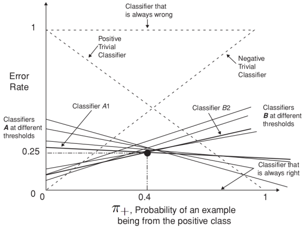
Introduction to Machine Learning

Evaluation: Cost Curves



Learning goals

- Understand cost curves
- As alternative to ROC curves

COST CURVES

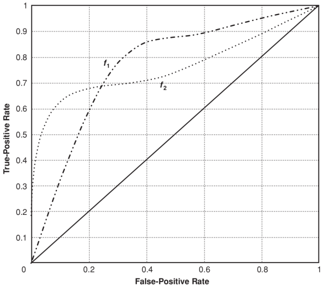
- Directly plot the misclassif costs / error
- Might be easier to interpret than ROC, especially in case of different misclassif costs or priors

Example:

- 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

ROC curves for f_1 and f_2



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

COST CURVES

With law total probab, can write misclassif rate as function of π_+:

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

Can do the same for costs:

$$Costs(\pi_+) = (1-\pi_+)\cdot FPR \cdot cost_{FP} + \pi_+ \cdot FNR \cdot cost_{FN}$$

$$Costs_{norm}(\pi_+) = \frac{(1-\pi_+)\cdot FPR\cdot cost_{FP} + \pi_+\cdot FNR\cdot cost_{FN}}{(1-\pi_+)\cdot cost_{FP} + \pi_+\cdot cost_{FN}} \in [0,1]$$

- Denominator = $\max(Costs)$ = all obs misclassified (i.e., $FPR = FNR = 1$).
- If $cost_{FN} = cost_{FP}$, than $Costs_{norm} = \rho_{MCE}$

Confusion matrix

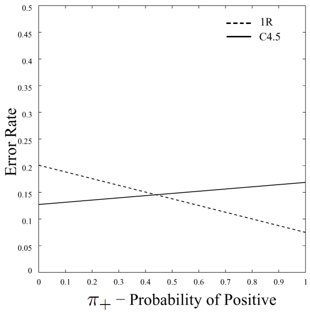
		True class	
		$y=1$	$y=0$
Pred.	$\hat{y}$ = 1	TP	FP
class	$\hat{y}$ = 0	FN	TN

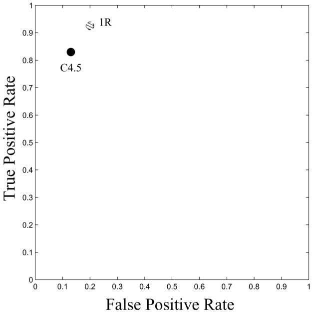
Cost matrix

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

COST CURVES

- Simplifying assumption: equal misclassification costs, i.e., $cost_{FN} = cost_{FP}$.
- Normalized costs (or error rate in the case of $cost_{FN} = cost_{FP}$) is plotted as a function of the proportion of positive instances, $\pi_+ = \mathbb{P}(y = 1)$.
- 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 the cost space.





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

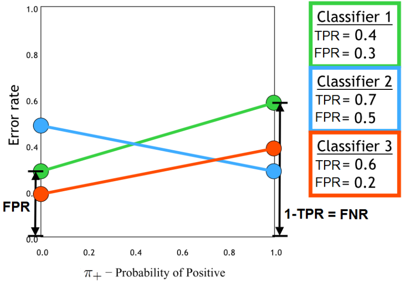
COST CURVES - EXAMPLE

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

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

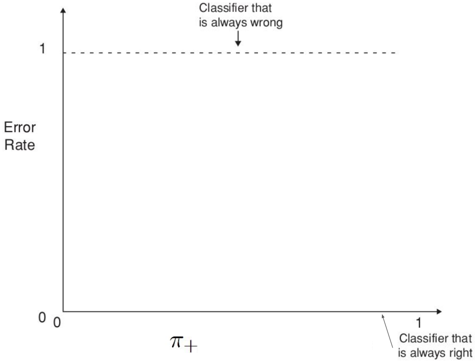
- Hard classifiers are points (TPR, FPR) in ROC space
- The cost curve of a classifier connects (π_+, ρ_{MCE})-points at $(0, FPR)$ and $(1, 1 - TPR)$
- Classifier 3 always dominates classifier 1
- Classifier 3 is better than classifier 2 when $\pi_+ < 0.7$

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



COST CURVES - EXAMPLE

- Horizontal dashed line: worst classifier (100% error rate for all π_+).
  $\Rightarrow$ *FNR* $=$ *FPR* $= 1$
- x-axis: perfect classifier (0% error rate for all π_+). $\Rightarrow$ *FNR* $=$ *FPR* $= 0$



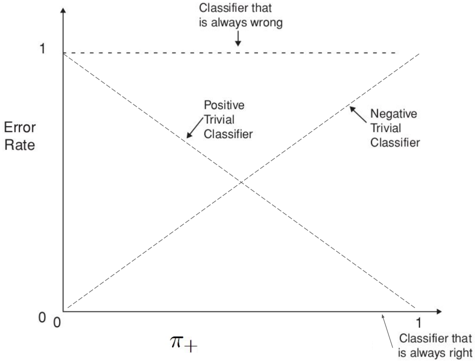
$$\rho_{MCE} = (FNR - FPR) \cdot \pi_+ + FPR$$

Confusion matrix

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

COST CURVES - EXAMPLE

- Horizontal dashed line: worst classifier (100% error rate for all π_+).
  $\Rightarrow$ *FNR* $=$ *FPR* $= 1$
- x-axis: perfect classifier (0% error rate for all π_+). $\Rightarrow$ *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.



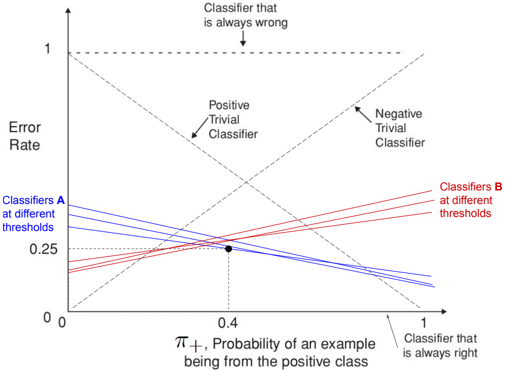
$$\rho_{MCE} = (FNR - FPR) \cdot \pi_+ + FPR$$

Confusion matrix

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

COST CURVES - EXAMPLE

- Horizontal dashed line: worst classifier (100% error rate for all π_+).
  $\Rightarrow FNR = FPR = 1$
- x-axis: perfect classifier (0% error rate for all π_+). $\Rightarrow FNR = FPR = 0$
- 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).



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

Confusion matrix

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

ROC CURVES VS. COST CURVES

- ROC curves do not indicate in which situations classifier A is superior to another classifier B.
- Cost curves can do exactly that and therefore provide practically more relevant information than ROC curves.
- For simplification, we focused on cost curves based on the misclassification error by assuming $cost_{FN} = cost_{FP}$. However, cost curves can also be defined for different misclassification costs.