SML Assignment - 2

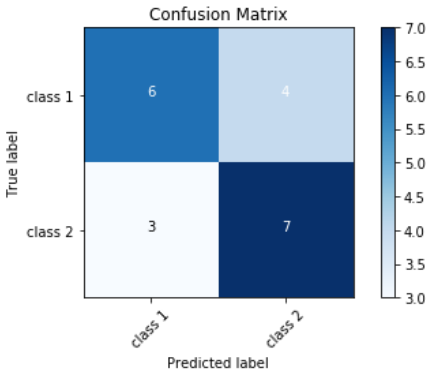
Kaustav Vats (2016048)

Question 1 (Book Questions)

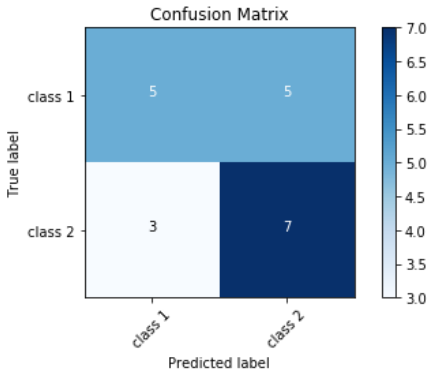
Classes- W1, W2

Confusion Matrix

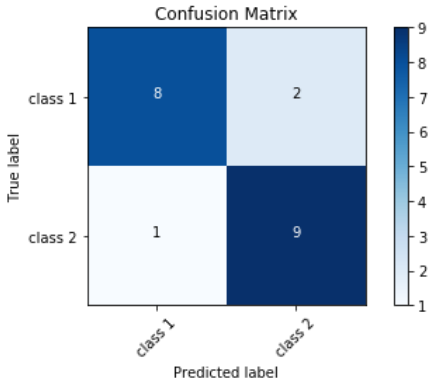
For feature X1 only



For feature X1 and X2



For feature X1, X2 and X3



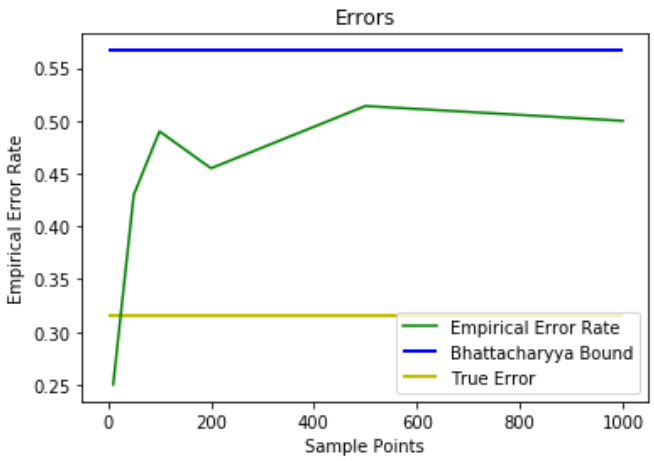
Features\Evaluation Metric	Training Accuracy	Empirical Error	Bhattacharyya Bound
X1	65.0	35.0	0.473996
X1 and X2	60.0	40.0	0.459847
X1, X2 and X3	85.0	15.0	0.411357

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 change the distribution, It might change in such a way that distribution is much better separated than before.

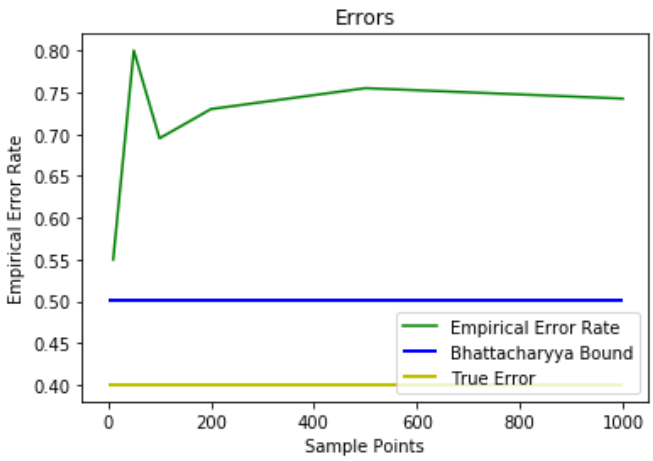
Question 2 & 3 Combined (Book Questions)

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

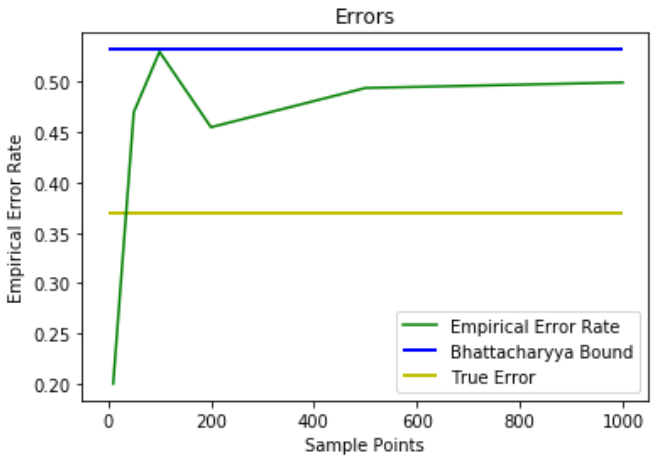
Error Curves



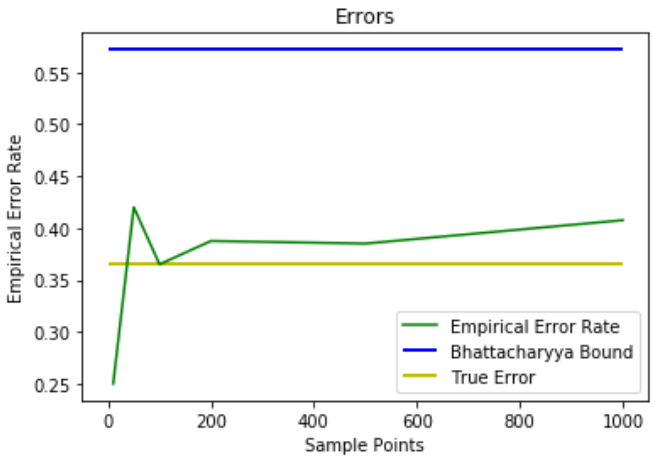
I



II



III

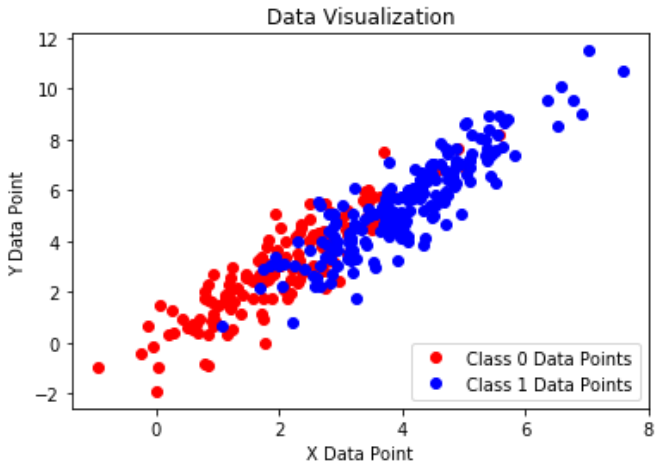


IV

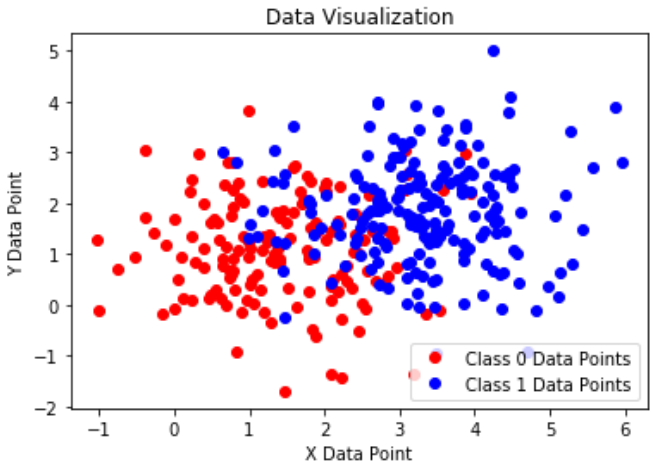
Part B

Data Visualization

Correlated Data

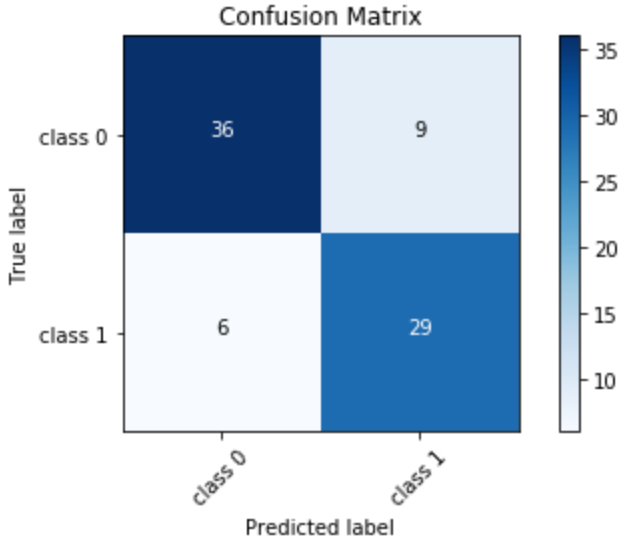


Decorrelated Data

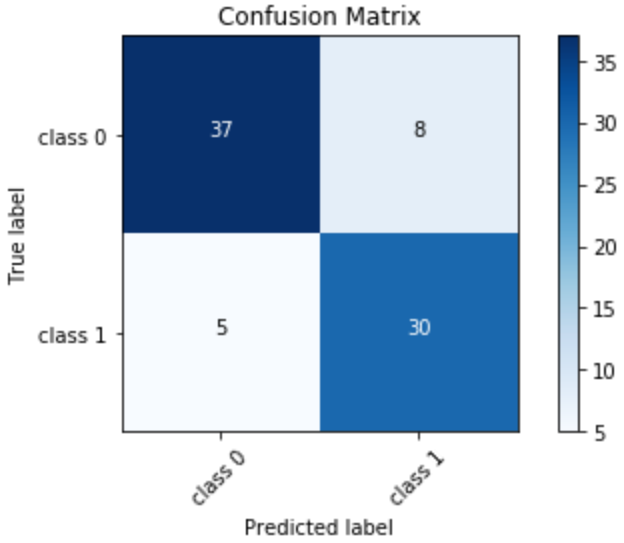


Confusion Matrix

Test Data Confusion Matrix (Correlated)



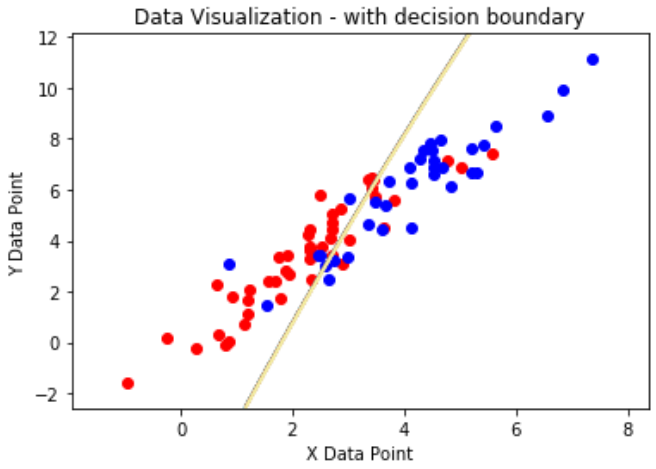
Test Data Confusion Matrix (Decorrelated)



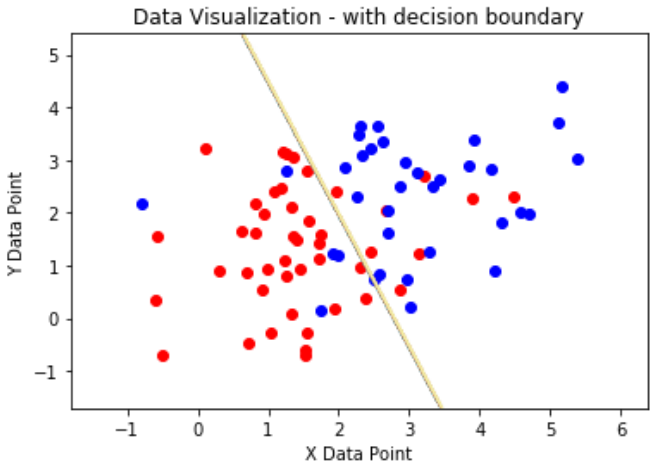
Evaluation Metric	Correlated Test Data	Decorrelated Test Data
Validation Accuracy	83.71	84.57
Testing Accuracy	81.25	83.75

Decision Boundary

Decision Boundary Correlated

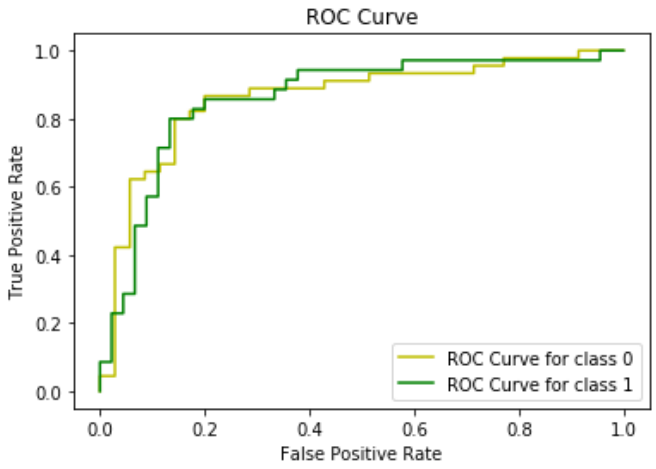


Decision Boundary with Decorrelated

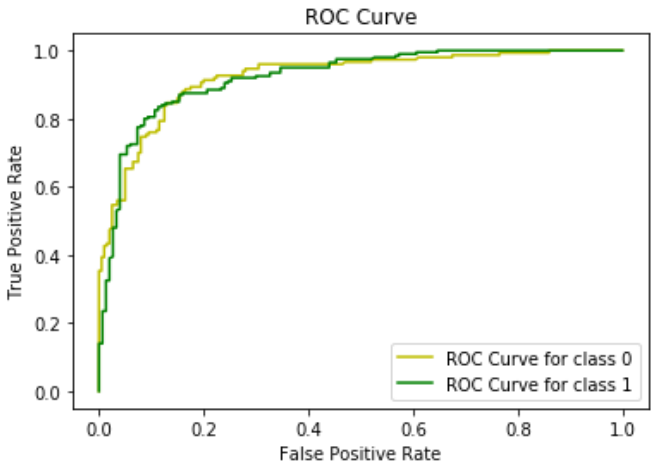


ROC Curves

ROC Curve

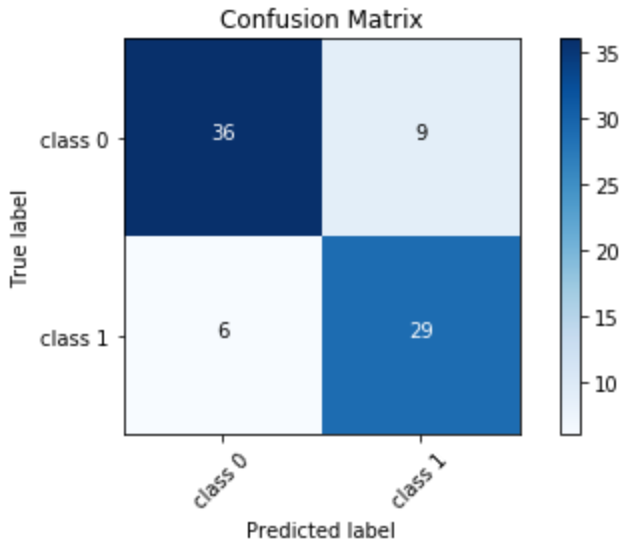


ROC Curve with Decorrelated Test Data

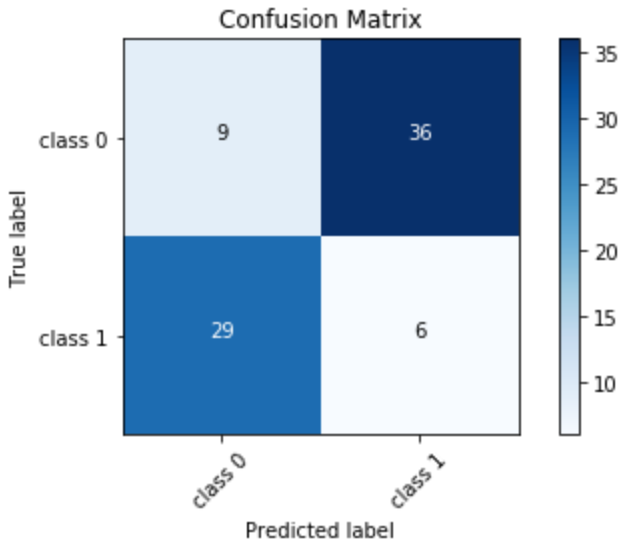


Risk Matrix Confusion Matrix Testing Data (Correlated)

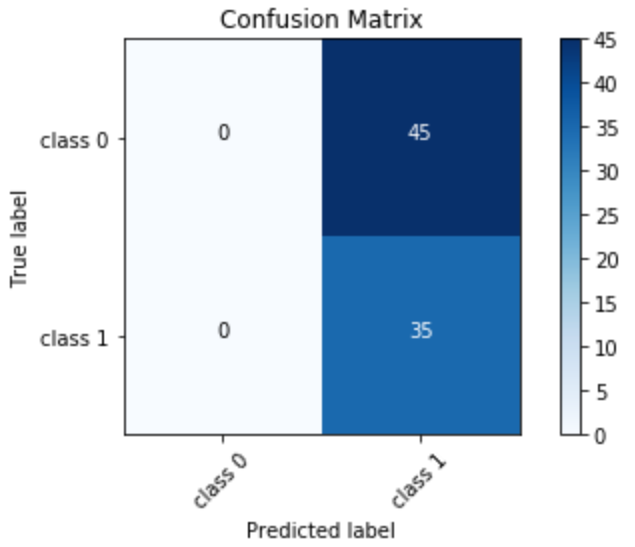
Risk 1



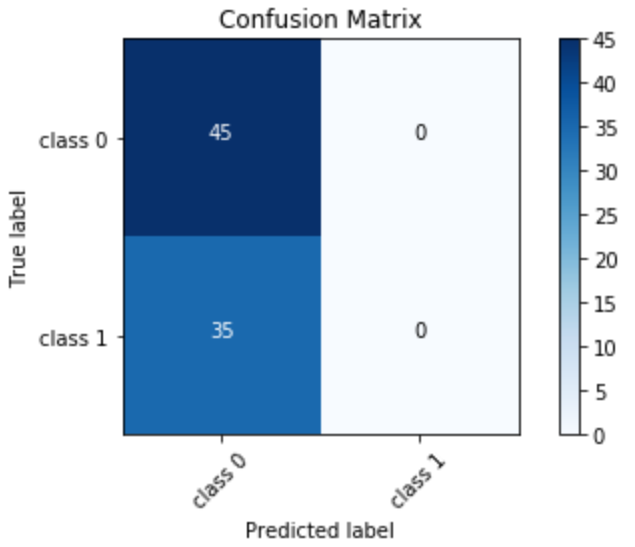
Risk 2



Risk 3

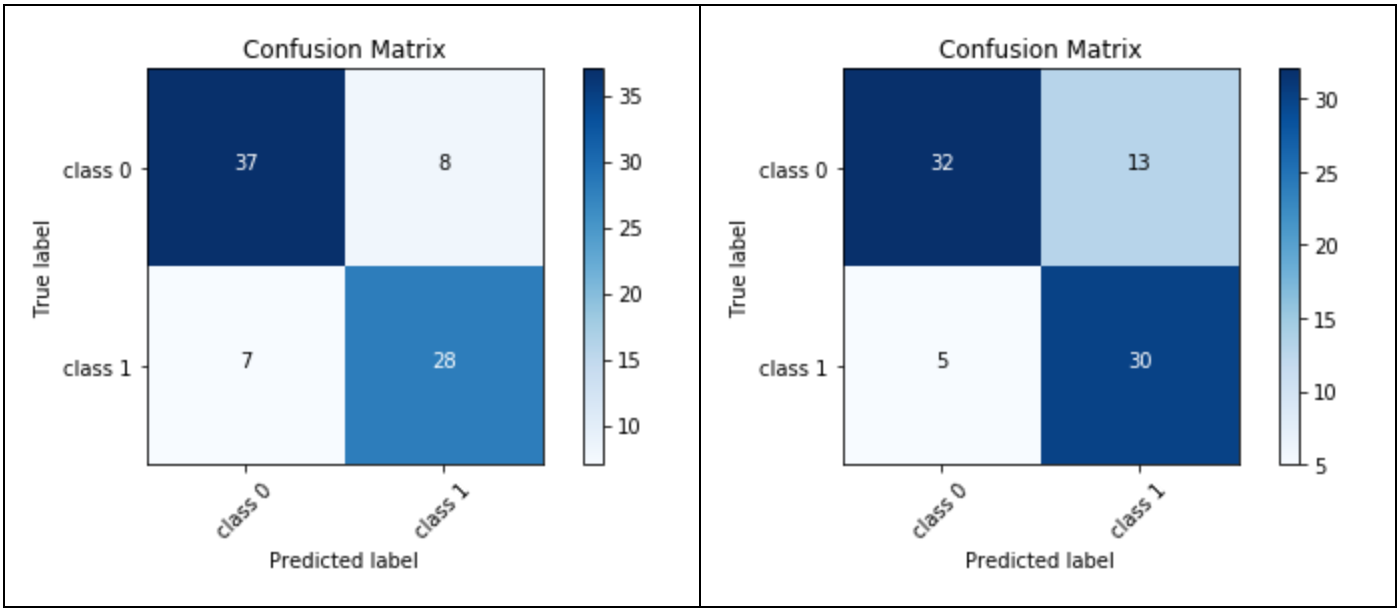


Risk 4

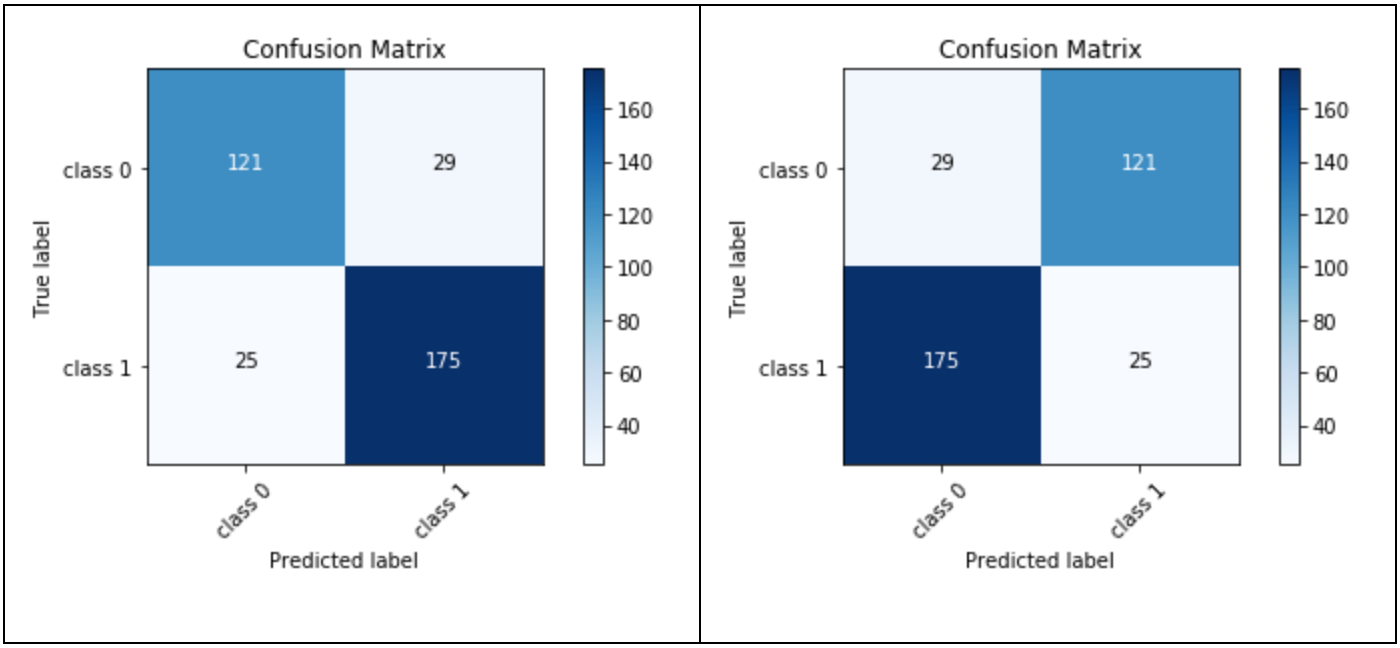


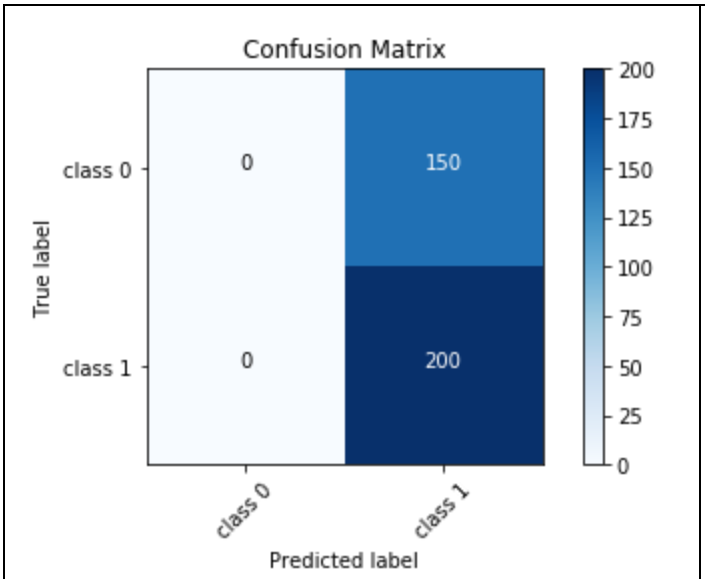
Risk 5

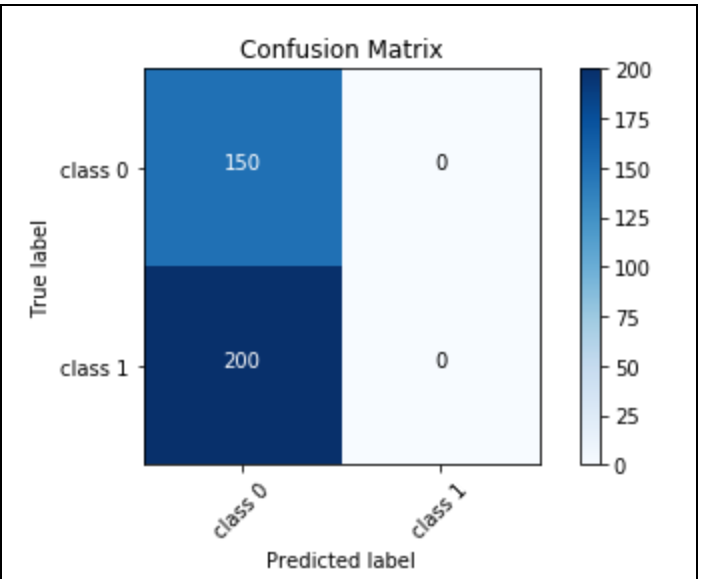
Risk 6

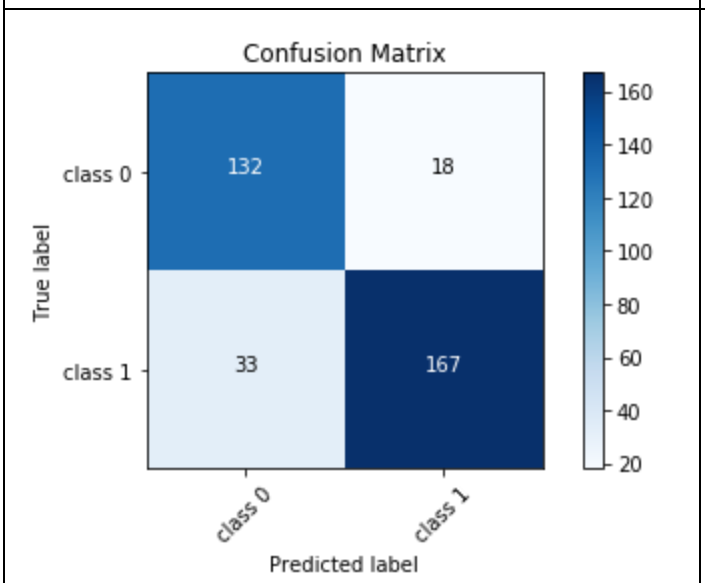


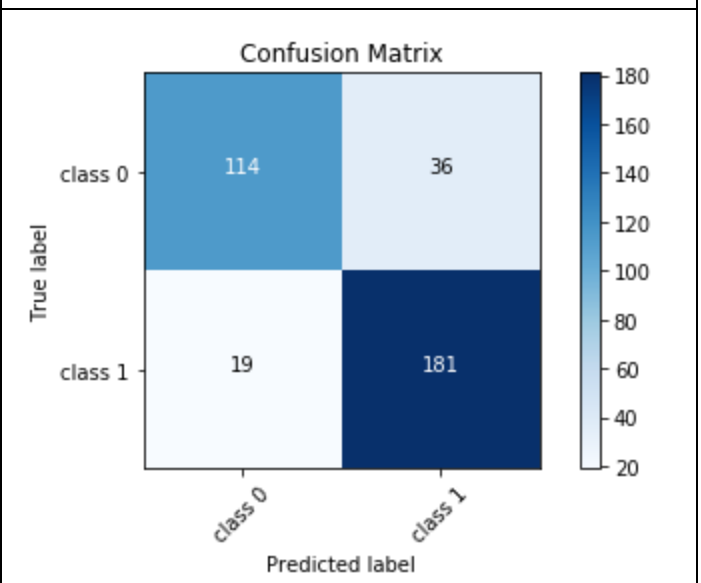
Risk Matrix Confusion Matrix Testing Data (Correlated)





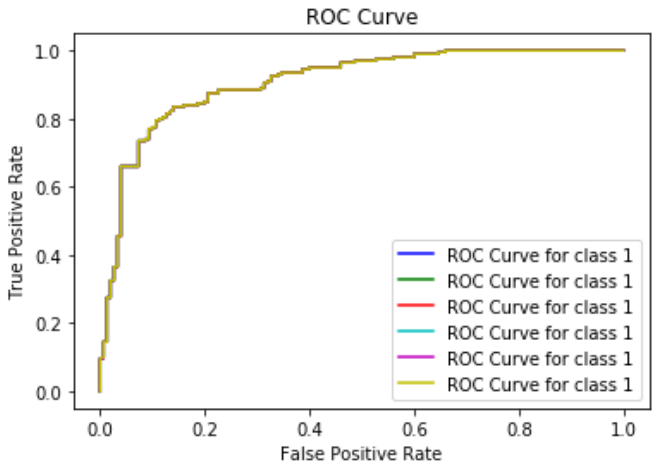
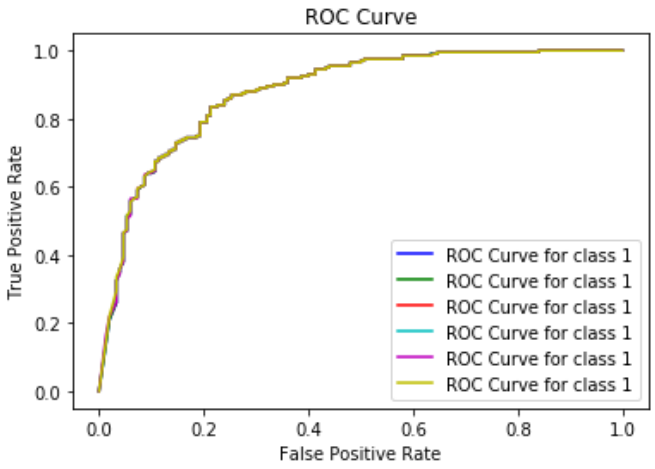






Accuracy	Risk 1	Risk 2	Risk 3	Risk 4	Risk 5	Risk 6
Correlated	60.0	40.0	43.75	56.25	60.0	58.75
Decorrelated	84.57	15.42	57.14	42.85	85.42	84.28

ROC Curves

Roc Correlated	ROC Decorrelated
	

Marginalisation