

# COST CURVES

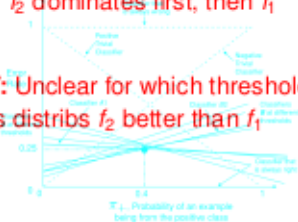
- Directly plot the misclassification costs / error (in terms of prior probs)
  - Might be easier to interpret than ROC, especially in case of different misclassification costs or priors
- ## Imbalanced Learning:



## Example: Cost Curves Part 1

- $f_1$  and  $f_2$  with intersecting ROC curves
- $f_2$  dominates first, then  $f_1$

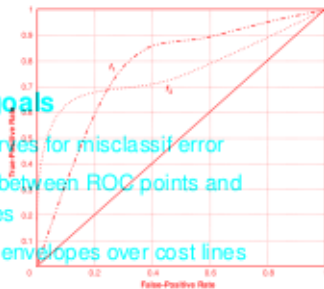
**BUT:** Unclear for which thresholds, costs or class distribs  $f_2$  better than  $f_1$



## Learning goals

- Cost curves for misclassification error
- Duality between ROC points and cost lines
- CCs as envelopes over cost lines

ROC curves for  $f_1$  and  $f_2$



Nathalie Japkowicz (2004): Evaluating Learning Algorithms : A Classification Perspective. (p. 125)

# COST CURVES

- Simplifying assumption: equal misclassif costs, i.e.,  $cost_{FN} = cost_{FP}$
- $\Rightarrow$  Expected misclassif cost reduces to misclassif error rate
- $\Rightarrow$  Might be easier to interpret than ROC, especially in case of different misclassif costs or priors

With law of total prob, we write error rate as function of  $\pi_+$ :

$$\begin{aligned} \rho_{MCE}(\pi_+) &= (1 - \pi_+) \cdot P(\hat{y} = 1 | y = 0) + \pi_+ \cdot P(\hat{y} = 0 | y = 1) \\ &= (1 - \pi_+) \cdot FPR + \pi_+ \cdot FNR \\ &= (FNR - FPR) \cdot \pi_+ + FPR \end{aligned}$$

Example:

- $f_1$  and  $f_2$  with intersecting ROC curves
- $f_2$  dominates  $f_1$  when  $f_1$

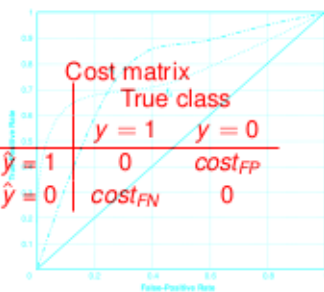
Confusion matrix  
True class

	$y = 1$	$y = 0$
Pred. $\hat{y} = 1$	TP	FP
Pred. $\hat{y} = 0$	FN	TN

But, unclear for which thresholds, costs or class distrib  $f_2$  better than  $f_1$

	True class $y = 1$	True class $y = 0$
Pred. $\hat{y} = 1$	0	$cost_{FP}$
Pred. $\hat{y} = 0$	$cost_{FN}$	0

ROC curves for  $f_1$  and  $f_2$



Nathalie Japkowicz (2004): Evaluating Learning Algorithms : A Classification Perspective. (p. 125)



# COST CURVES

Simplifying assumption: equal misclassif costs, i.e.  $cost_{FN} = cost_{FP}$

⇒ Expected misclassif cost reduces to misclassif error rate

With law of total prob, we write error rate as function of  $\pi_+$ :

- Cost curves are point-line duals of ROC curves, i.e., a single classifier is represented by a point in the ROC space and by a line in cost space

$$PMCE(\pi_+) = (1 - \pi_+) \cdot (y = 1, \hat{y} = 0) + \pi_+ \cdot (y = 0, \hat{y} = 1)$$

$$= (1 - \pi_+) \cdot FPR + \pi_+ \cdot FNR$$

$$= (FNR - FPR) \cdot \pi_+ + FPR$$

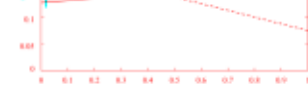
Confusion matrix

True class

$y = 1$     $y = 0$

Pred.  $\hat{y} = 1$    TP   FP

class  $\hat{y} = 0$    FN   TN



$\pi_+$  - Probability of Positive

Cost matrix

True class

$y = 1$     $y = 0$

Pred.  $\hat{y} = 1$    0    $cost_{FP}$

class  $\hat{y} = 0$     $cost_{FN}$    0



False Positive Rate



Chris Drummond and Robert C. Holte (2006): Cost curves: An improved method for visualizing classifier performance. Machine Learning, 65, 95-130 ([URL](#)).

# COST LINES

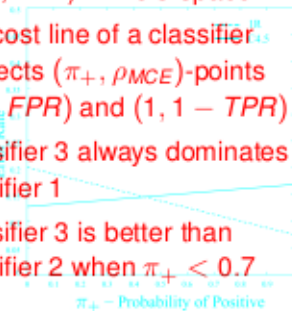
Cost line of a classifier with slope  $(FNR - FPR)$  and intercept  $FPR$ :

$$\rho_{MCE}(\pi_+) = (FNR - FPR) \cdot \pi_+ + FPR$$

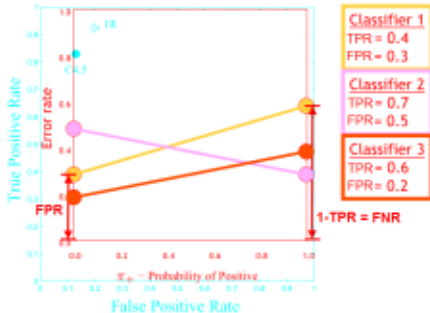
• Cost curves are point-line duals of ROC curves, i.e., a single classifier is

- Hard classifiers are points  $(TPR, FPR)$  in ROC space
- Cost lines plot different values of  $\pi_+$  vs.  $\rho_{MCE}(\pi_+)$

- The cost line of a classifier connects  $(\pi_+, \rho_{MCE})$ -points at  $(0, FPR)$  and  $(1, 1 - TPR)$
- Classifier 3 always dominates classifier 1
- Classifier 3 is better than classifier 2 when  $\pi_+ < 0.7$



$\pi_+$  vs.  $\rho_{MCE}(\pi_+)$



Chris Drummord and Robert C. Holte (2006): Cost curves: An improved method for visualizing classifier performance. Machine Learning, 65, 95-130 ([URL](#)).

# COST LINES - EXAMPLE

• Horizontal dashed line: worst classifier (100% error rate for all  $\pi_+$ )  $\Rightarrow FNR = FPR = 1$

$\Rightarrow FNR = FPR = 1$

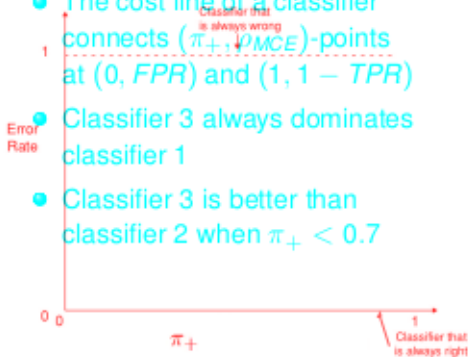
• x-axis: perfect classifier (0% error rate for all  $\pi_+$ )  $\Rightarrow FNR = FPR = 0$

• Hard classifiers are points (TPR, FPR) in ROC space

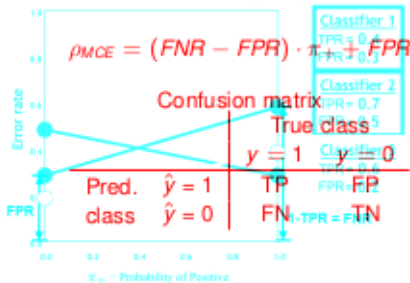
• The cost line of a classifier connects  $(\pi_+, \rho_{MCE})$ -points at  $(0, FPR)$  and  $(1, 1 - TPR)$

• Classifier 3 always dominates classifier 1

• Classifier 3 is better than classifier 2 when  $\pi_+ < 0.7$

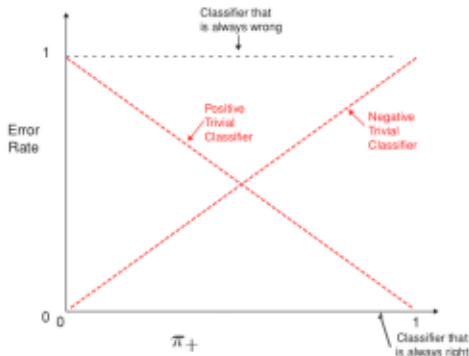


Cost lines plot different values of  $\pi_+$  vs.  $\rho_{MCE}(\pi_+)$



# COST LINES - EXAMPLE

- Horizontal dashed line: worst classifier (100% error rate for all  $\pi_+$ )  
 $\Rightarrow FNR = FPR = 1$
- x-axis: perfect classifier (0% error rate for all  $\pi_+$ )  $\Rightarrow FNR = FPR = 0$
- Dashed diagonal lines: trivial classifiers, i.e., ascending diagonal always predicts negative instances ( $\Rightarrow FNR = 1$  and  $FPR = 0$ ) and vice versa

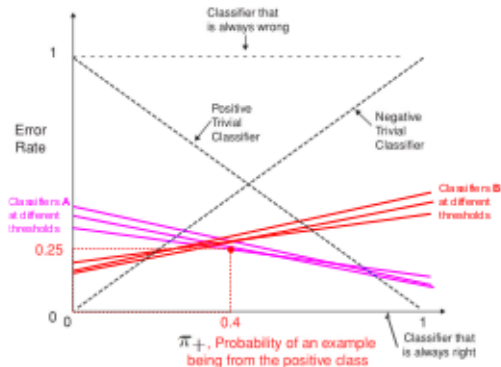


$$\rho_{MCE} = (FNR - FPR) \cdot \pi_+ + FPR$$

		True class	
		$y = 1$	$y = 0$
Pred. class	$\hat{y} = 1$	TP	FP
	$\hat{y} = 0$	FN	TN

# COST LINES - EXAMPLE

- Horizontal dashed line: worst classifier (100% error rate for all  $\pi_+$ )  
 $\Rightarrow FNR = FPR = 1$
- x-axis: perfect classifier (0% error rate for all  $\pi_+$ )  $\Rightarrow FNR = FPR = 0$
- Dashed diagonal lines: trivial classifiers, i.e., ascending diagonal always predicts negative instances ( $\Rightarrow FNR = 1$  and  $FPR = 0$ ) and vice versa
- Descending/ascending bold lines: two families of classifiers *A* and *B* (represented by points in their respective ROC curves)

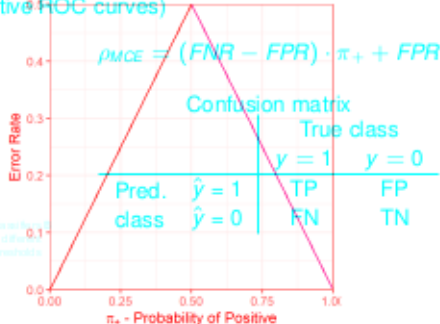
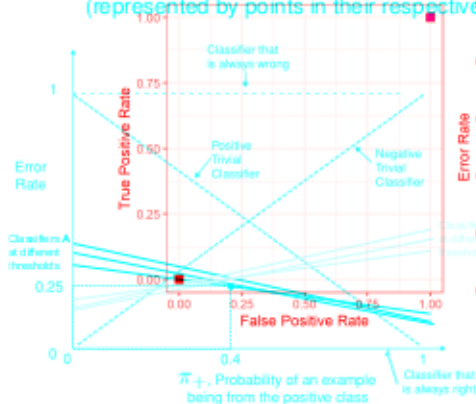


$$PMCE = (FNR - FPR) \cdot \pi_+ + FPR$$

		Confusion matrix	
		True class	
		$y = 1$	$y = 0$
Pred. class	$\hat{y} = 1$	TP	FP
	$\hat{y} = 0$	FN	TN

# VISUALIZE COST CURVE - LOWER ENVELOPE

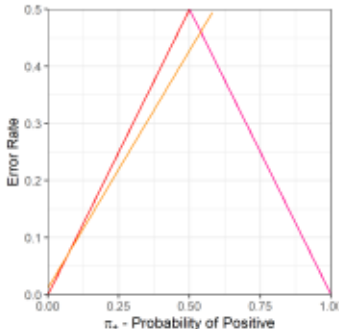
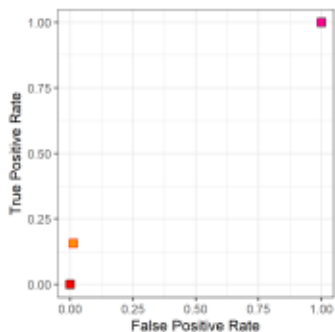
- Left: ROC diagram for best classifier (100% error rate for all  $\pi_+$ )  
 $\Rightarrow FNR = FPR = 1$
- Right: Corresponding cost lines
- x-axis: perfect classifier (0% error rate for all  $\pi_+$ )  $\Rightarrow FNR = FPR = 0$
- Duality: For every ROC point we can construct the CC line, and vice versa.
- Dashed diagonal lines: trivial classifiers, i.e., ascending diagonal always predicts negative instances ( $\Rightarrow FNR = 1$  and  $FPR = 0$ ) and vice versa
- Descending/ascending bold lines: two families of classifiers A and B (represented by points in their respective ROC curves)





# VISUALIZE COST CURVE - LOWER ENVELOPE

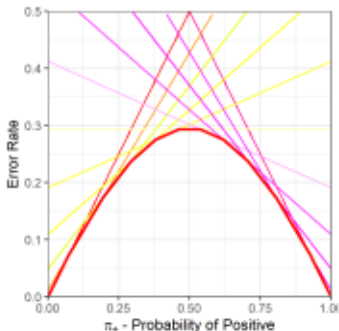
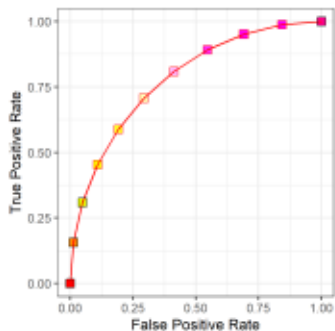
- Left: ROC = TPR & FPR of a classifier for different prob thresholds
- Right: Corresponding cost lines
- Duality: For every ROC point we can construct the CC line, and vice versa.





# VISUALIZE COST CURVE - LOWER ENVELOPE

- Left: ROC = TPR & FPR of a classifier for different prob thresholds
- Right: Corresponding cost lines
- Duality: For every ROC point we can construct the CC line, and vice versa.
- **Cost curve (right, black) is lower envelope of cost lines**  
≙ pointwise minimum of error rate (as function of  $\pi_+$ )



# VISUALIZE COST CURVE - LOWER ENVELOPE

- Left: ROC = TPR & FPR of a classifier for different prob thresholds
- Right: Corresponding cost lines
- Duality: For every ROC point we can construct the CC line, and vice versa.
- **Cost curve** (right, black) is lower envelope of **cost lines**  
≙ pointwise minimum of error rate (as function of  $\pi_+$ )

