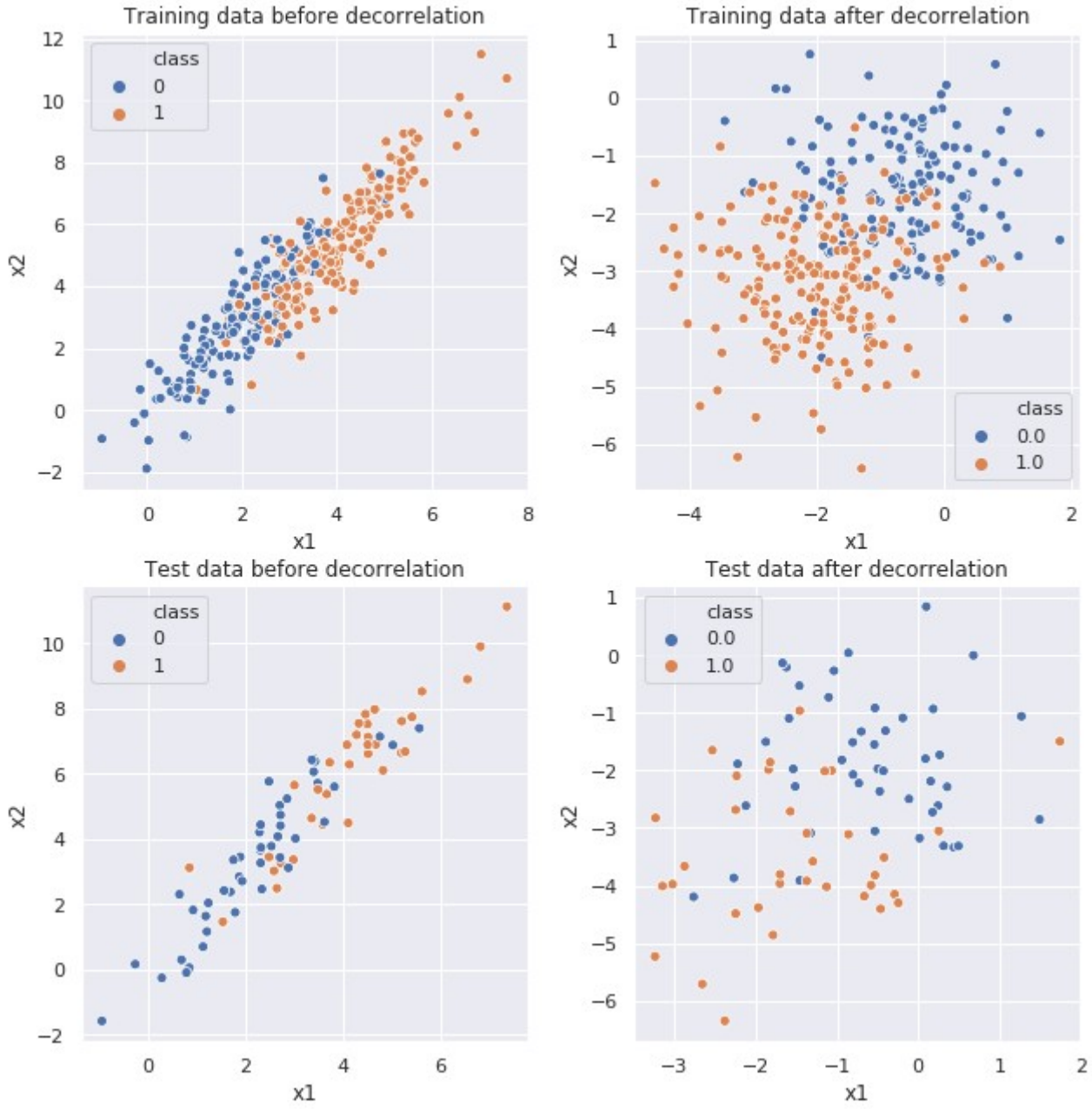
Assignment 2

Statistical Machine Learning

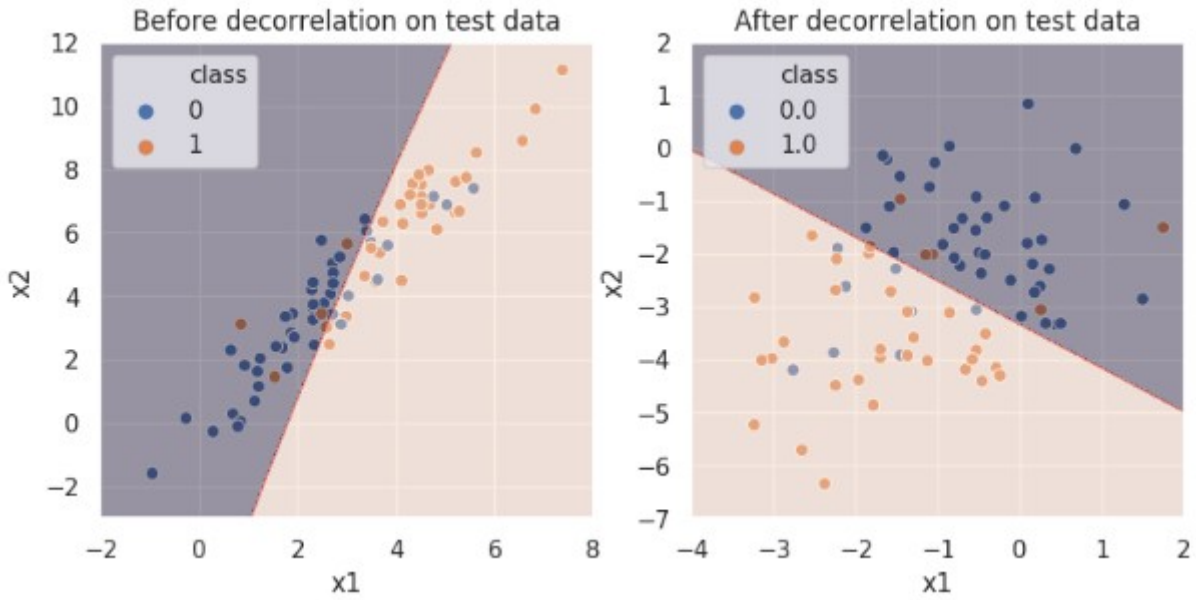
- Mayank Chauhan, MT18008

1. Visualize Dataset

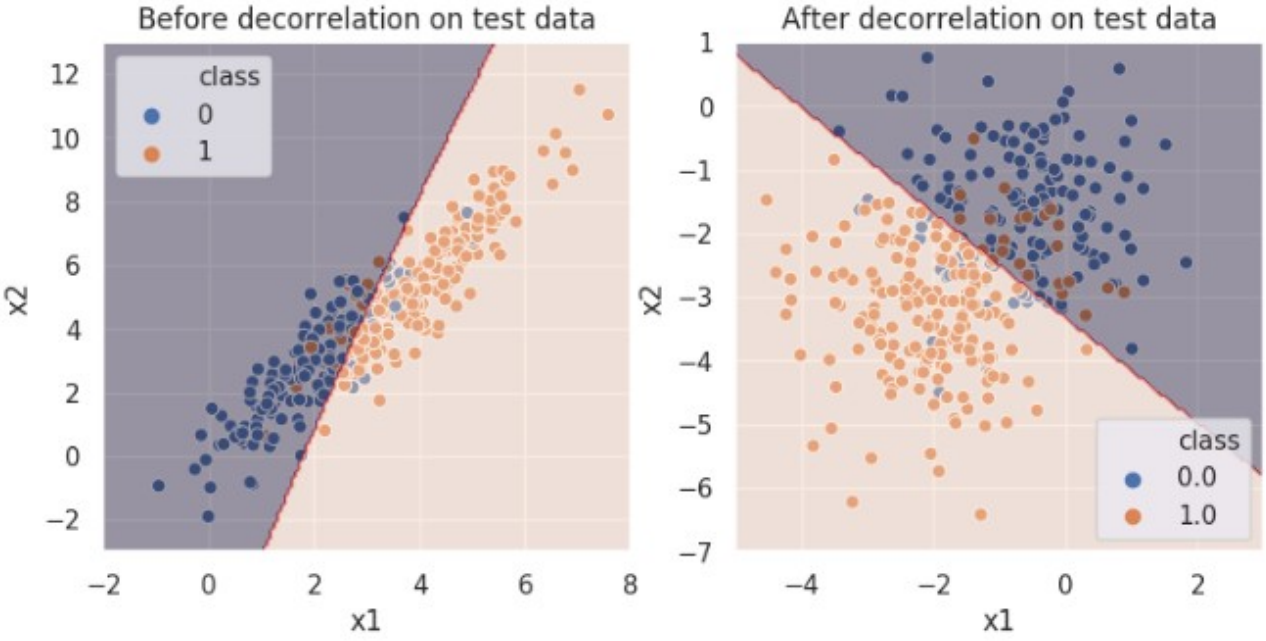


2. Decision boundaries

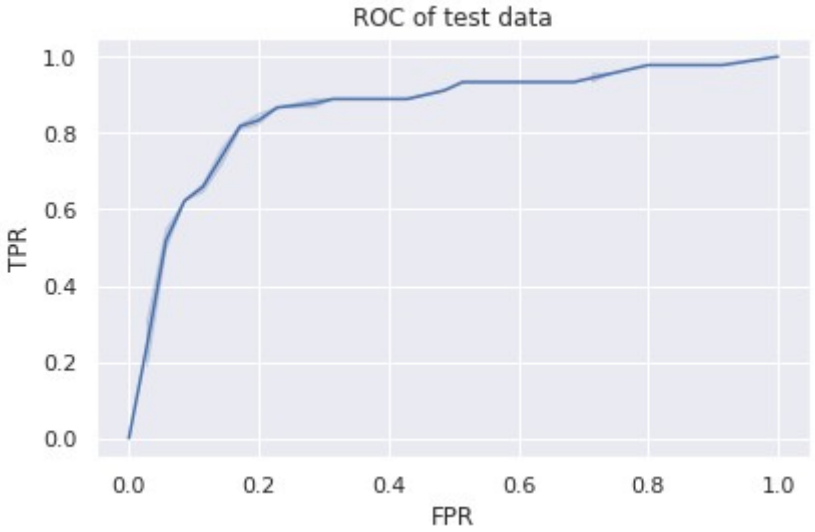
Test Data



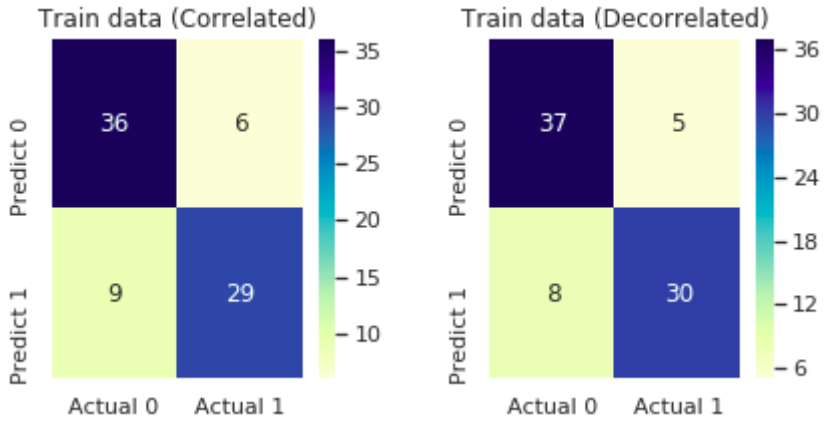
Training Data



ROC



Confusion Matrix

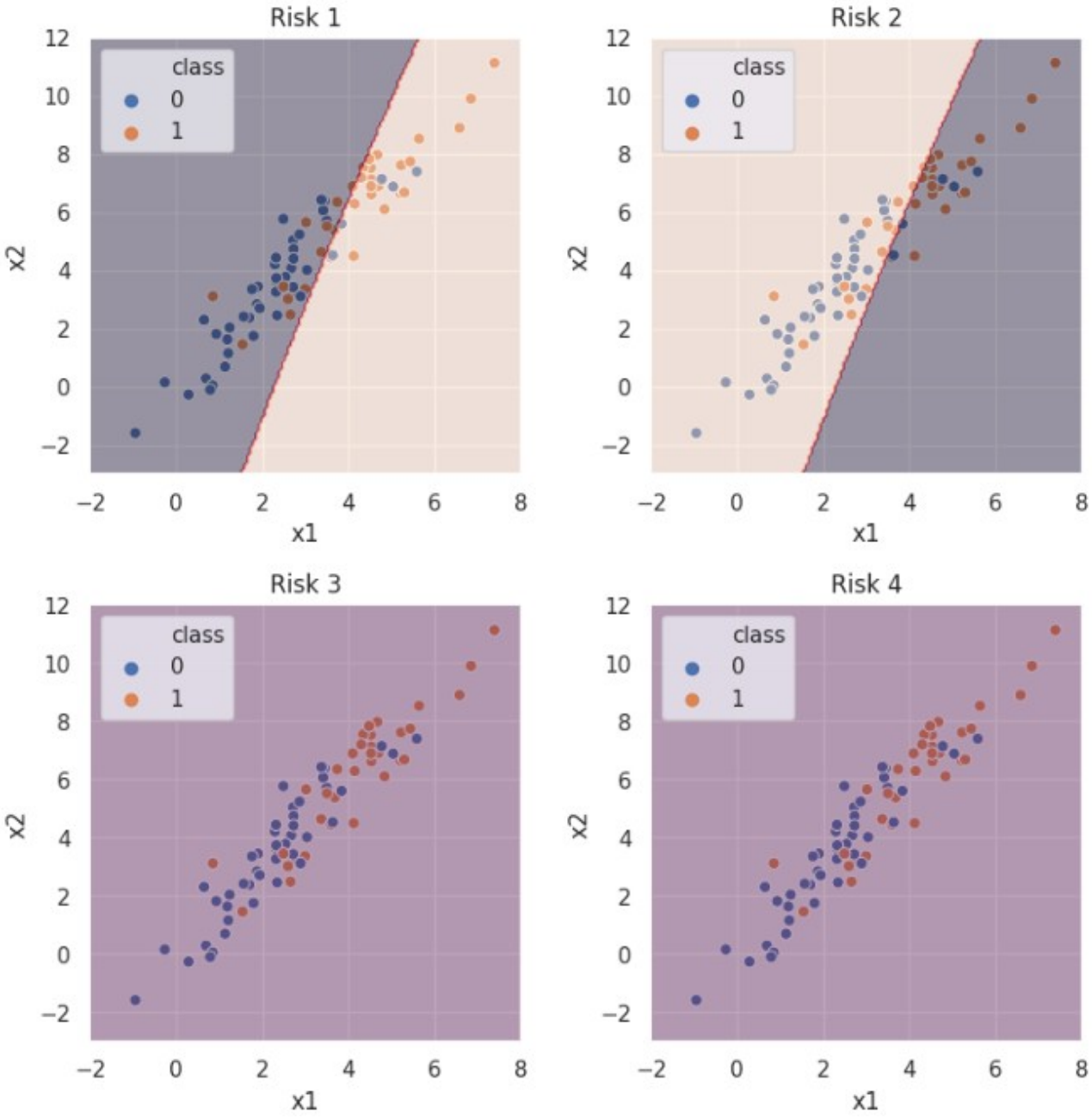


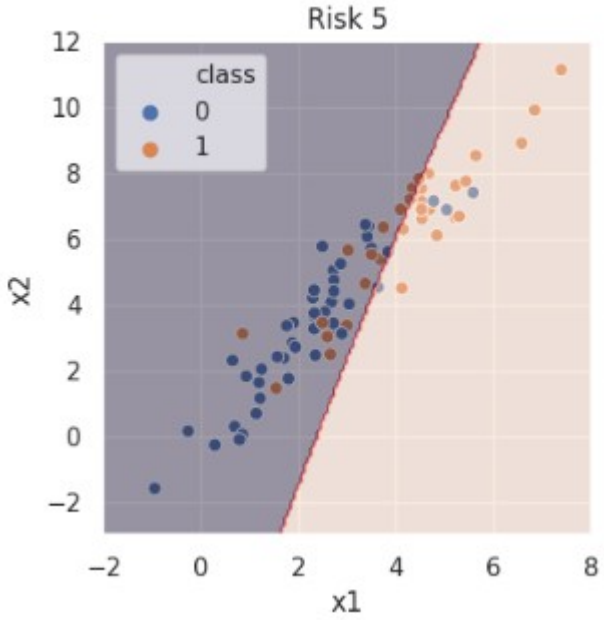
Accuracy

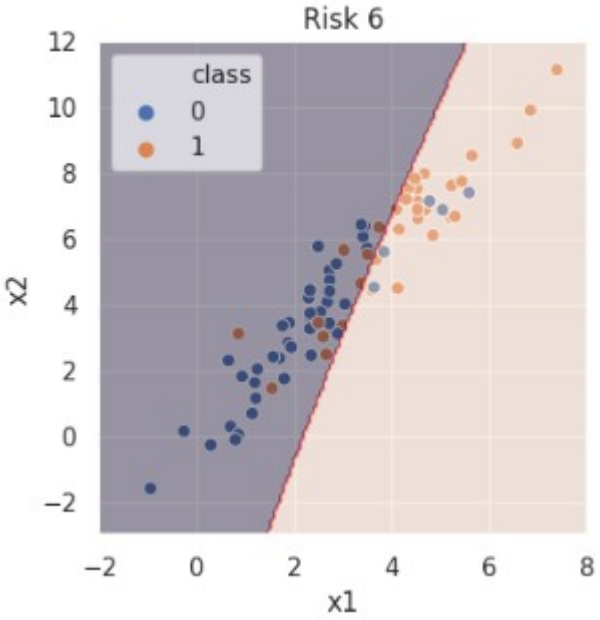
Test data	
Correlated	0.81
Decorrelated	0.83

3. Decision boundaries with Risk

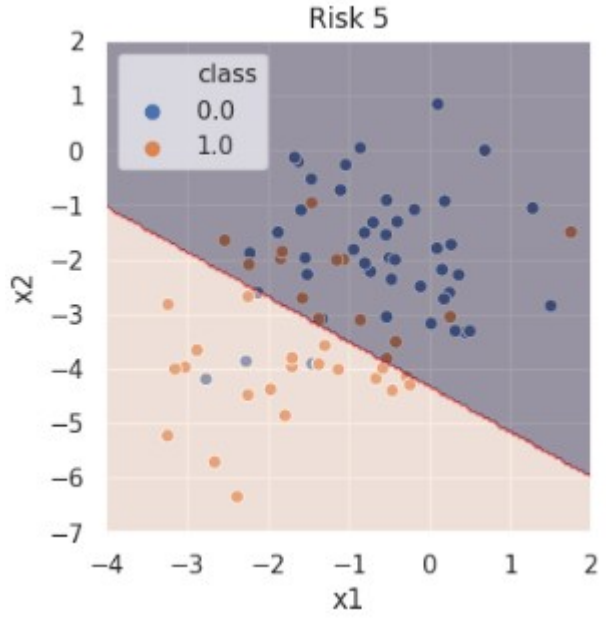
Correlated data

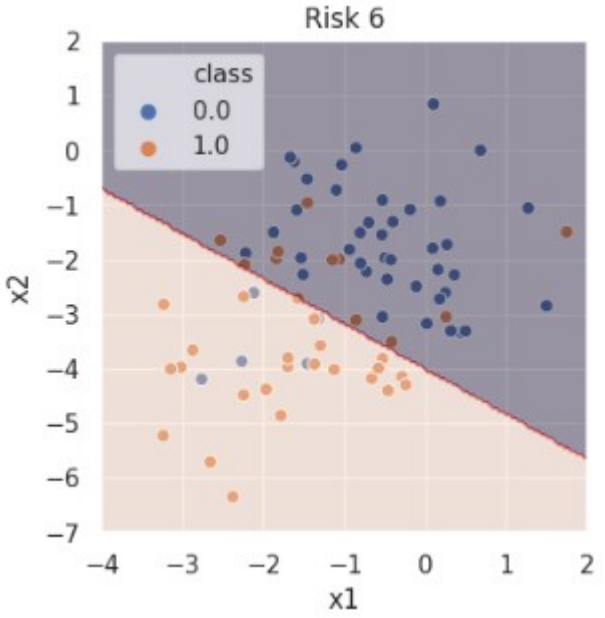


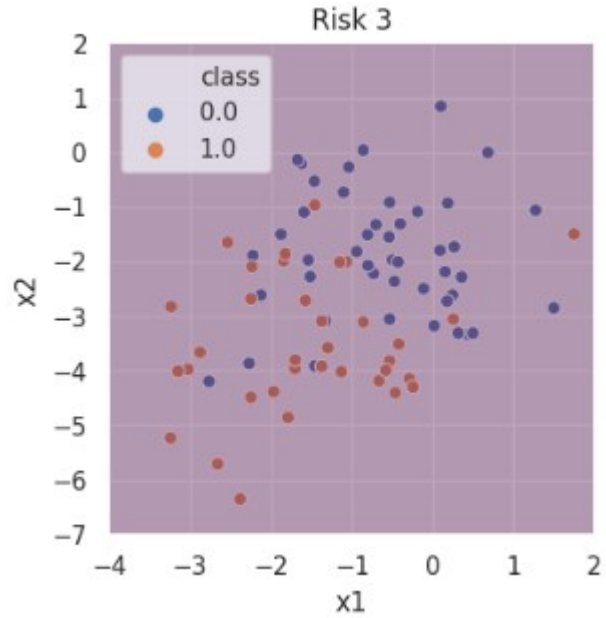

Risk 5
class
0
1
x2
x1
12
10
8
6
4
2
0
−2
−2
0
2
4
6
8

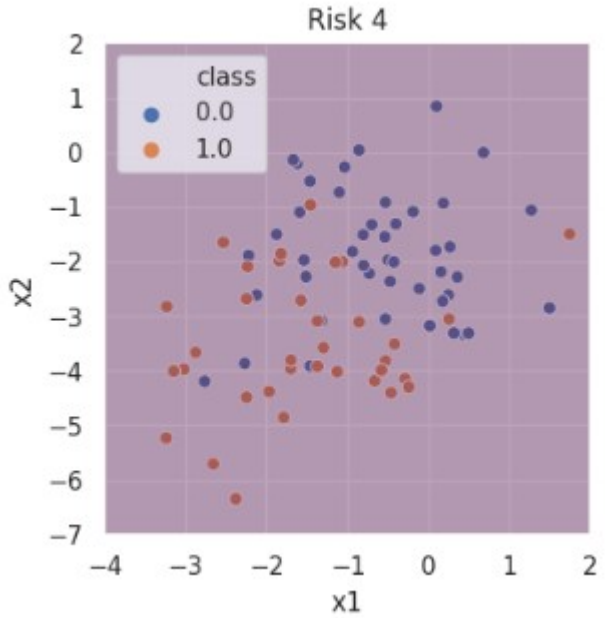

Risk 6
class
0
1
x2
x1
12
10
8
6
4
2
0
−2
−2
0
2
4
6
8

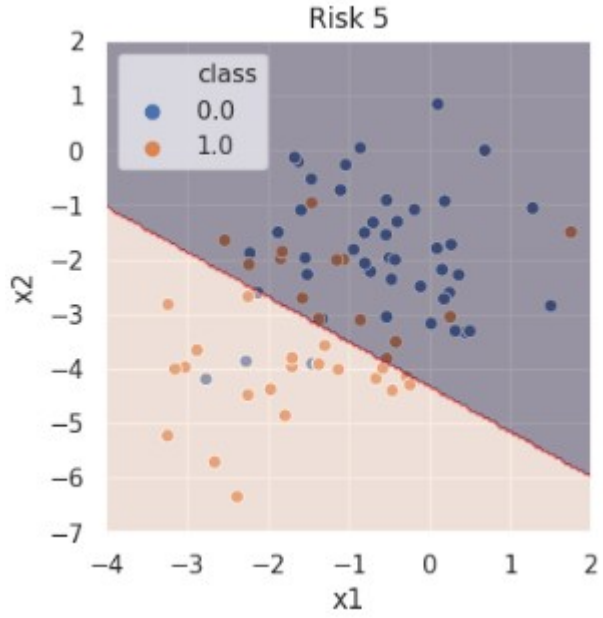
Decorrelated Case

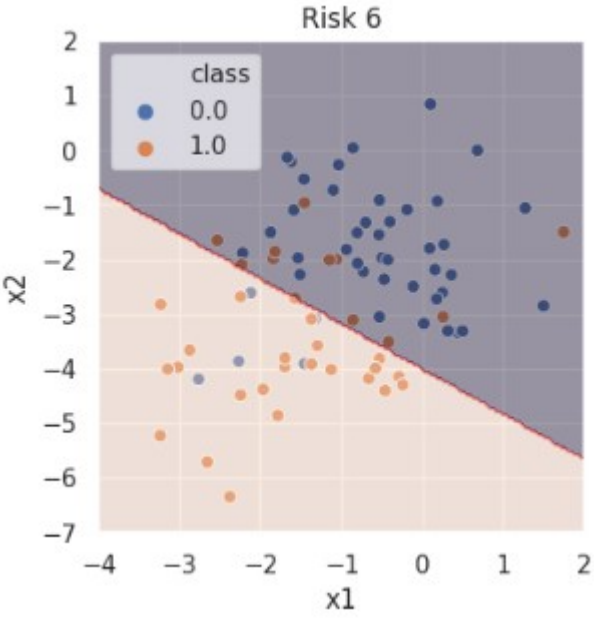






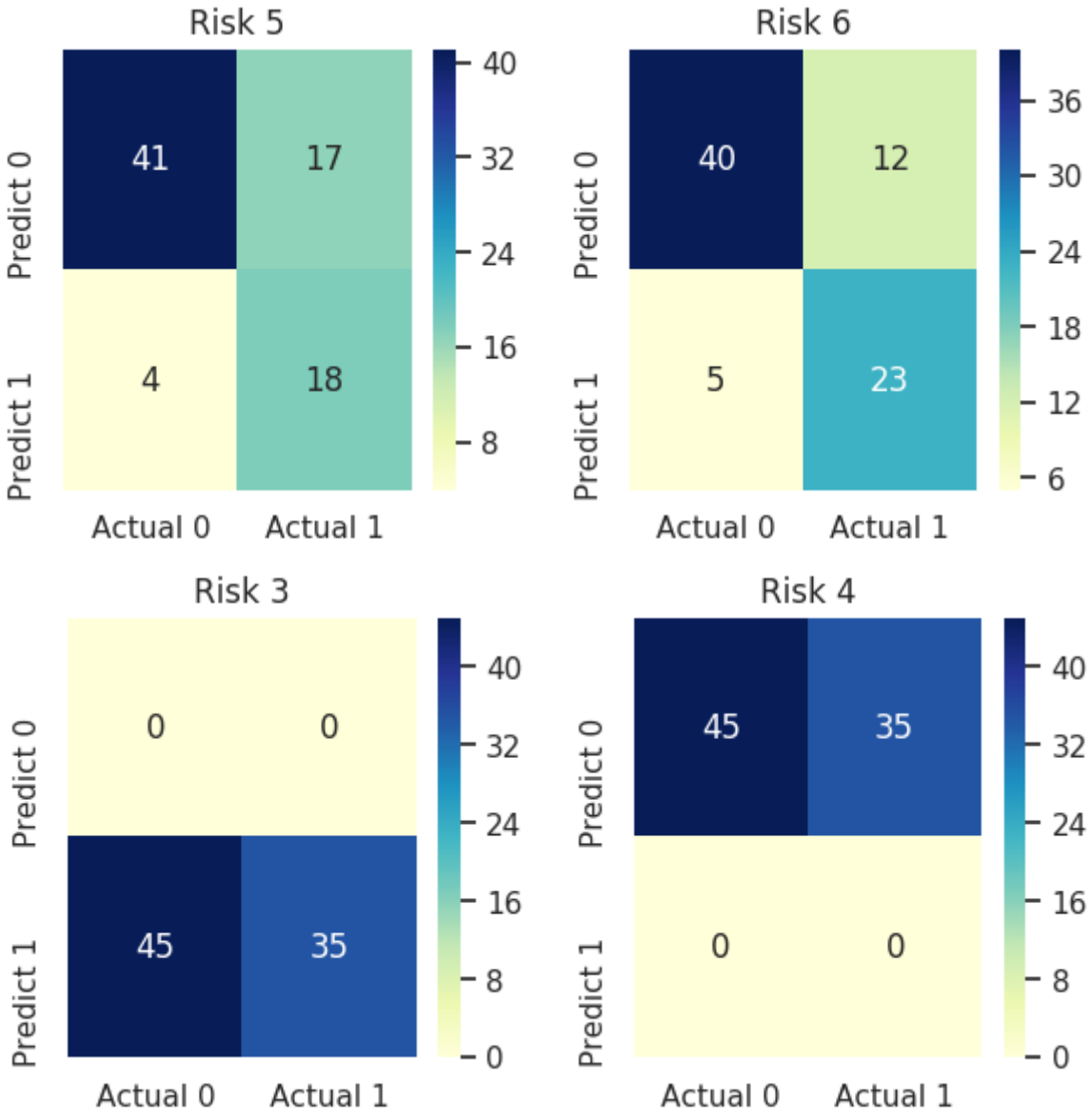


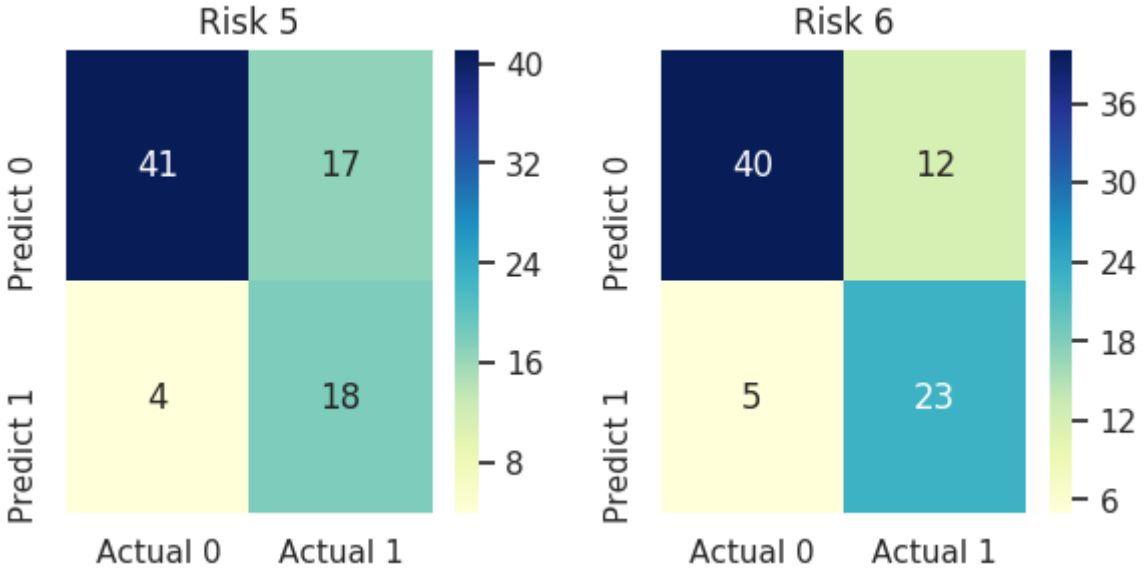

Risk 5
class
0.0
1.0
x2
x1
2
1
0
−1
−2
−3
−4
−5
−6
−7
−4
−3
−2
−1
0
1
2


Risk 6
class
0.0
1.0
x2
x1
2
1
0
−1
−2
−3
−4
−5
−6
−7
−4
−3
−2
−1
0
1
2

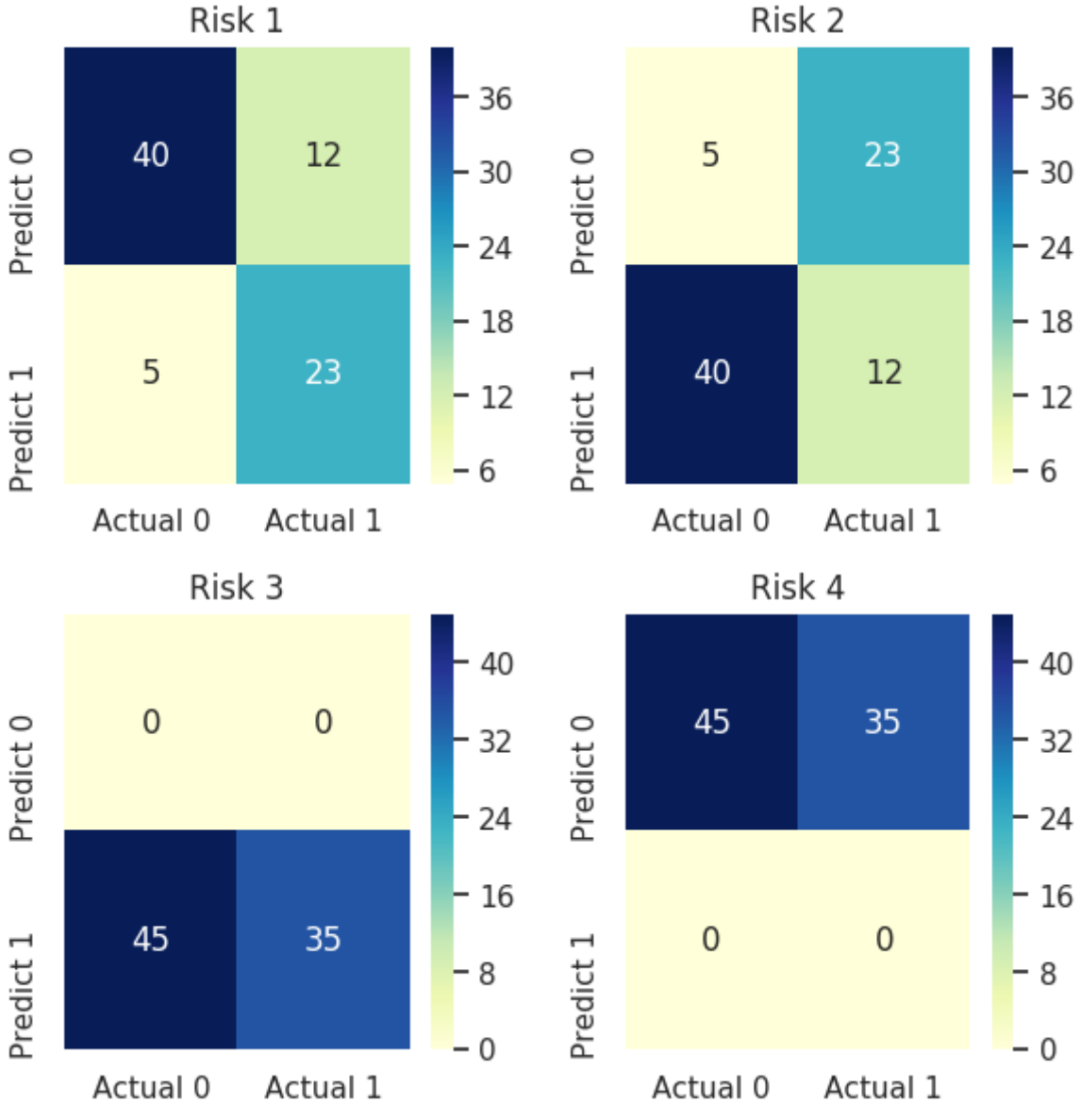
Confusion Matrices

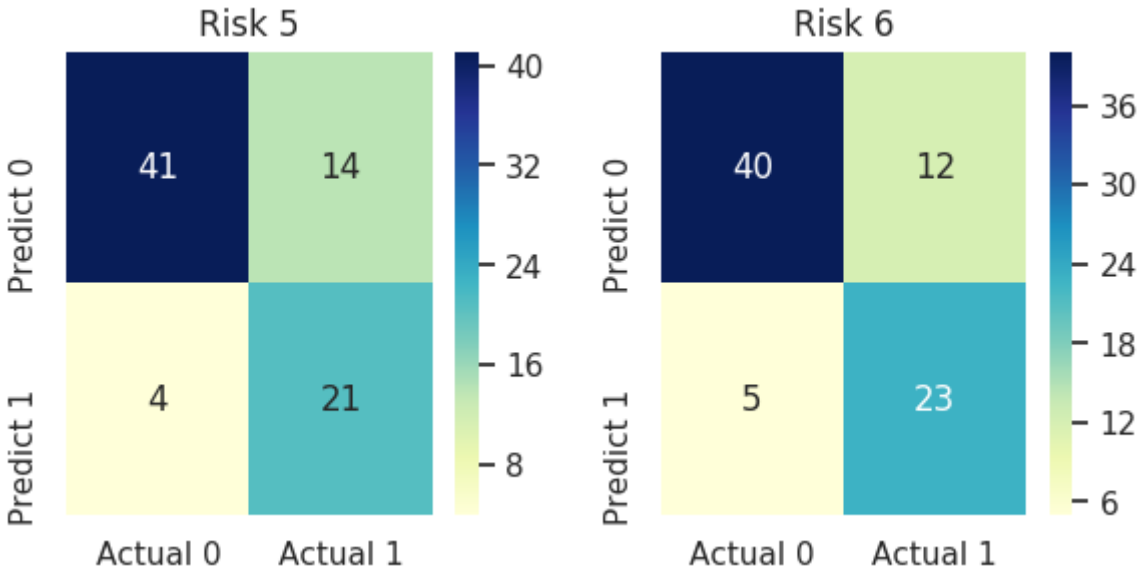
Test data without decorrelation





Test data with decorrelation



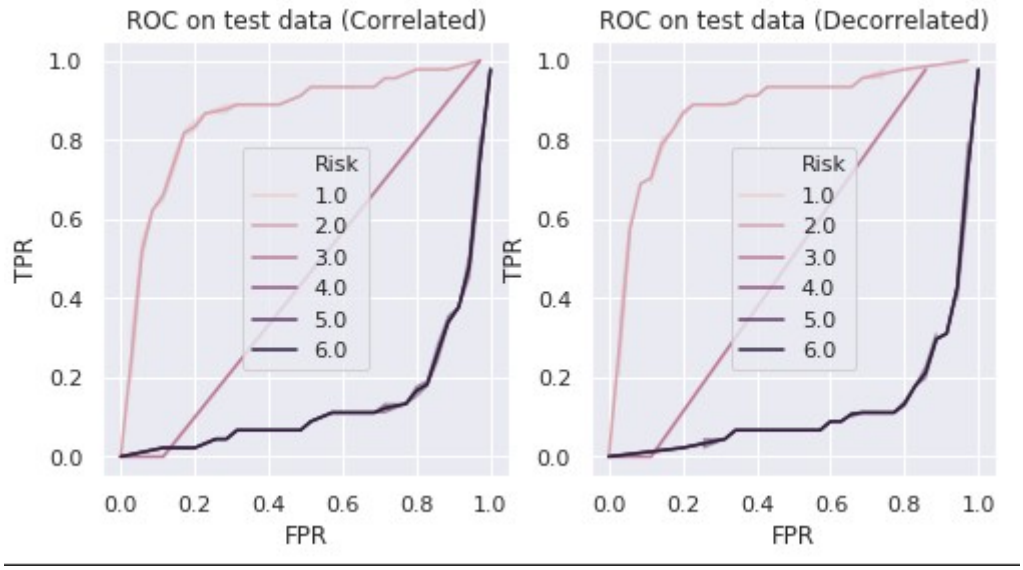


Accuracy

Test Data	Risk 1	Risk 2	Risk 3	Risk 4	Risk 5	Risk 6
Without decorrelation	0.76	0.23	0.43	0.56	0.73	0.78
With decorrelation	0.78	0.21	0.43	0.56	0.77	0.78

ROC Curve

Test Data with Risk Included



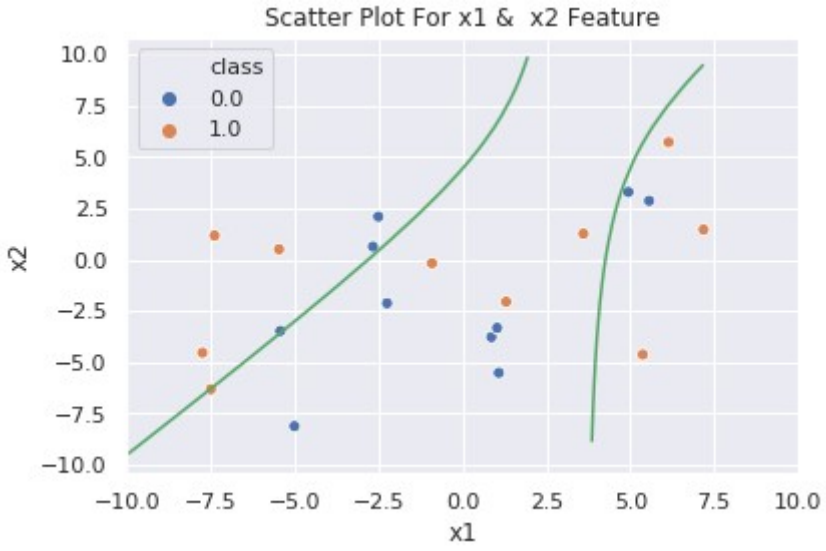
Computer Excerises

Q.2

Empirical Error Rate

Features	Error rate	Bhattacharya bound
X1	0.35	0.473
X1, X2	0.45	0.460
X1, X2, X3	0.15	0.411

Q.7

	True Error rate	Error rate	Bhattacharya bound
Q7.	0.308	0.32	0.44
Q8.a	0.36	0.39	0.43
Q8.b	0.36	0.34	0.45

Decision boundary Q2. With two features x1 and x2