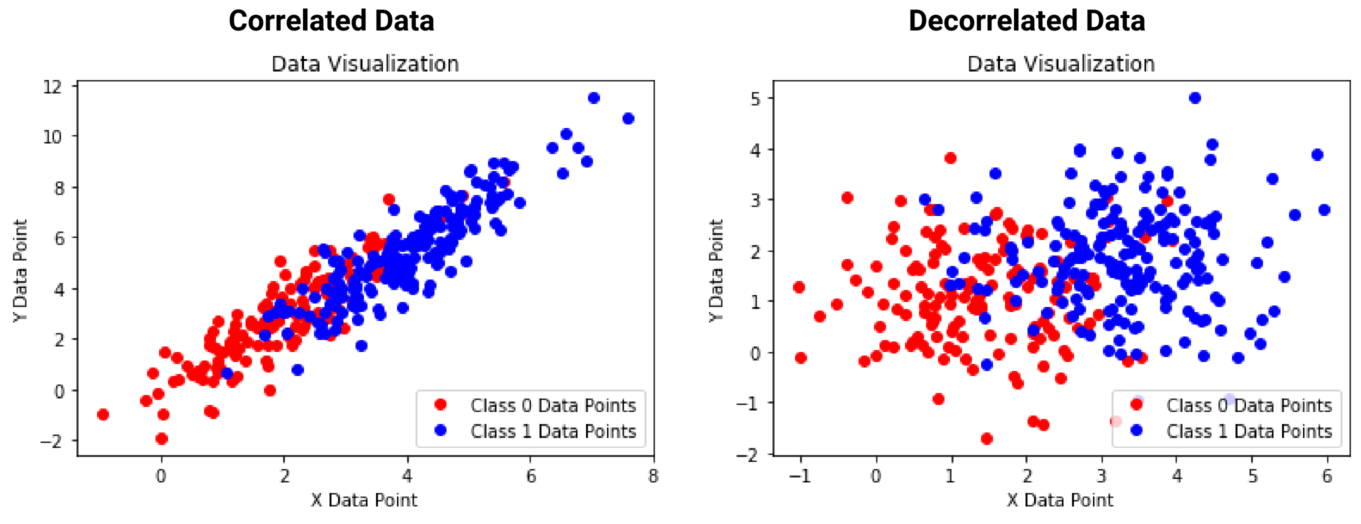
II

III

IV

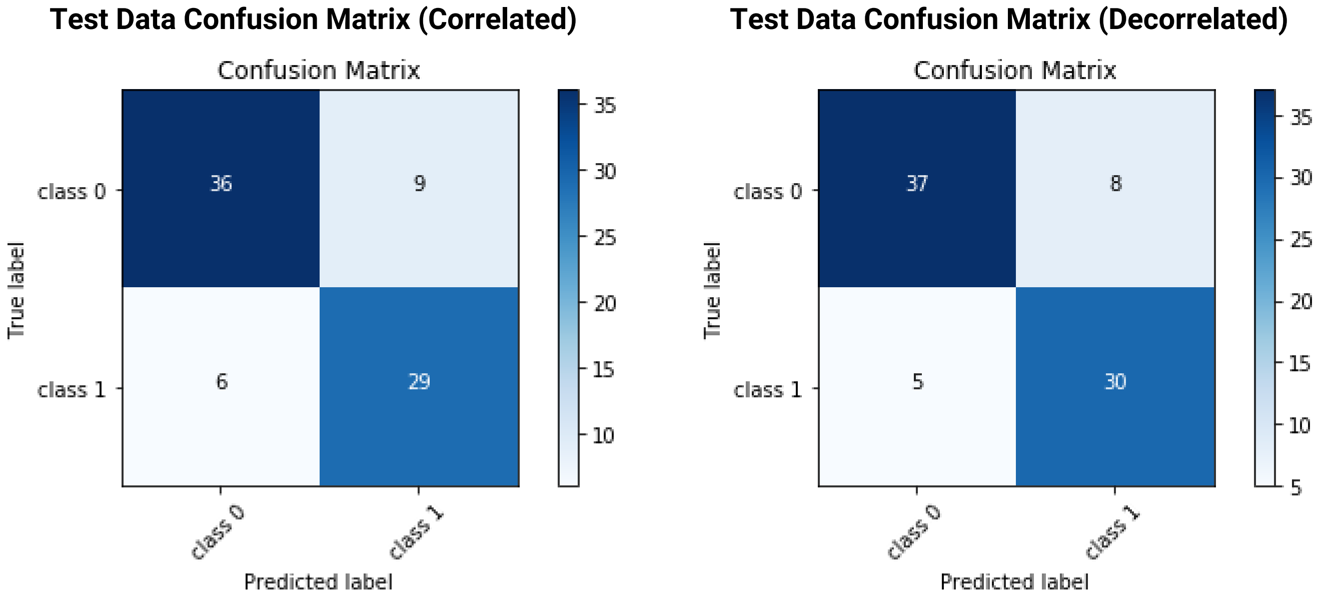
Part B

Data Visualization


Correlated Data

Decorrelated Data

Confusion Matrix

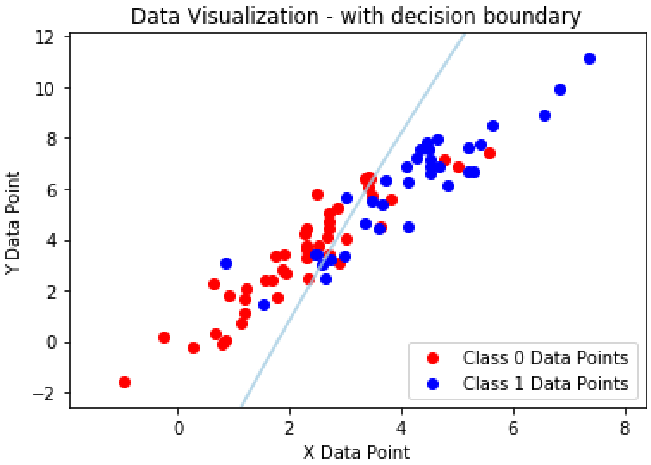

Test Data Confusion Matrix (Correlated)

Test Data Confusion Matrix (Decorrelated)

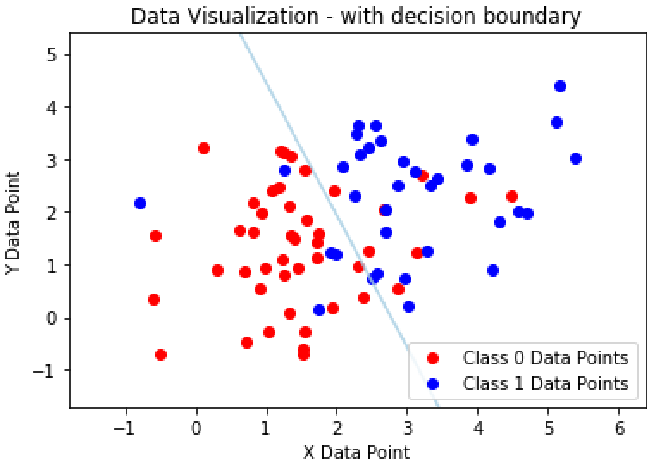
Evaluation Metric	Correlated Test Data	Decorrelated Test Data
Validation Accuracy	83.71	84.57
Testing Accuracy	81.25	83.75

Decision Boundary

Decision Boundary Correlated

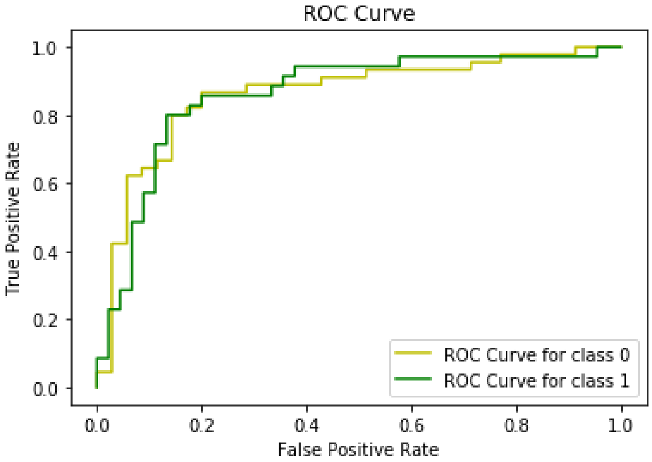


Decision Boundary with Decorrelated

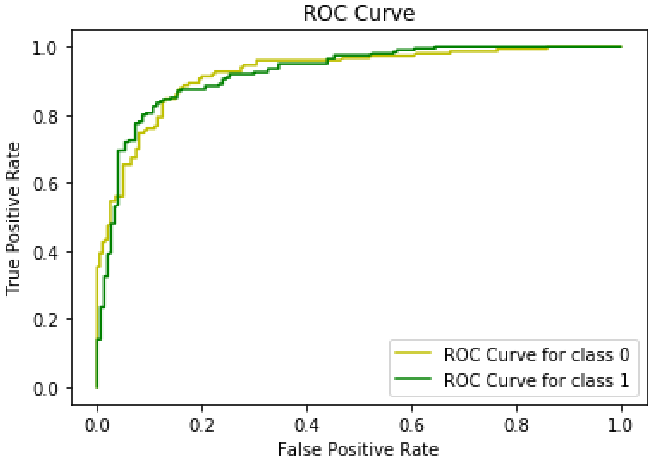


ROC Curves

ROC Curve

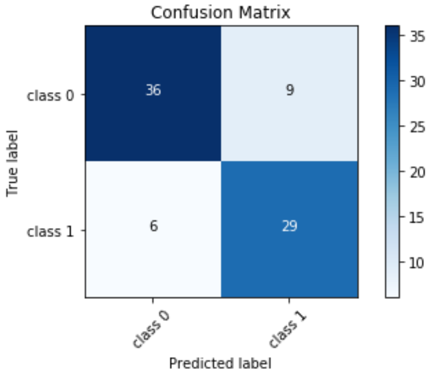
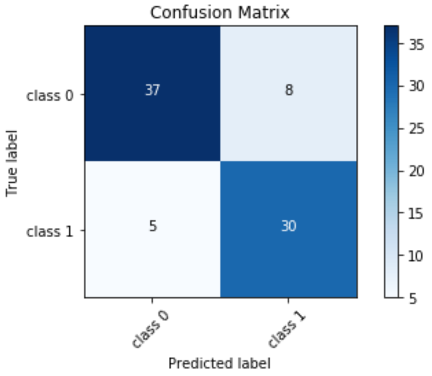
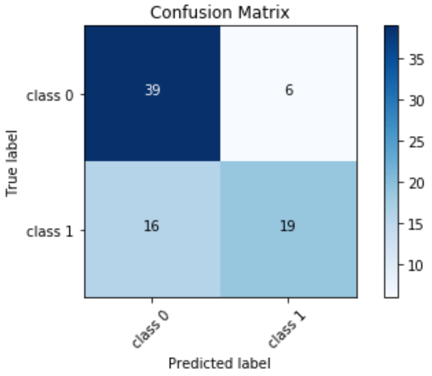
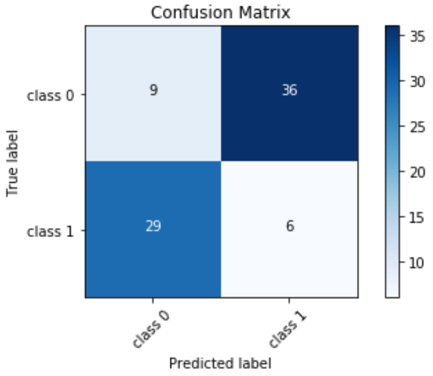
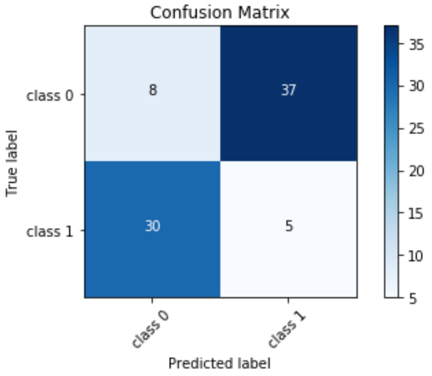
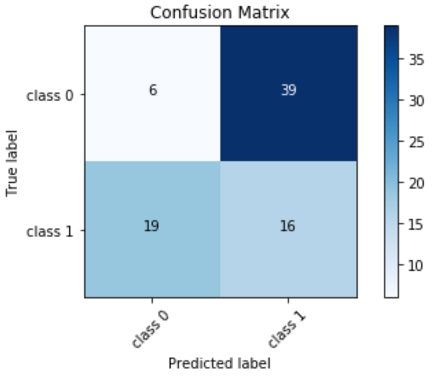
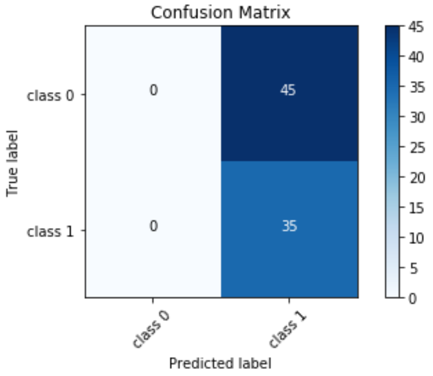
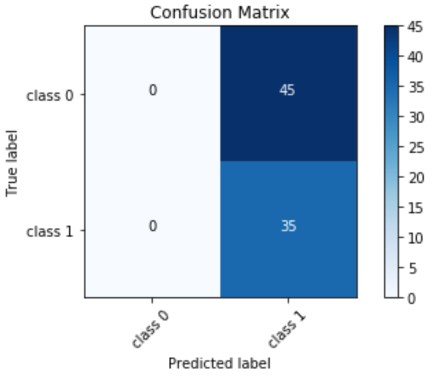
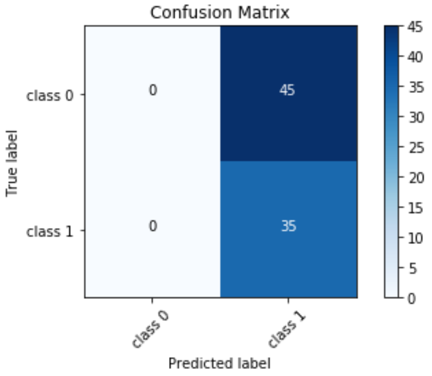


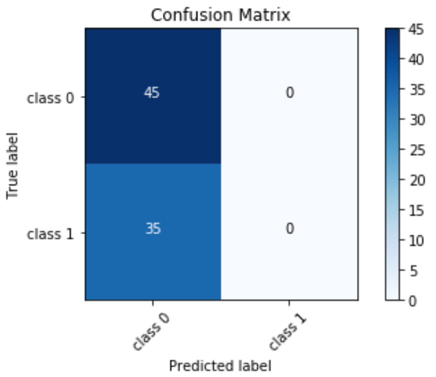
ROC Curve with Decorrelated Test Data

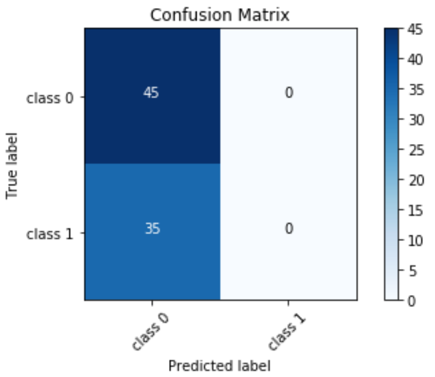


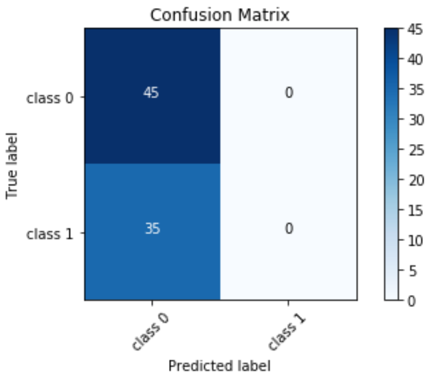
Decorrelated data is giving much better result. False positives have decreased after doing decorrelation as compared to correlated data. Data points are much better separated than before.

Risk Matrix Confusion Matrix Testing Data (Correlated) , (Decorrelated) and Missing points

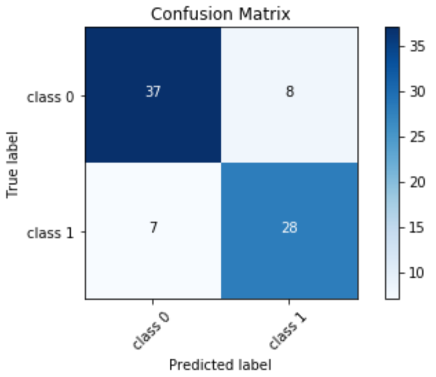
Risk 1  	**Risk 1**  	**Risk 1** 
Risk 2  	**Risk 2**  	**Risk 2** 
Risk 3  	**Risk 3**  	**Risk 3** 
Risk 4	**Risk 4**	**Risk 4**



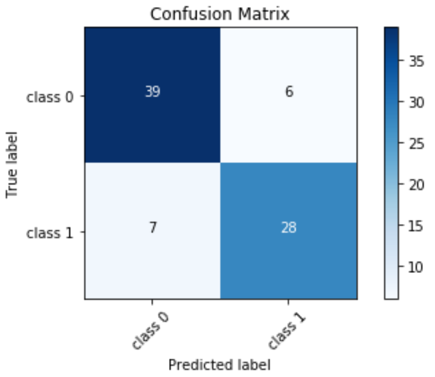




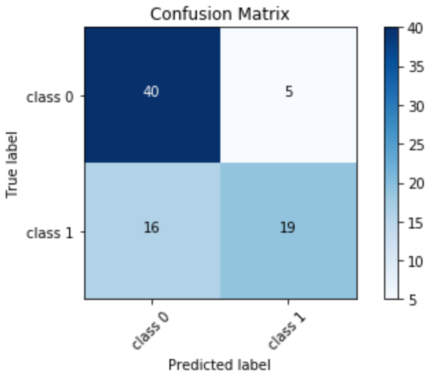
Risk 5



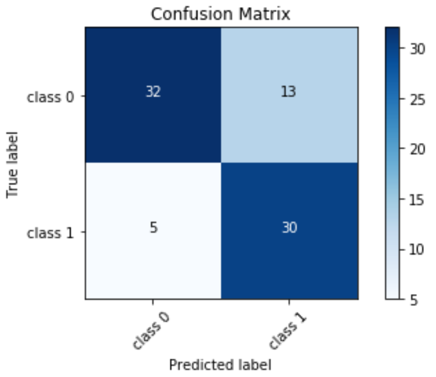
Risk 5



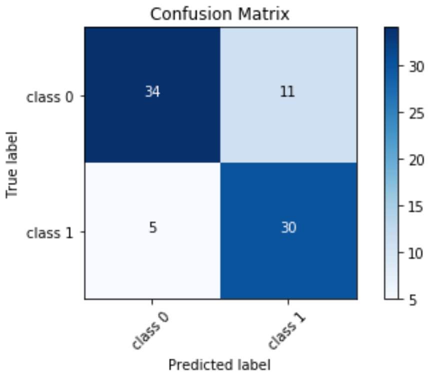
Risk 5



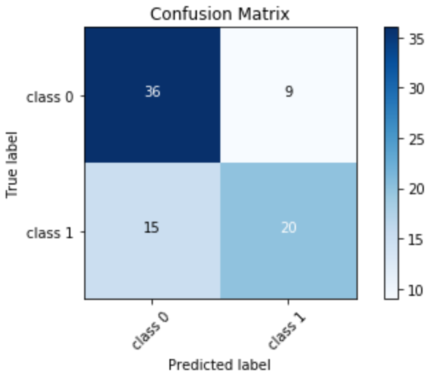
Risk 6



Risk 6



Risk 6



For Missing Values, it seems like good points were more towards one Normal Distribution.

Evaluation Metric

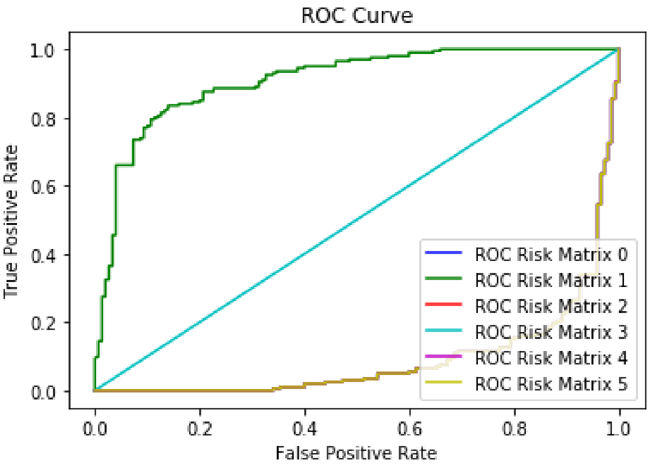
Accuracy	**Risk 1**	**Risk 2**	**Risk 3**	**Risk 4**	**Risk 5**	**Risk 6**
Correlated	81.25	18.75	43.75	56.25	81.25	77.5
Decorrelated	83.75	16.25	43.75	56.25	83.75	80.0
Missing Points	72.5	27.5	43.75	56.25	73.75	70.0

Missing Test Points

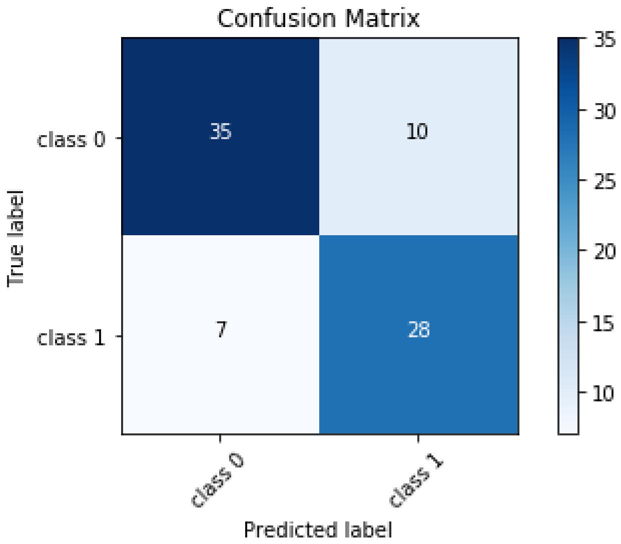
Accuracy	**78.75**

Roc with Risk Matrix Incorporation

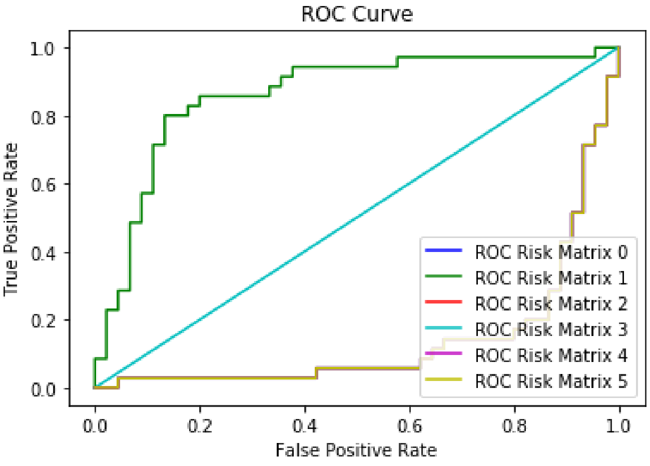
Roc of Train Data



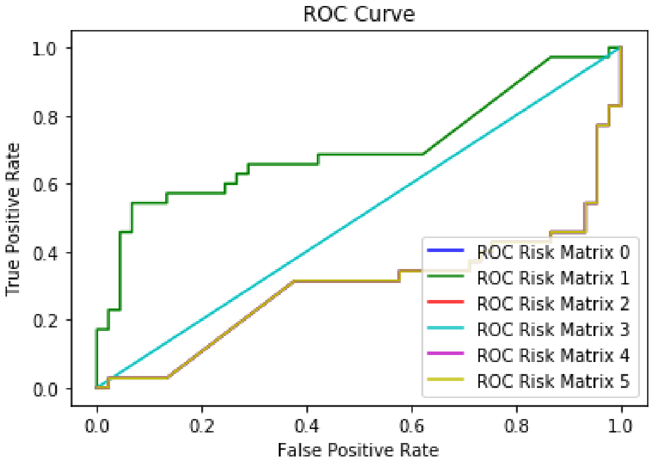
Confusion Matrix



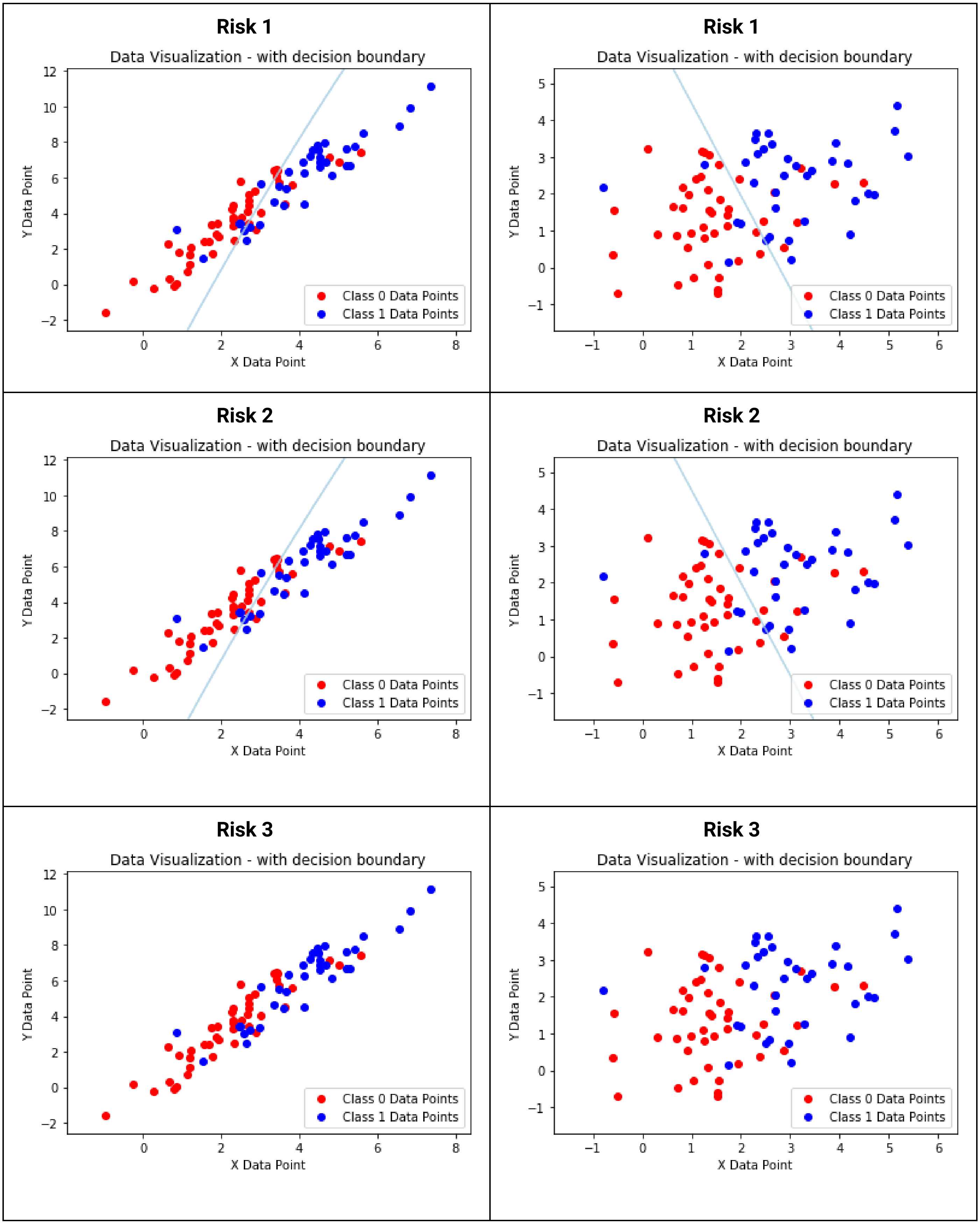
Roc of Test Data



Roc of Missing Test Data



Decision Boundary After incorporating Risk Matrix with Correlated Data and Decorrelated Data



Risk 1

Data Visualization - with decision boundary

Risk 1

Data Visualization - with decision boundary

Risk 2

Data Visualization - with decision boundary

Risk 2

Data Visualization - with decision boundary

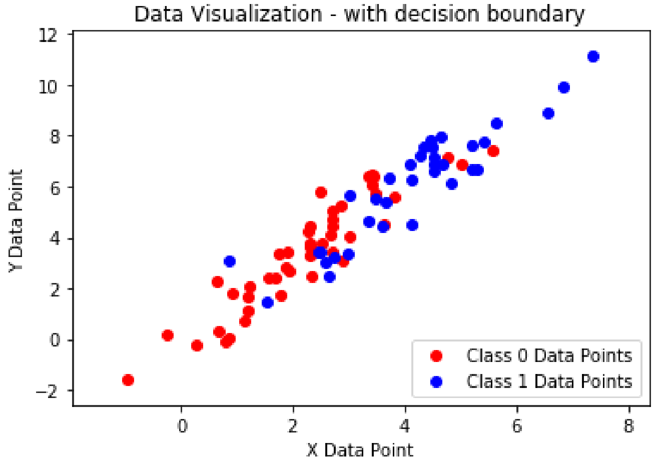
Risk 3

Data Visualization - with decision boundary

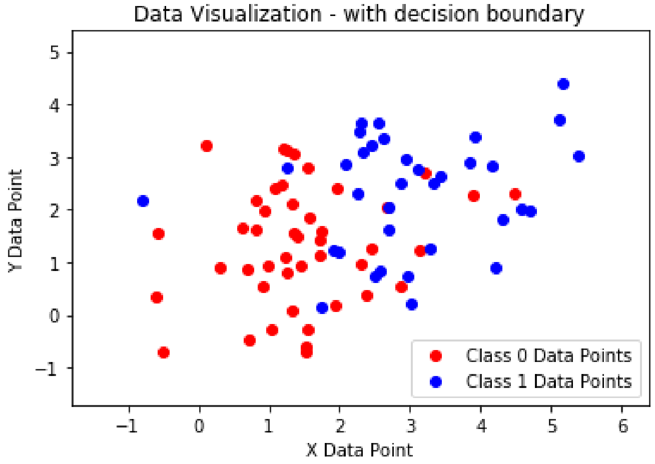
Risk 3

Data Visualization - with decision boundary

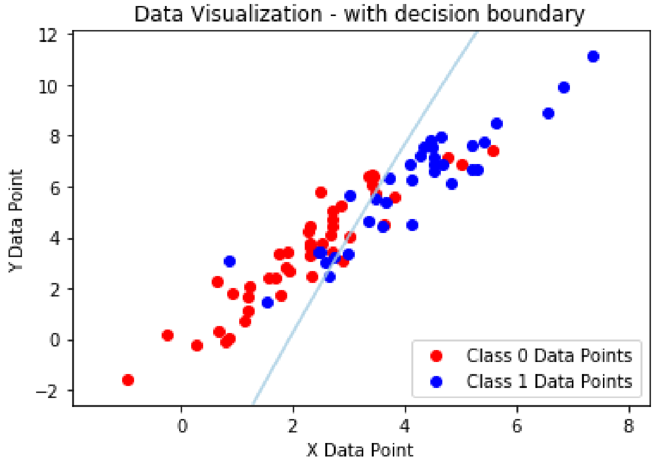
Risk 4



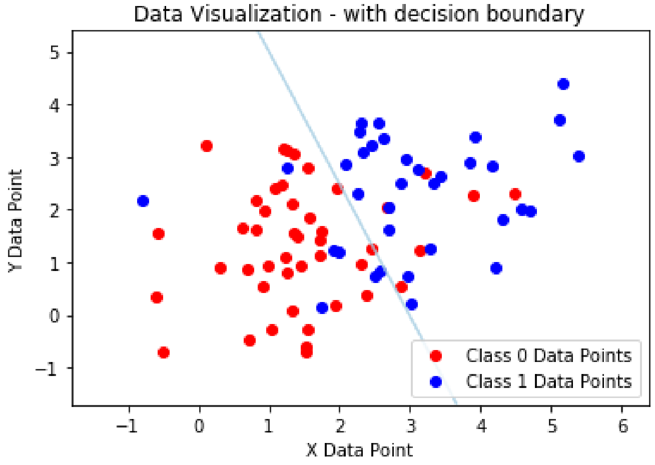
Risk 4



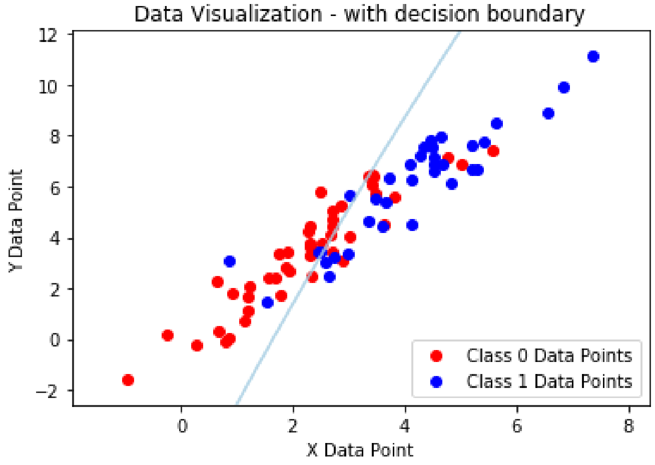
Risk 5



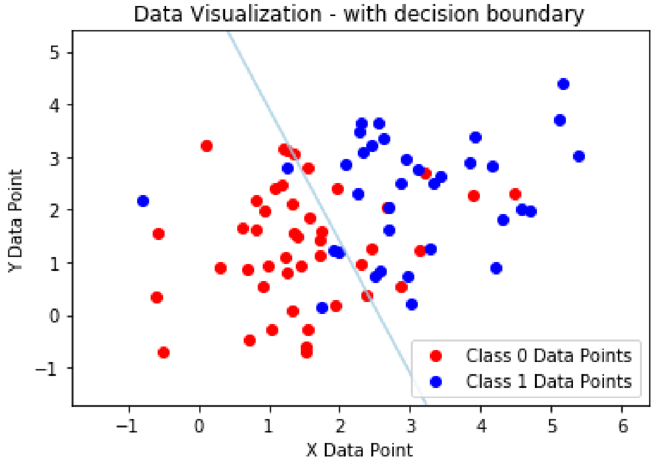
Risk 5



Risk 6



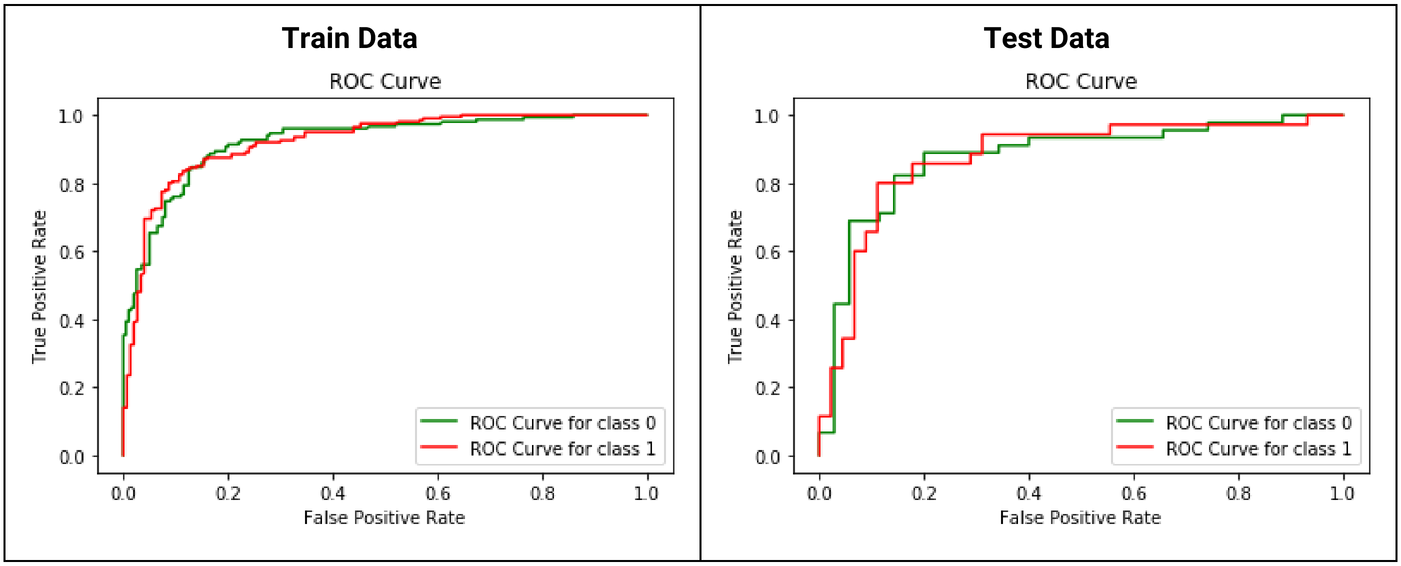
Risk 6



Decision Boundary are shifted, normal axis to the boundary.
After incorporating Risk Matrix, decision boundary for Risk 3 and Risk 4, are not visible on given grid.
On increasing Grid size, It takes lot of time to plot decision boundary with meshgrid.

For Risk 3 and Risk 4, All points are predicted as class 1, that's why boundaries are shifted and not visible in meshgrid size.

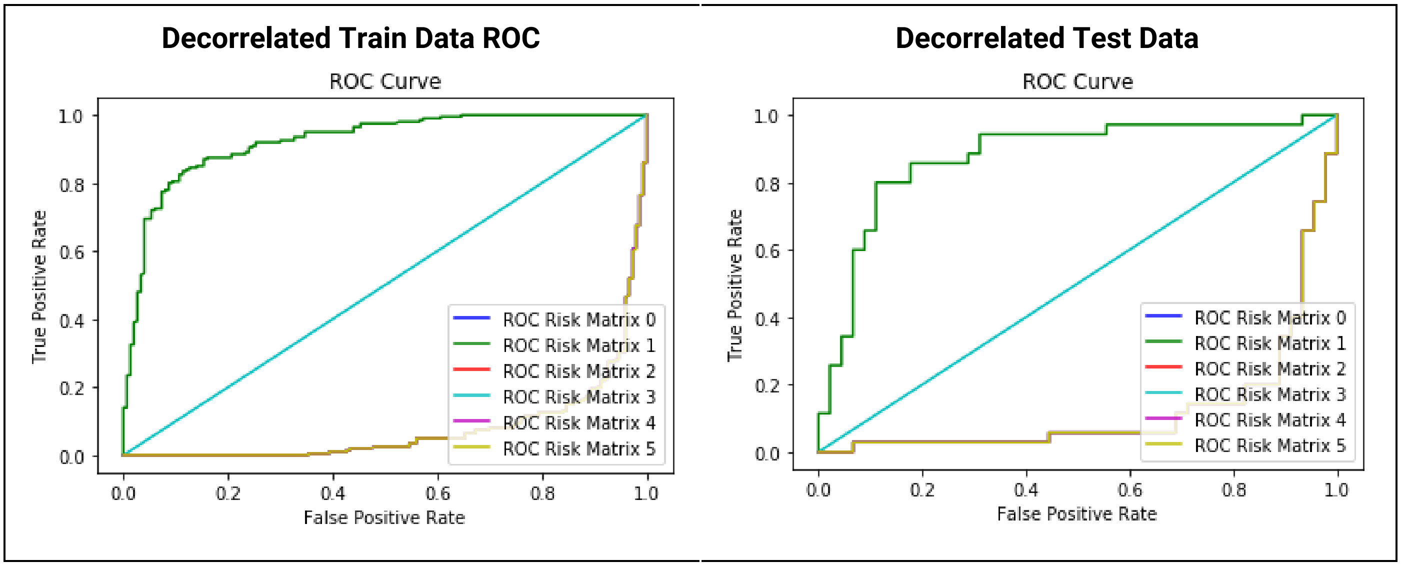
Roc Curve for Decorrelated Data



Train Data

Test Data

ROC For Decorrelated Data with Risk Matrix



Decorrelated Train Data ROC

Decorrelated Test Data