BETA CALIBRATION

MEELIS KULL, TELMO M. SILVA FILHO AND PETER FLACH
meelis.kull@ut.ee, tmsf@cin.ufpe.br, peter.flach@bristol.ac.uk

CALIBRATED CLASSIFIER

A binary classifier is calibrated, if:

- **it outputs a probability** of the instance to be positive instead of outputting a class label;
- **and that probability is calibrated**, i.e. it is equal to the proportion of positives among all instances with the same predicted probability;

WORKS IN ANY COST CONTEXT

The same calibrated classifier works for any false positive and false negative cost context without retraining:

1. Learn a calibrated classifier;
2. Apply the classifier on the given test instance to obtain an estimate $\hat{p}$ of its probability to be positive;
3. Determine the costs c_{FP} and c_{FN} per false positive and false negative;
4. Predict positive if $\hat{p} > c_{FP}/(c_{FP}+c_{FN})$, otherwise predict negative.

Prediction and cost-sensitive decision making have been separated. This is required when the misclassification costs are not known during model training.

CLASSIFIER CALIBRATION

If the classifier outputs non-calibrated probabilities or any real-valued scores, then it can still be calibrated:

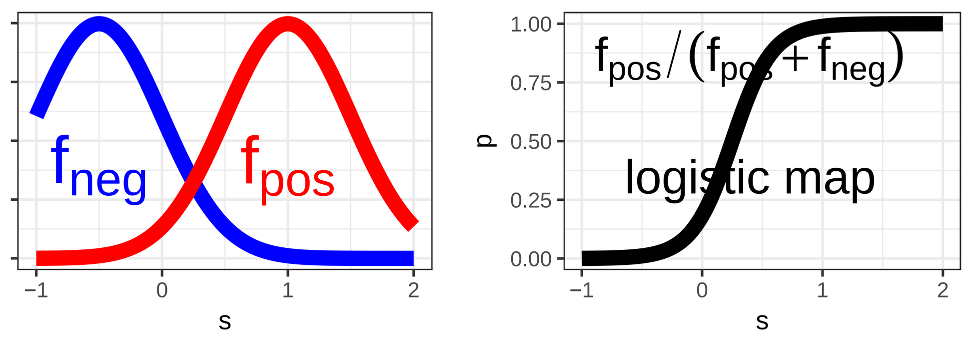
1. Learn a calibration map μ from classifier outputs to calibrated probabilities;
2. Apply the classifier on the given test instance to obtain the non-calibrated score s;
3. Remap the score into a calibrated probability $\hat{p} = \mu(s)$.

LOGISTIC CALIBRATION

- Also known as Platt scaling [Platt 2000]
- Fits a parametric family with 2 parameters:

$$\mu_{logistic}(s;\gamma,\delta) = \frac{1}{1+1/(e^{\gamma\cdot s+\delta})}$$

- Family contains only sigmoids.
- Logistic calibration is perfect if the class-conditional score densities f_{neg} and f_{pos} are Gaussian with equal variance.



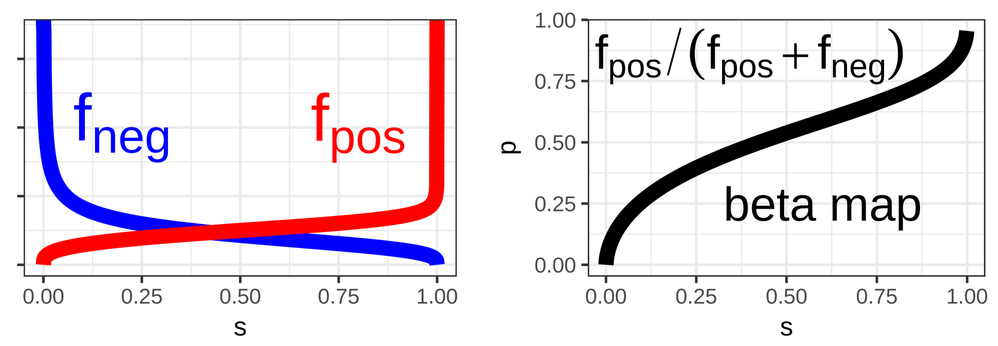
- Easily implemented by fitting logistic regression on the single feature s.

BETA CALIBRATION

- Our novel contribution.
- Fits a parametric family with 3 parameters:

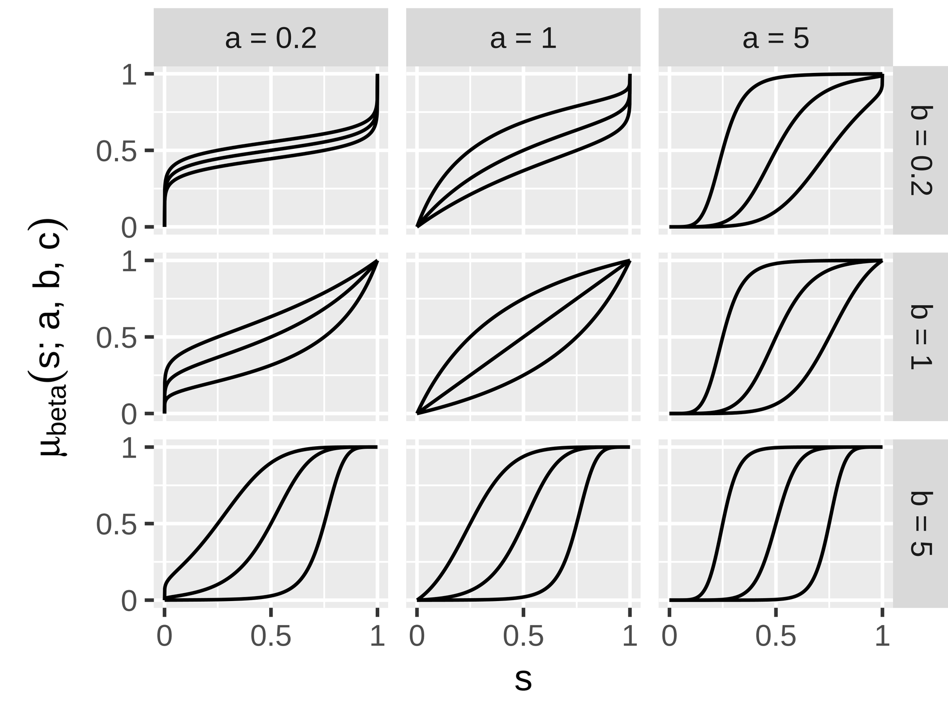
$$\mu_{beta}(s;a,b,c) = \frac{1}{1+1\Big/\left(e^{c}\frac{s^{a}}{(1-s)^{b}}\right)}$$

- Sigmoids, inverse sigmoids, identity and more.
- Beta calibration is perfect if the class-conditional score densities f_{neg} and f_{pos} are beta distributions.

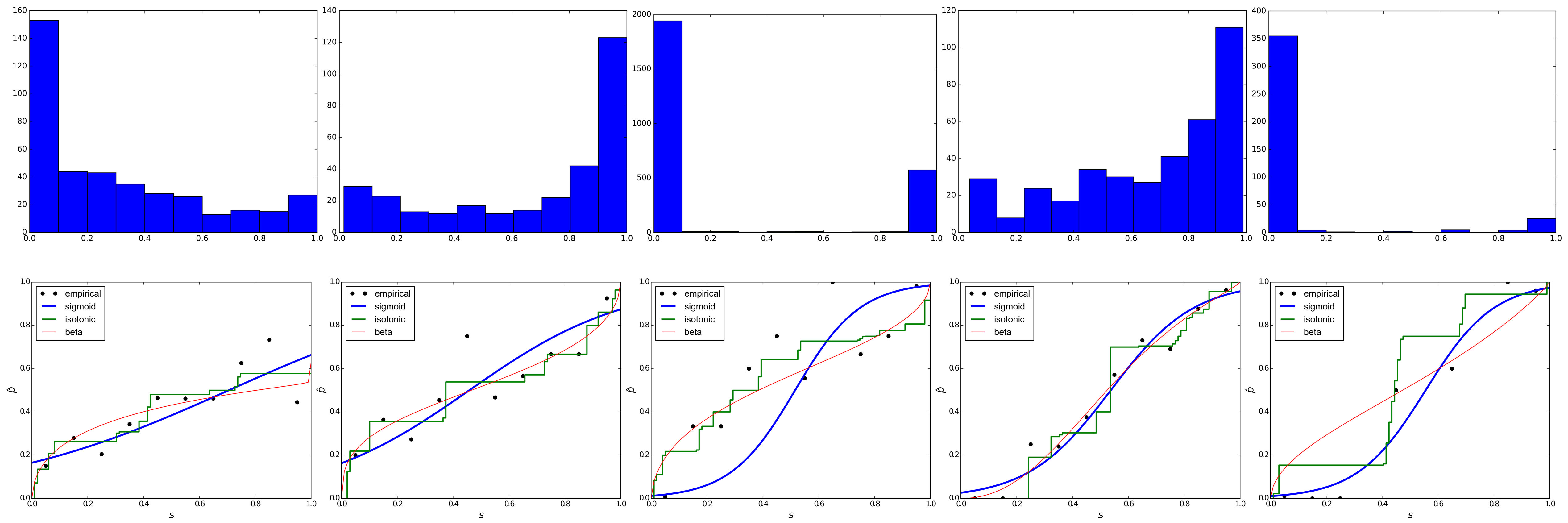


- Easily implemented by fitting logistic regression on two features $\ln(s)$ and $-\ln(1-s)$.

BETA CALIBRATION FAMILY

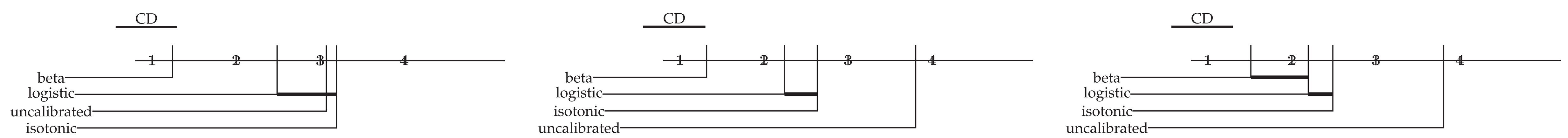


SCORE HISTOGRAMS, CALIBRATION MAPS (ADABOOST- ORIGINAL)



German Diabetes Landsat Tic-tac Vowel

EXPERIMENTS ON 41 DATASETS (LOG-LOSS)



Adaboost-Original (p=1.01e−12) Naive Bayes (p=6.88e−17) Adaboost-SAMME (p=4.73e−15)

TAKE HOME MESSAGES

Beta calibration:

- **Well-founded**: derived from beta distribution;
- **Easily-implemented**: logistic regression after log-transform;
- **Better calibrated probabilities than from logistic** in our experiments on 3 model classes.

CODE AND PACKAGES

The source code for experiments, beta calibration packages for Python and R and tutorials for both languages:

https://betacal.github.io

ACKNOWLEDGEMENTS

We were supported by the SPHERE Interdisciplinary Research Collaboration, funded by the UK Engineering and Physical Sciences Research Council under grant EP/K031910/1. TSF was financially supported by CNPq (Brazilian National Council for Scientific and Technological Development).

MULTICLASS?

Dirichlet calibration is the upcoming generalisation to multi-class classifier calibration available at https://dircal.github.io