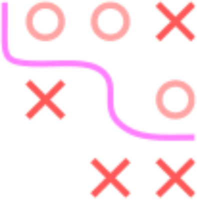
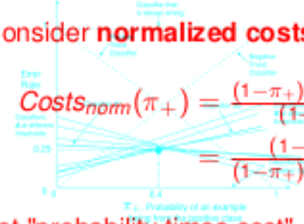
Introduction to Machine Learning

Imbalanced Learning: Cost Curves Part 2





Learning goals

- CCs with cost matrices
- Comparing classifiers
- Wrap-Up comparison to ROC

CCS WITH TRUE COSTS

Assume unequal misclassif costs, i.e., $cost_{FN} \neq cost_{FP}$ and generalize error rate to **expected costs** (as function of π_+):

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

Maximum of expected costs happens when

$$FPR = FNR = 1 \;\Rightarrow\; Costs_{max} = (1-\pi_+) \cdot cost_{FP} + \pi_+ \cdot cost_{FN}$$

Consider **normalized costs** (as function of π_+):

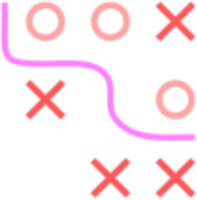
$$Costs_{norm}(\pi_+) = \frac{(1-\pi_+) \cdot FPR \cdot cost_{FP} + \pi_+ \cdot FNR \cdot cost_{FN}}{(1-\pi_+) \cdot cost_{FP} + \pi_+ \cdot cost_{FN}}$$

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

Let "probability times cost" $PC(+)$ be normalized version of $\pi_+ \cdot cost_{FN}$:

$$PC(+) = \frac{\pi_+ \cdot cost_{FN}}{(1-\pi_+) \cdot cost_{FP} + \pi_+ \cdot cost_{FN}} \text{ and } 1 - PC(+) = \frac{(1-\pi_+) \cdot cost_{FP}}{(1-\pi_+) \cdot cost_{FP} + \pi_+ \cdot cost_{FN}}$$

CCS WITH TRUE COSTS / 2



To obtain cost lines, we need a function with slope $(FNR - FPR)$ and intercept FPR ⇒ Rewrite $Costs_{norm}(\pi_+)$ as function of $PC(+)$:

$$
\begin{aligned}
Costs_{norm}(PC(+)) &= (1 - PC(+)) \cdot FPR + PC(+) \cdot FNR \\
&= (FNR - FPR) \cdot PC(+) + FPR \\
&= \begin{cases} FPR, \text{if } PC(+) = 0 \\ FNR, \text{if } PC(+) = 1 \end{cases}
\end{aligned}
$$

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

Maximum of expected costs happens when

$$FPR = FNR = 1 \;\Rightarrow\; Costs_{max} = (1 - \pi_+) \cdot cost_{FP} + \pi_+ \cdot cost_{FN}$$

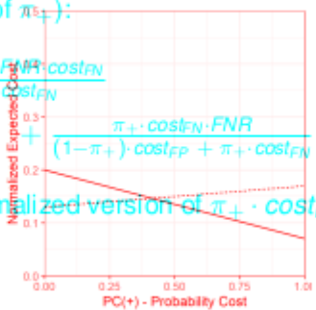
Consider **normalized costs** (as function of π_+):

$$Costs_{norm}(\pi_+) = \frac{(1-\pi_+) \cdot FPR \cdot cost_{FP} + \pi_+ \cdot FNR \cdot cost_{FN}}{(1-\pi_+) \cdot cost_{FP} + \pi_+ \cdot cost_{FN}} = FPR \ldots + \frac{\pi_+ \cdot cost_{FN} \cdot FNR}{(1-\pi_+) \cdot cost_{FP} + \pi_+ \cdot cost_{FN}}$$

- Plot is similar to simplified case with $cost_{FN} = cost_{FP}$
- Axes' labels and their interpretation have changed
- Normalized cost vs. "probability times cost"

Let "probability times cost" $PC(+)$ be normalized version of $\pi_+ \cdot cost_{FN}$:

$$PC(+) = \frac{\pi_+ \cdot cost_{FN}}{(1-\pi_+) \cdot cost_{FP} + \pi_+ \cdot cost_{FN}}$$ and



COMPARE WITH TRIVIAL CLASSIFIERS

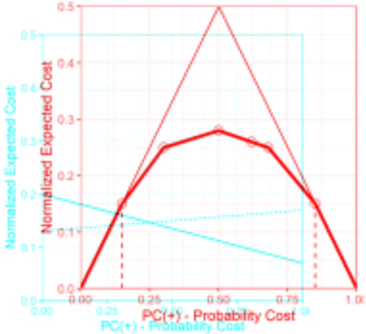
To obtain cost lines, we need a function with slope $(FNR - FPR)$ and intercept FPR → Rewrite $Cost_{norm}(\pi_+)$ as function of $PC(+)$:

$$Cost_{norm}(PC(+)) = (1 - PC(+)) \cdot FPR + PC(+) \cdot FNR$$
$$= (FNR - FPR) \cdot PC(+) + FPR$$
$$= \begin{cases} FPR & \text{if } PC(+) = 0 \\ FNR & \text{if } PC(+) = 1 \end{cases}$$

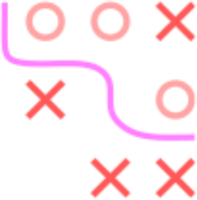
- Operating range of a classifier is a set of $PC(+)$ values (operating points) where classifier performs better than both trivial classifiers
- Intersection of cost curves and trivial classifiers' diagonals determine operating range
- At any $PC(+)$ value, the vertical distance of trivial diagonal to a classifer's cost curve within operating range shows advantage in performance (normalized costs) of classifier

Example: Dotted lines are operating range of a classifier (here: [0.14, 0.85])

- Plot is similar to simplified case with $cost_{FN} = cost_{FP}$
- Axes' labels and their interpretation have changed
- Normalized cost vs. "probability times cost"

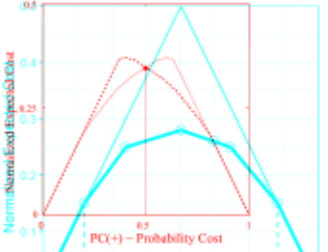


COMPARING CLASSIFIERS



- If classifier C1's expected cost is lower than classifier C2's at a $PC(+)$ value, C1 outperforms C2 at that operating point
- The two cost curves of C1 and C2 may cross, which indicates C1 outperforms C2 for a certain operating range and vice versa
- The vertical distance between the two cost curves of C1 and C2 at any $PC(+)$ value directly indicates the performance difference between them at that operating point

Example: Dotted cost curve has lower expected cost as dashed cost curve for $PC(+) < 0.5$ and hence outperforms dashed one in this operating range and vice versa



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

COMPARING CLASSIFIERS

- If classifier C1's expected cost is lower than classifier C2's at a $PC(+)$ value, C1 outperforms C2 at that operating point
- The two cost curves of C1 and C2 may cross, which indicates C1 outperforms C2 for a certain operating range and vice versa
- The vertical distance between the two cost curves of C1 and C2 at any $PC(+)$ value directly indicates the performance difference between them at that operating point

Example: Dotted cost curve has lower expected cost as dashed cost curve for $PC(+) < 0.5$ and hence outperforms dashed one in this operating range and vice versa

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