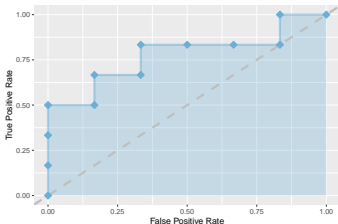


# Introduction to Machine Learning

## Evaluation: Measures for Binary Classification: ROC Visualization

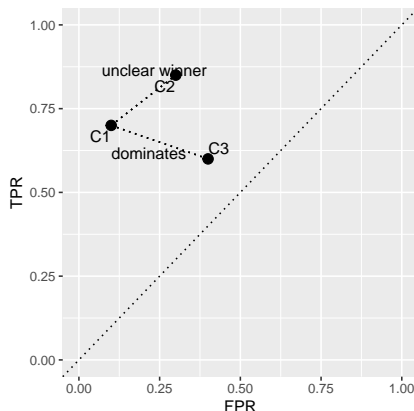


### Learning goals

- Understand the ROC curve
- Be able to compute a ROC curve manually

# 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.



		True Class $y$	
		+	-
Pred. $\hat{y}$	+	TP	FP
	-	FN	TN

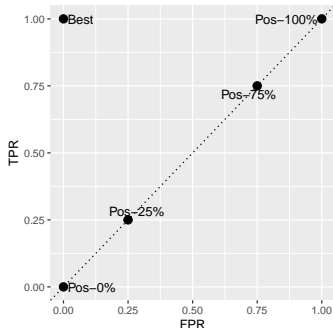
$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

# 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).

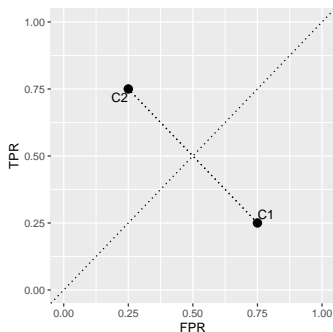
- If each positive  $x$  will be randomly classified with 25% as "pos",  $TPR = 0.25$ .
- If we assign each negative  $x$  randomly to "pos",  $FPR = 0.25$ .



# 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.

$\Rightarrow \text{TPR}_{\text{new}} = 1 - \text{TPR}$  and  $\text{FPR}_{\text{new}} = 1 - \text{FPR}$ .



# 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  $n_+/n_- = 1$

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

$$\text{MCE} = 35/100 = 0.35$$

$$\text{TPR} = 0.8$$

$$\text{FPR} = 0.5$$

## Example 2:

Proportion  $n_+/n_- = 2$

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

$$\text{MCE} = 45/150 = 0.3$$

$$\text{TPR} = 0.8$$

$$\text{FPR} = 0.5$$

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

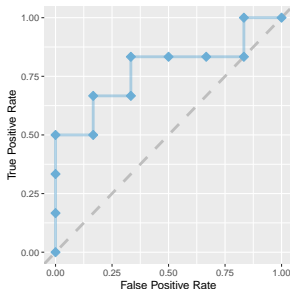
# FROM PROBABILITIES TO LABELS: ROC CURVE

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

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

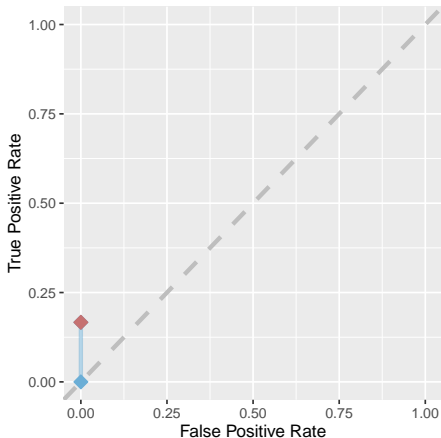
## To draw a ROC curve:

- ❶ Rank test observations on decreasing score.
- ❷ Start with  $c = 1$ , so we start in  $(0, 0)$ ; we predict everything as negative.
- ❸ 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.



# DRAWING ROC CURVES

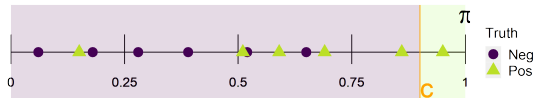
#	Truth	Score
1	Pos	0.95
2	Pos	0.86
3	Pos	0.69
4	Neg	0.65
5	Pos	0.59
6	Neg	0.52
7	Pos	0.51
8	Neg	0.39
9	Neg	0.28
10	Neg	0.18
11	Pos	0.15
12	Neg	0.06



$$c = 0.9$$

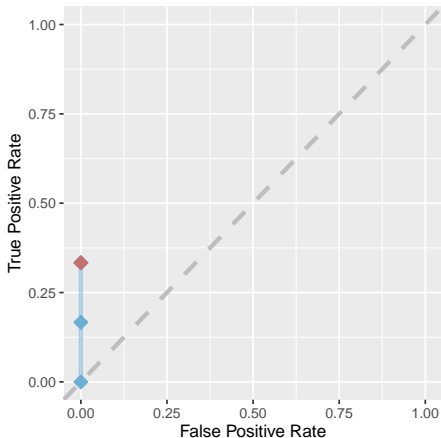
$$\rightarrow \text{TPR} = 0.167$$

$$\rightarrow \text{FPR} = 0$$



# DRAWING ROC CURVES

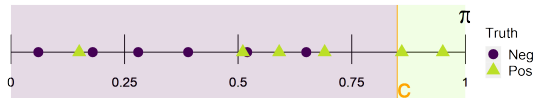
#	Truth	Score
1	Pos	0.95
2	Pos	0.86
3	Pos	0.69
4	Neg	0.65
5	Pos	0.59
6	Neg	0.52
7	Pos	0.51
8	Neg	0.39
9	Neg	0.28
10	Neg	0.18
11	Pos	0.15
12	Neg	0.06



$$c = 0.85$$

$$\rightarrow \text{TPR} = 0.333$$

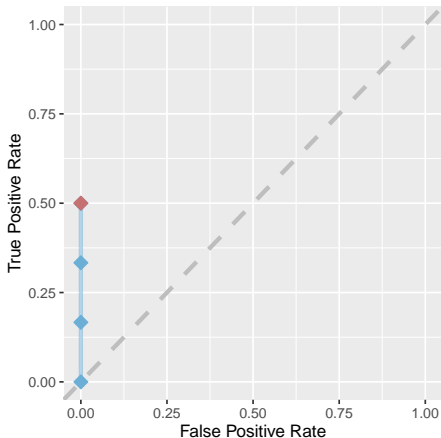
$$\rightarrow \text{FPR} = 0$$





# DRAWING ROC CURVES

#	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



$$c = 0.66$$

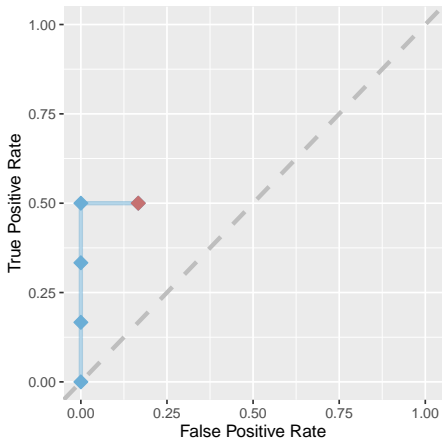
$$\rightarrow \text{TPR} = 0.5$$

$$\rightarrow \text{FPR} = 0$$



# DRAWING ROC CURVES

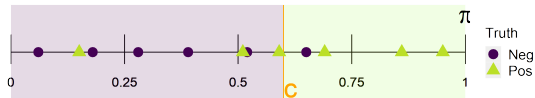
#	Truth	Score
1	Pos	0.95
2	Pos	0.86
3	Pos	0.69
4	Neg	0.65
5	Pos	0.59
6	Neg	0.52
7	Pos	0.51
8	Neg	0.39
9	Neg	0.28
10	Neg	0.18
11	Pos	0.15
12	Neg	0.06



$c = 0.6$

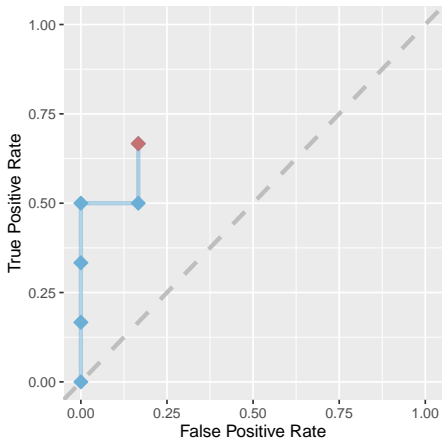
→  $TPR = 0.5$

→  $FPR = 0.167$



# DRAWING ROC CURVES

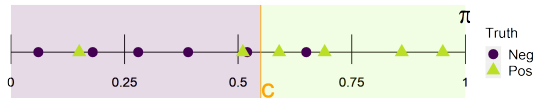
#	Truth	Score
1	Pos	0.95
2	Pos	0.86
3	Pos	0.69
4	Neg	0.65
5	Pos	0.59
6	Neg	0.52
7	Pos	0.51
8	Neg	0.39
9	Neg	0.28
10	Neg	0.18
11	Pos	0.15
12	Neg	0.06



$c = 0.55$

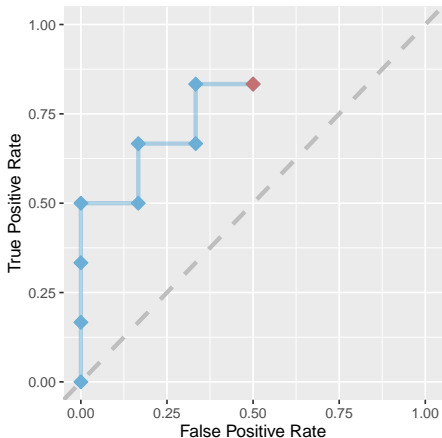
→ TPR = 0.667

→ FPR = 0.167



# DRAWING ROC CURVES

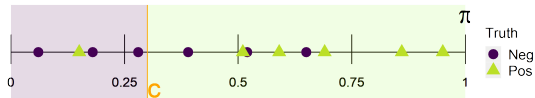
#	Truth	Score
1	Pos	0.95
2	Pos	0.86
3	Pos	0.69
4	Neg	0.65
5	Pos	0.59
6	Neg	0.52
7	Pos	0.51
8	Neg	0.39
9	Neg	0.28
10	Neg	0.18
11	Pos	0.15
12	Neg	0.06



$c = 0.3$

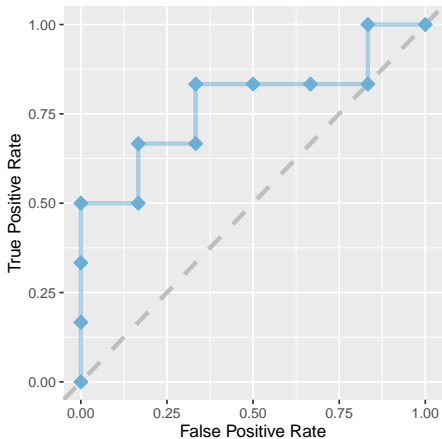
→  $TPR = 0.833$

→  $FPR = 0.5$



# DRAWING ROC CURVES

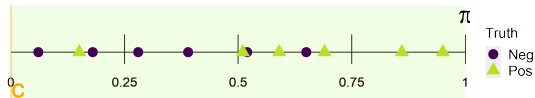
#	Truth	Score
1	Pos	0.95
2	Pos	0.86
3	Pos	0.69
4	Neg	0.65
5	Pos	0.59
6	Neg	0.52
7	Pos	0.51
8	Neg	0.39
9	Neg	0.28
10	Neg	0.18
11	Pos	0.15
12	Neg	0.06



$c = 0$

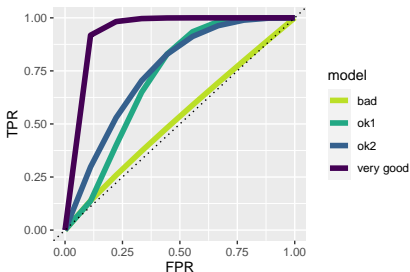
→ TPR = 1

→ FPR = 1



# ROC CURVE PROPERTIES

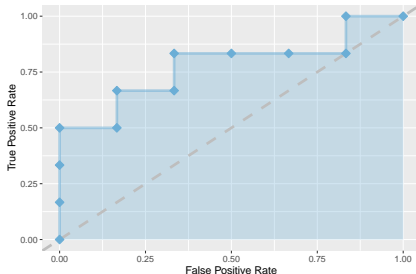
- The closer the curve to the top-left corner, the better.
- If ROC curves cross, a different model might be better in different parts of the ROC space.



- 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 positives and false negative costs, we cannot easily select the "best" threshold.  
→ Visual inspection of all possible results seems useful.

# AUC: AREA UNDER ROC CURVE

- $AUC \in [0, 1]$  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  $AUC = 1$ .

Truth	Score
1	0.9
1	0.76
1	0.7
0	0.5
1	0.45
0	0.3
0	0.1

Choose a random positive



1	0.76
---	------

Choose a random negative



0	0.3
---	-----

AUC = 0.9167

- Classifier ranks the positive higher than the negative
- This happens with a mean probability of 0.9167

