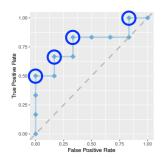
Introduction to Machine Learning

Evaluation Measures for Binary Classification: ROC Visualization



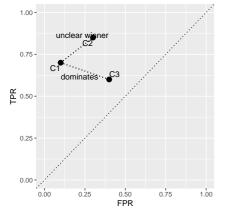
Learning goals

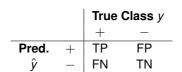
- Understand ROC curve
- Be able to compute a ROC curve manually
- Understand that ROC curve is invariant to class priors at test-time
- Discuss threshold selection
- Understand AUC



LABELS: ROC SPACE

- For comparing classifiers, we characterize them by their TPR and FPR values and plot them in a coordinate system.
- We could also use two different ROC metrics which define a trade-off, for instance, TPR and PPV.





$$\mathsf{TPR} = \frac{\mathsf{TP}}{\mathsf{TP} + \mathsf{FN}}$$

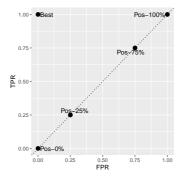
$$\mathsf{FPR} = rac{\mathsf{FP}}{\mathsf{FP} + \mathsf{TN}}$$



LABELS: ROC SPACE

- The best classifier lies on the top-left corner, where FPR equals 0 and TPR is maximal.
- The diagonal is worst as it corresponds to a classifier producing random labels (with different proportions).

- If each positive x will be randomly classified with 25% as "pos", TPR = 0.25.
- If we assign each negative x randomly to "pos",
 FPR = 0.25.

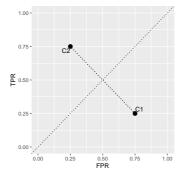




LABELS: ROC SPACE

- In practice, we should never obtain a classifier below the diagonal.
- Inverting the predicted labels (0 \mapsto 1 and 1 \mapsto 0) will result in a reflection at the diagonal.

 \Rightarrow TPR_{new} = 1 - TPR and FPR_{new} = 1 - FPR.





LABEL DISTRIBUTION IN TPR AND FPR

TPR and FPR (ROC curves) are insensitive to the class distribution in the sense that they are not affected by changes in the ratio n_+/n_- (at prediction).



Example 1:

Proportion $n_+/n_-=1$

Example	e 2:
---------	------

Proportion $n_+/n_-=2$

	Actual Positive	Actual Negative
Pred. Positive	40	25
Pred. Negative	10	25

	Actual Positive	Actual Negative
Pred. Positive	80	25
Pred. Negative	20	25

$$MCE = 35/100 = 0.35$$

$$TPR = 0.8$$

 $FPR = 0.5$

$$MCE = 45/150 = 0.3$$

 $TPR = 0.8$

$$FPR = 0.5$$

Note: If class proportions differ during training, the above is not true. Estimated posterior probabilities can change!

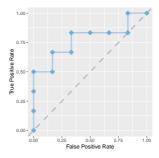
FROM PROBABILITIES TO LABELS: ROC CURVE

Remember: Both probabilistic and scoring classifiers can output classes by thresholding:

$$h(\mathbf{x}) = [\pi(\mathbf{x}) \geq c]$$
 or $h(\mathbf{x}) = [f(\mathbf{x}) \geq c_f]$.

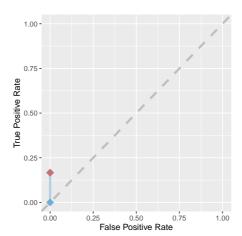


- Rank test observations on decreasing score.
- Start with c = 1, so we start in (0,0); we predict everything as negative.
- 3 Iterate through all possible thresholds *c* and proceed for each observation *x* as follows:
 - If x is positive, move TPR $1/n_+$ up, as we have one TP more.
 - If x is negative, move FPR 1/n_− right, as we have one FP more.

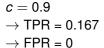


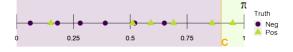


#	Truth	Score
1	Pos	0.95
2	Pos	0.86
3	Pos	0.69
4	Neg	0.65
5	Pos	0.59
6	Neg	0.52
7	Pos	0.51
8	Neg	0.39
9	Neg	0.28
10	Neg	0.18
11	Pos	0.15
12	Neg	0.06

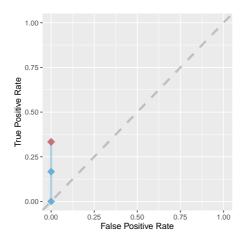




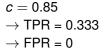


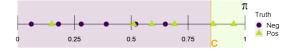


#	Truth	Score
1	Pos	0.95
2	Pos	0.86
3	Pos	0.69
4	Neg	0.65
5	Pos	0.59
6	Neg	0.52
7	Pos	0.51
8	Neg	0.39
9	Neg	0.28
10	Neg	0.18
11	Pos	0.15
12	Neg	0.06

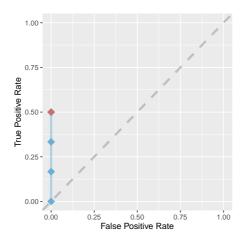








_		
#	Truth	Score
1	Pos	0.95
2	Pos	0.86
3	Pos	0.69
4	Neg	0.65
5	Pos	0.59
6	Neg	0.52
7	Pos	0.51
8	Neg	0.39
9	Neg	0.28
10	Neg	0.18
11	Pos	0.15
12	Nea	0.06



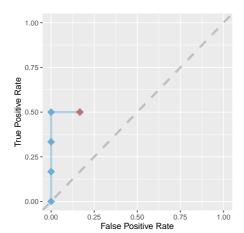








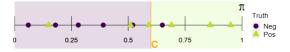
#	Truth	Score
1	Pos	0.95
2	Pos	0.86
3	Pos	0.69
4	Neg	0.65
5	Pos	0.59
6	Neg	0.52
7	Pos	0.51
8	Neg	0.39
9	Neg	0.28
10	Neg	0.18
11	Pos	0.15
12	Neg	0.06





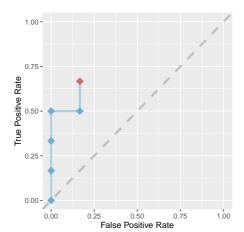
$$\rightarrow$$
 FPR = 0.167







#	Truth	Score
1	Pos	0.95
2	Pos	0.86
3	Pos	0.69
4	Neg	0.65
5	Pos	0.59
6	Neg	0.52
7	Pos	0.51
8	Neg	0.39
9	Neg	0.28
10	Neg	0.18
11	Pos	0.15
12	Nea	0.06

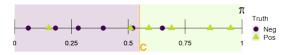




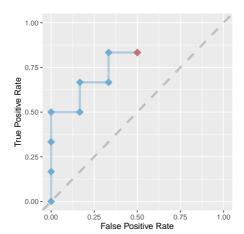


 \rightarrow TPR = 0.667

 \rightarrow FPR = 0.167



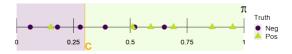
#	Truth	Score
1	Pos	0.95
2	Pos	0.86
3	Pos	0.69
4	Neg	0.65
5	Pos	0.59
6	Neg	0.52
7	Pos	0.51
8	Neg	0.39
9	Neg	0.28
10	Neg	0.18
11	Pos	0.15
12	Neg	0.06





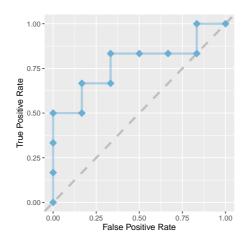


$$\rightarrow$$
 FPR = 0.5



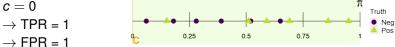


#	Truth	Score
1	Pos	0.95
2	Pos	0.86
3	Pos	0.69
4	Neg	0.65
5	Pos	0.59
6	Neg	0.52
7	Pos	0.51
8	Neg	0.39
9	Neg	0.28
10	Neg	0.18
11	Pos	0.15
12	Neg	0.06



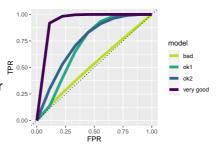






ROC CURVE PROPERTIES

- The closer the curve to the top-left corner, the better.
- If ROC curves cross, a different model might be better in different parts of the ROC space.



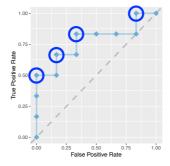


- Small thresholds will very liberally predict the positive class, and result in a potentially higher FPR, but also higher TPR.
- High thresholds will very conservatively predict the positive class, and result in a lower FPR and TPR.
- As we have not defined the trade-off between false positive and false negative costs, we cannot easily select the "best" threshold.
 - \rightarrow Visual inspection of all possible results seems useful.

CHOOSING THRESHOLD / OPERATING POINT

Often done visually and post-hoc, as class imbalances or costs are unknown a-priori.

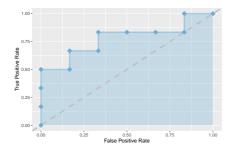
- Identify non-dominated points
- Assess TPR / FPR
- Decide which combo is best for task
- Pick associated threshold





AUC: AREA UNDER ROC CURVE

- AUC ∈ [0, 1] is a single metric to evaluate scoring classifiers independent of the chosen threshold.
 - AUC = 1: perfect classifier
 - AUC = 0.5: random, non-discriminant classifier
 - AUC = 0: perfect, with inverted labels





AUC AS A RANK-BASED METRIC

- We can also interpret the AUC as the probability of our classifier ranking a random positive observation higher than a random negative one.
- A perfect classifier will rank all positive above all negative observations, achieving AUC = 1.

