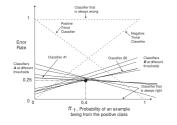
Introduction to Machine Learning

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_+):

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

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_+):

$$\begin{aligned} \textit{Costs}_{\textit{norm}}(\pi_{+}) &= \frac{(1-\pi_{+}) \cdot \textit{FPR} \cdot \textit{cost}_\textit{FP} + \pi_{+} \cdot \textit{FNR} \cdot \textit{cost}_\textit{FN}}{(1-\pi_{+}) \cdot \textit{cost}_\textit{FP} + \pi_{+} \cdot \textit{cost}_\textit{FN}} \\ &= \frac{(1-\pi_{+}) \cdot \textit{cost}_\textit{FP} \cdot \textit{FPR}}{(1-\pi_{+}) \cdot \textit{cost}_\textit{FP} + \pi_{+} \cdot \textit{cost}_\textit{FN}} + \frac{\pi_{+} \cdot \textit{cost}_\textit{FN} \cdot \textit{FNR}}{(1-\pi_{+}) \cdot \textit{cost}_\textit{FP} + \pi_{+} \cdot \textit{cost}_\textit{FN}} \end{aligned}$$

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



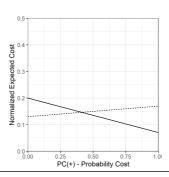
CCS WITH TRUE COSTS

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

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



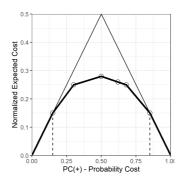
- Plot is similar to simplified case with cost_{FN} = cost_{FP}
- Axes' labels and their interpretation have changed
- Normalized cost vs.
 "probability times cost"



COMPARE WITH TRIVIAL CLASSIFIERS

- 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])

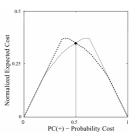




COMPARING CLASSIFIERS

- 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)

