Cost-Sensitive Classification Meets Proper Scoring Rules

Peter A. Flach

Intelligent Systems Laboratory, University of Bristol, United Kingdom

11 September 2015

Talk outline

Introduction and motivation

Brier score and Brier curve

Harmonic cost contexts and Log-Loss

A family of cost contexts

Discussion

Cost context

In cost-sensitive classification we penalise misclassified positives with a cost c_0 and misclassified negatives with a cost c_1, jointly referred to as the *cost context*.

It is usually assumed that only the *cost proportion* $c = c_0/(c_0 + c_1)$ matters, so that the following are all equivalent:

$c_0 = 1, c_1 = 3$;

$c_0 = 2/3, c_1 = 2$; and

$c_0 = 1/2, c_1 = 3/2$.

Furthermore, the latter cost context has the advantage of leading to cost-sensitive loss being expressed on a scale commensurate with error rate, which has cost context $c_0 = 1, c_1 = 1$.

From additive to harmonic cost contexts

This carries an implicit assumption that costs are expressed on an additive scale, summing up to a fixed budget. It is then natural to investigate what happens when we assume different scales. For example, the cost context $c_0 = 2/3, c_1 = 2$ is commensurate with error rate if we measure costs on a harmonic scale, since $1/c_0 + 1/c_1 = 2$.

The main technical result of the paper is that, for harmonic cost contexts, expected loss of a probabilistic classifier which sets its decision threshold equal to c, averaged over uniform c, is equal to the model's Log-Loss (while it is equal to the model's Brier score for additive cost contexts).

Both Brier score and Log-Loss are so-called *proper scoring rules* used to evaluate probability estimators. The cost-based perspective then allows to enumerate a family of candidate proper scoring rules.

Cost-sensitive loss

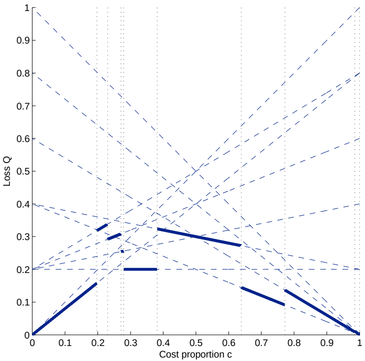
$$b = c_0 + c_1$$
$$c_0 = bc, c_1 = b(1-c)$$
$$Q(t;\pi_0,b,c) = b\{c\pi_0(1-F_0(t)) + (1-c)\pi_1 F_1(t)\}$$

Expected loss for b, c independent and π_0 fixed

$$L = \mathbb{E}\{b\}\int_0^1 \{c\pi_0(1-F_0(t)) + (1-c)\pi_1 F_1(t)\}\, w(c)dc$$

Brier score obtains if $\mathbb{E}\{b\} = 2$ and c uniform

$$BS = \pi_0 \int_0^1 s^2 f_0(s)ds + \pi_1 \int_0^1 (1-s)^2 f_1(s)ds$$



(left) The Brier curve is a piecewise linear cost curve (solid lines) jumping between different cost lines (dashed lines). The vertical dotted lines denote actual scores assigned by the model, and hence determine the values of c where the operating point changes. Score-driven thresholds are sub-optimal whenever the Brier curve departs from the lower envelope. The area under the Brier curve is the Brier score. **(right)** Optimal Brier curve resulting from perfectly calibrated scores (dashed verticals).

Harmonic cost contexts

$$1/d = 1/c_0 + 1/c_1$$
$$c_0 = d/(1-c), c_1 = d/c$$
$$Q(t;\pi_0,d,c) = d\left\{\frac{1}{1-c}\pi_0(1-F_0(t)) + \frac{1}{c}\pi_1 F_1(t)\right\}$$

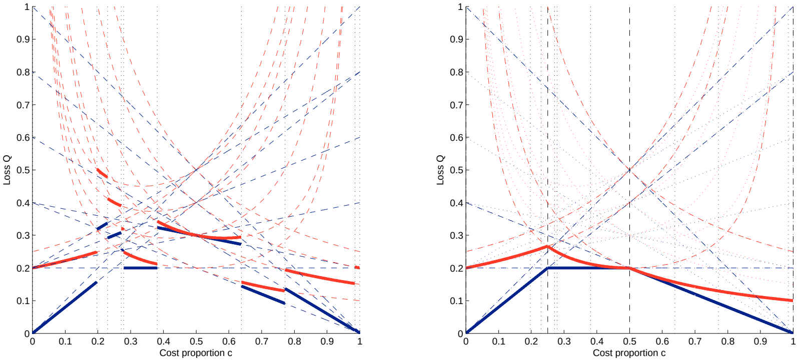
Expected loss for d, c independent

$$L = \mathbb{E}\{d\}\int_0^1 \left\{\frac{1}{1-c}\pi_0(1-F_0(t)) + \frac{1}{c}\pi_1 F_1(t)\right\} w(c)dc$$

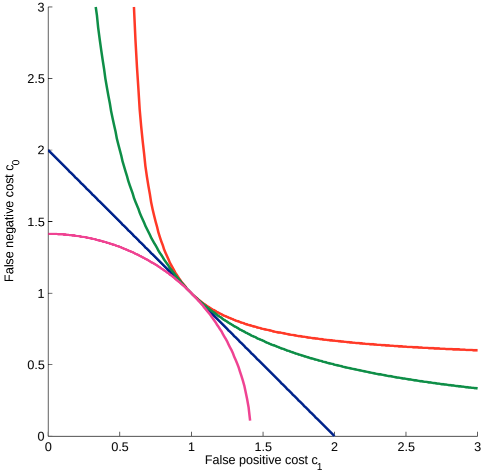
Theorem

If $\mathbb{E}\{d\} = 1/2$ and c uniform, expected loss is equal to Log-Loss:

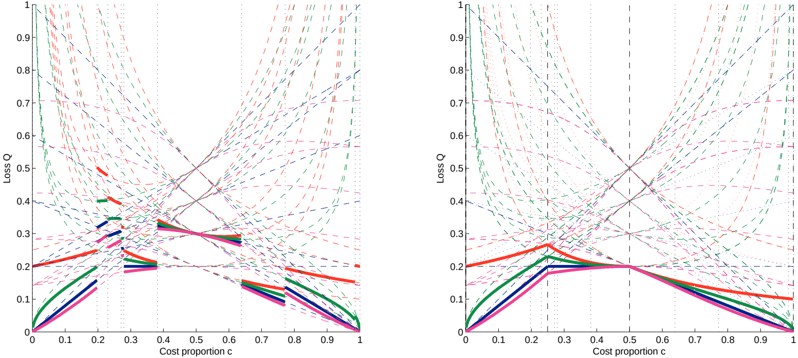
$$LL = \pi_0/2\int \ln\frac{1}{1-s} f_0(s)ds + \pi_1/2\int \ln\frac{1}{s} f_1(s)ds$$



(left) Comparison between Brier curve (in blue) and Log-Loss curve (in **red**). Comparatively, the Log-Loss curve is more affected by the model's behaviour at extreme values of c. **(right)** Optimal Brier and Log-Loss curves.



Different cost contexts: additive (in blue), harmonic (in **red**), geometric (in green) and Euclidean (in violet). All of these are scaled to be commensurate to error rate, which means that they all go through $(c_0 = 1, c_1 = 1)$.



(left) In addition to the Brier curve for additive cost contexts (in blue) and the Log-Loss curve for harmonic contexts (in **red**), this plot shows cost curves for geometric cost contexts (in green) and Euclidean cost contexts (in violet).
(right) Corresponding optimal cost curves.

Discussion

The cost-sensitive perspective helps to understand the difference between scoring rules and evaluation metrics.

Ultimately, the ‘correct’ parametrisation of cost contexts, and the corresponding evaluation metric, depends on the domain of application.

Log-Loss is usually justified by information-theoretic considerations, but from a cost-sensitive perspective the justification for a harmonic cost scale is less clear.

Euclidean cost contexts de-emphasise extreme values of c, which might be a good thing.