PS5841

Data Science in Finance & Insurance

Classification Measures

Yubo Wang

Spring 2022

Classification Quality

- Error Rate
- Accuracy
- Confusion Matrix
- ROC (Receiver Operating Characteristics)
- …

Confusion Matrix

		Predicted class		
		− or Null	+ or Non-null	Total
True	− or Null	True Neg. (TN)	False Pos. (FP)	N
class	+ or Non-null	False Neg. (FN)	True Pos. (TP)	P
	Total	N*	P*	

Name	Definition	Synonyms
False Pos. rate	FP/N	Type I error, 1−Specificity
True Pos. rate	TP/P	1−Type II error, power, sensitivity, recall
Pos. Pred. value	TP/P*	Precision, 1−false discovery proportion
Neg. Pred. value	TN/N*	

FN/P   Type II error

Error Rate = (FP + FN) / (N + P)   off-diagonal
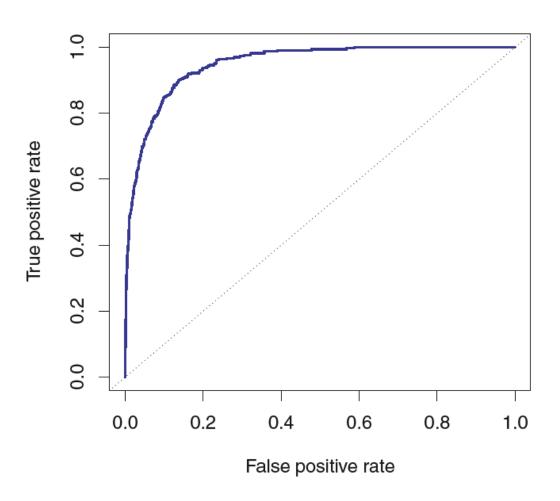Accuracy = (TN + TP) / (N + P)   diagonal
Sensitivity = TP / P   along row P, % of P correctly classified
Specificity = TN / N   along row N, % of N correctly classified
Precision = TP/ P*   along col P*, % of P* correctly classified

ROC

ROC Example

					TP/N+	FP/N-
Positive Class	2	2	<- N+			
Negative Class	1	2	<- N-			
True Class	1	1	2	2		
Pred Prob of Class 2	0.1	0.4	0.35	0.8		
					TP/N+	FP/N-
Threshold >= x	Predicted Class				TPR	FPR
sklearn specific 1.8	1	1	1	1	0	0
0.8	1	1	1	2	0.5	0
0.4	1	2	1	2	0.5	0.5
0.35	1	2	2	2	1	0.5
0.1	2	2	2	2	1	1

That was