

Introduction to ROC Curve

Lu Wang

February 9, 2016

Data and A Binary Classifier

Subject	Disease	Predicted probability
1	0	0.513
2	1	0.921
3	1	0.203
4	1	0.840
5	0	0.143
⋮	⋮	⋮
459	1	0.397

⇒

	Disease present	Disease absent
Test positive	207	42
Test negative	88	122

- ▶ True positive rate (TPR) = $TP / (TP + FN)$
- ▶ False positive rate (FPR) = $FP / (FP + TN)$

Data and A Binary Classifier

Subject	Disease	Predicted probability
1	0	0.513
2	1	0.921
3	1	0.203
4	1	0.840
5	0	0.143
⋮	⋮	⋮
459	1	0.397

⇒

	Disease present	Disease absent
Test positive	207	42
Test negative	88	122

- ▶ True positive rate (TPR) = $TP / (TP + FN)$
- ▶ False positive rate (FPR) = $FP / (FP + TN)$

Data and A Binary Classifier

Subject	Disease	Predicted probability
1	0	0.513
2	1	0.921
3	1	0.203
4	1	0.840
5	0	0.143
⋮	⋮	⋮
459	1	0.397

True positives (TP)

⇒

	Disease present	Disease absent
Test positive	207	42
Test negative	88	122

- ▶ True positive rate (TPR) = $TP / (TP + FN)$
- ▶ False positive rate (FPR) = $FP / (FP + TN)$

Data and A Binary Classifier

Subject	Disease	Predicted probability
1	0	0.513
2	1	0.921
3	1	0.203
4	1	0.840
5	0	0.143
⋮	⋮	⋮
459	1	0.397

⇒

	Disease present	Disease absent
Test positive	207	42
Test negative	88	122

True positives (TP) is indicated by a blue arrow pointing to the cell (Test positive, Disease present) containing 207.

False positives (FP) is indicated by a blue arrow pointing to the cell (Test positive, Disease absent) containing 42.

- ▶ True positive rate (TPR) = $TP / (TP + FN)$
- ▶ False positive rate (FPR) = $FP / (FP + TN)$

Data and A Binary Classifier

Subject	Disease	Predicted probability
1	0	0.513
2	1	0.921
3	1	0.203
4	1	0.840
5	0	0.143
⋮	⋮	⋮
459	1	0.397

⇒

	Disease present	Disease absent
Test positive	207	42
Test negative	88	122

True positives (TP) points to 207

False positives (FP) points to 42

True negatives (TN) points to 122

- ▶ True positive rate (TPR) = $TP / (TP + FN)$
- ▶ False positive rate (FPR) = $FP / (FP + TN)$

Data and A Binary Classifier

Subject	Disease	Predicted probability
1	0	0.513
2	1	0.921
3	1	0.203
4	1	0.840
5	0	0.143
⋮	⋮	⋮
459	1	0.397

⇒

	Disease present	Disease absent
Test positive	207	42
Test negative	88	122

True positives (TP) points to 207
False positives (FP) points to 42
False negatives (FN) points to 88
True negatives (TN) points to 122

► True positive rate (TPR) = $TP / (TP + FN)$

► False positive rate (FPR) = $FP / (FP + TN)$

Data and A Binary Classifier

Subject	Disease	Predicted probability
1	0	0.513
2	1	0.921
3	1	0.203
4	1	0.840
5	0	0.143
⋮	⋮	⋮
459	1	0.397

⇒

	Disease present	Disease absent
Test positive	207	42
Test negative	88	122

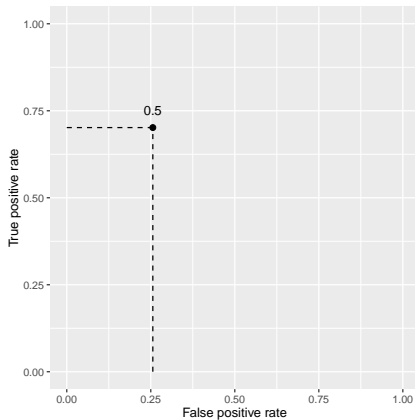
True positives (TP) points to 207
False positives (FP) points to 42
False negatives (FN) points to 88
True negatives (TN) points to 122

- ▶ True positive rate (TPR) = $TP / (TP + FN)$
- ▶ False positive rate (FPR) = $FP / (FP + TN)$

True positive rate and False positive rate

Cutpoint	TPR	FPR
0.3	0.831	0.433
0.5	0.702	0.256
0.8	0.325	0.067

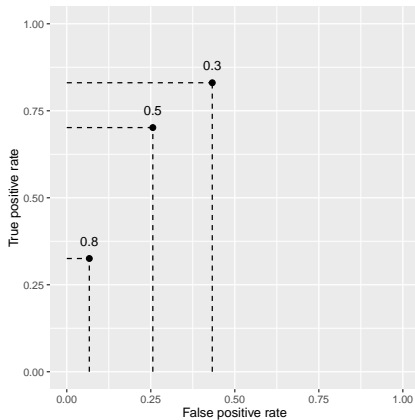
⇒



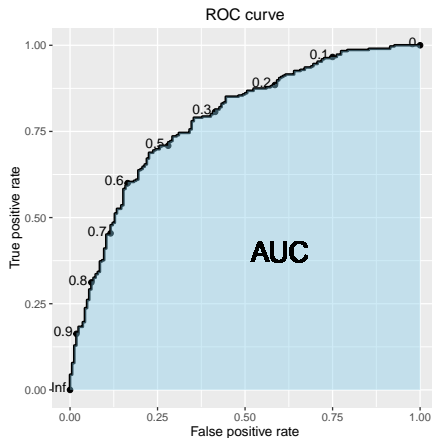
True positive rate and False positive rate

Cutpoint	TPR	FPR
0.3	0.831	0.433
0.5	0.702	0.256
0.8	0.325	0.067

⇒

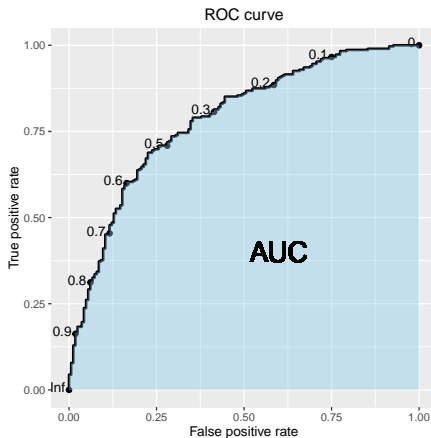


Receiver Operating Characteristic (ROC) Curve



The area under the curve (AUC) is the probability that the model correctly classifies the two subjects in a randomly drawn pair.

Receiver Operating Characteristic (ROC) Curve



The area under the curve (AUC) is the probability that the model correctly classifies the two subjects in a randomly drawn pair.

Receiver Operating Characteristic (ROC) Curve

