Classifier Calibration

MaVi Research Group Meeting

Miquel Perelló Nieto

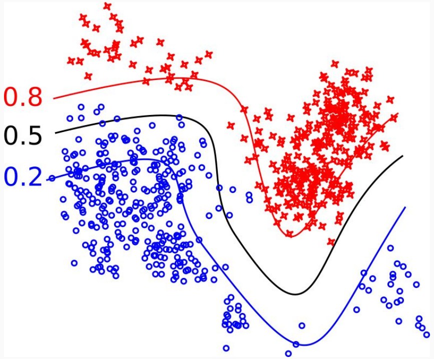
19th January 2023

Classifier Calibration

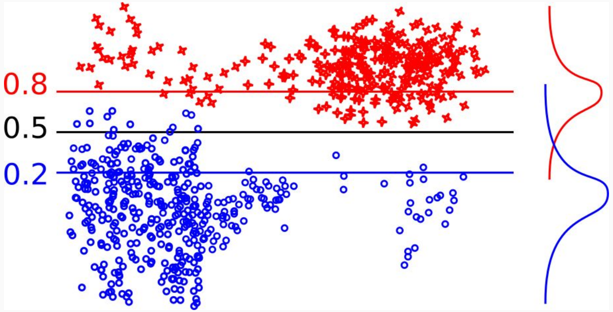
Table of content

1. What is Classification Calibration?

(a) Decision boundary and iso-hyperplanes of a probability estimator.

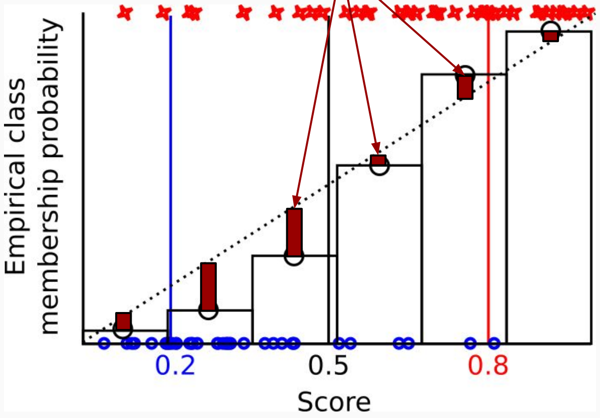


(b) Projection to an orthogonal vector



(c) Reliability diagram

Deviation to diagonal indicates miss-calibration



A probabilistic classifier $\hat{\mathbf{p}}$ is: (A) binary-calibrated if for any prediction q the proportion of positives among all instances x getting the same prediction $\hat{p}(x) = q$ are $P(Y = 2|\hat{p}(X) = q) = q$

(B) multiclass-calibrated if for any prediction vector $\mathbf{q} = (q_1, \ldots, q_k) \in \Delta_k$, the proportions of classes among all possible instances **x** getting the same prediction $\hat{\mathbf{p}}(\mathbf{x}) = \mathbf{q}$ are $P(Y = i \mid \hat{\mathbf{p}}(X) = \mathbf{q}) = q_i$ *for* $i = 1, \ldots, k.$

2. Why is calibration important? Optimal decision making

Cost Matrix	Predicted True	P. False
Actual True	-1£	100£
Actual False	50£	0£

Population proportion
50%
50%

Changes on costs

Changes on class proportions

It also allows the decision to abstain

Cost Matrix	Predicted True	P. False	**Abstain**
Actual True	-10£	90£	50£
Actual False	40£	-5£	30£

Population proportion
70%
30%

University of BRISTOL

MaVi Machine Learning and Computer Vision

3. How to measure calibration?

Confidence-calibrated if

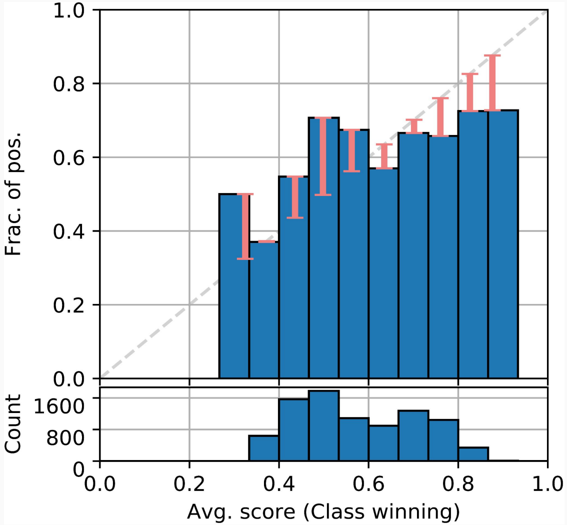
$$P\Big(Y = argmax\big(\hat{\mathbf{p}}(X)\big)\ \Big|\ max\big(\hat{\mathbf{p}}(X)\big) = c\Big) = c.$$

Empirically measured as

$$\text{confidence-ECE} = \sum_{i=1}^{m} \frac{|B_i|}{n} |y_j(B_i) - \hat{p}_j(B_i)|$$

Confidence Calibration Error

Maximum Calibration Error counterparts



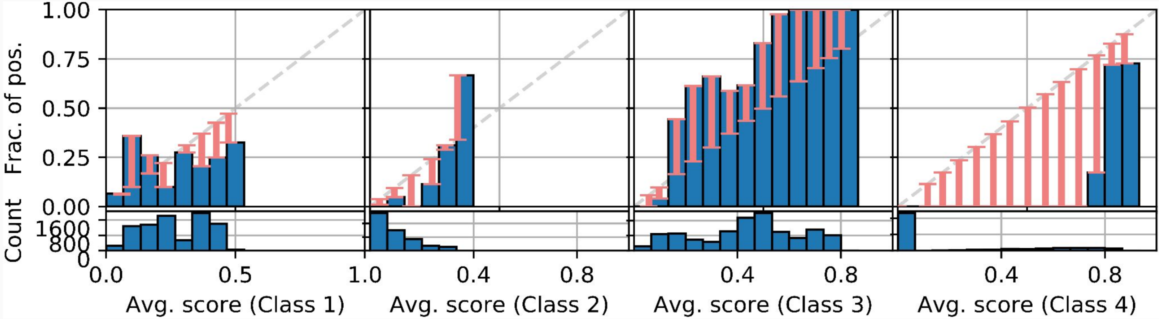
Classwise-calibrated if

$$P(Y = i \mid \hat{p}_i(X) = q_i) = q_i.$$

Empirically measured as

$$\text{classwise}-\text{ECE} = \frac{1}{k}\sum_{j=1}^{k}\sum_{i=1}^{m}\frac{|B_{i,j}|}{n}|y_j(B_{i,j}) - \hat{p}_j(B_{i,j})|$$

Calibration Error Per class

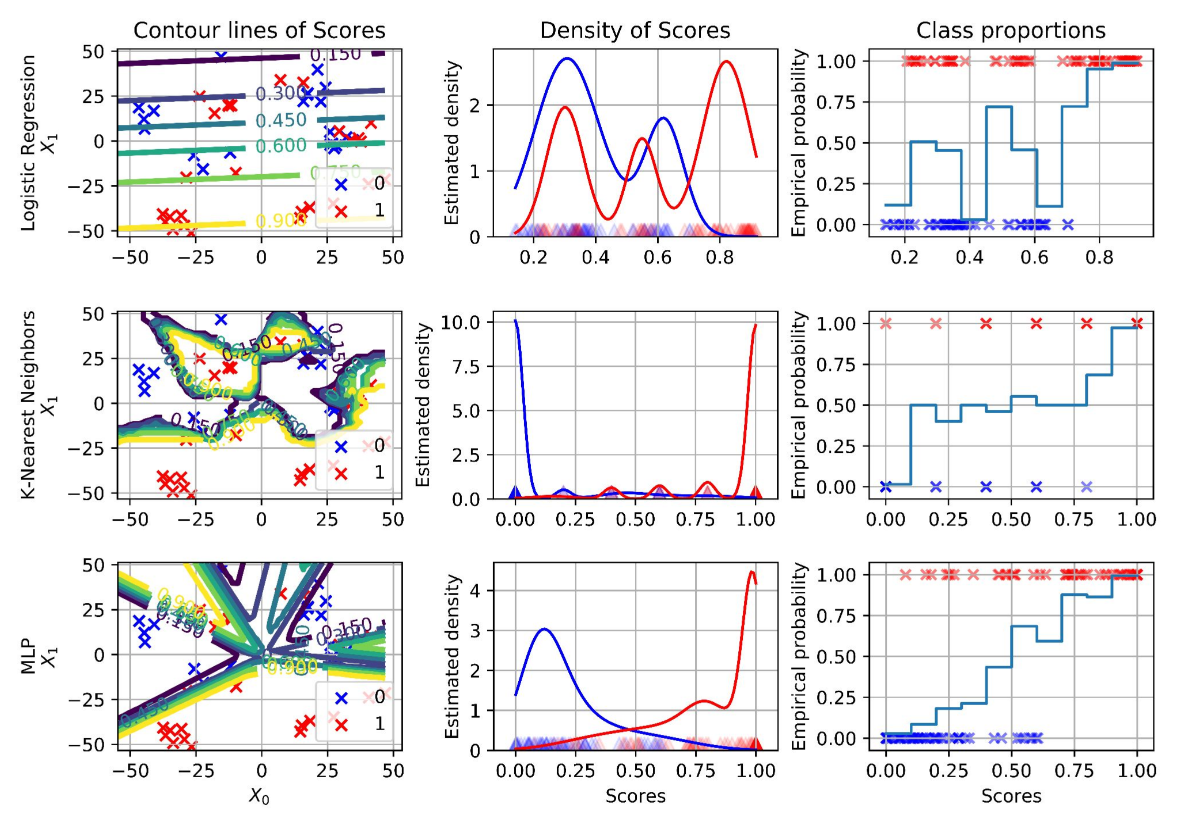


Every proper loss is minimised by the canonical calibration function (eg. log-loss and Brier score).

University of BRISTOL

MaVi Machine Learning and Computer Vision

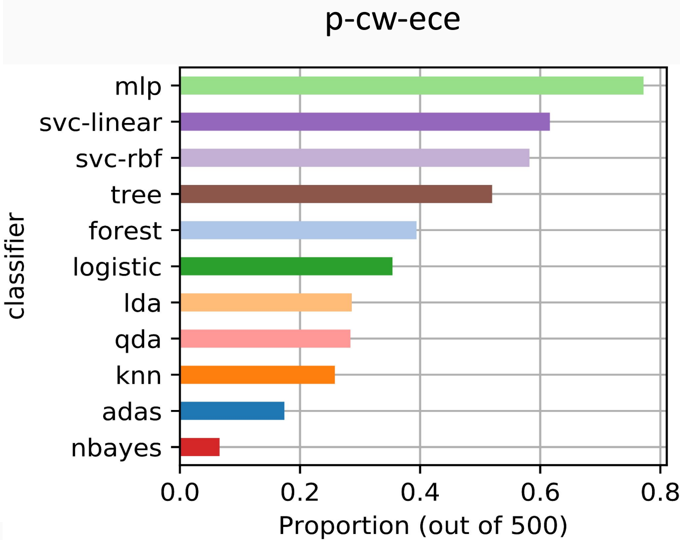
4. Are classifiers calibrated? Simple examples.



- Empirical evaluation:

- 21 datasets and 11 classifiers

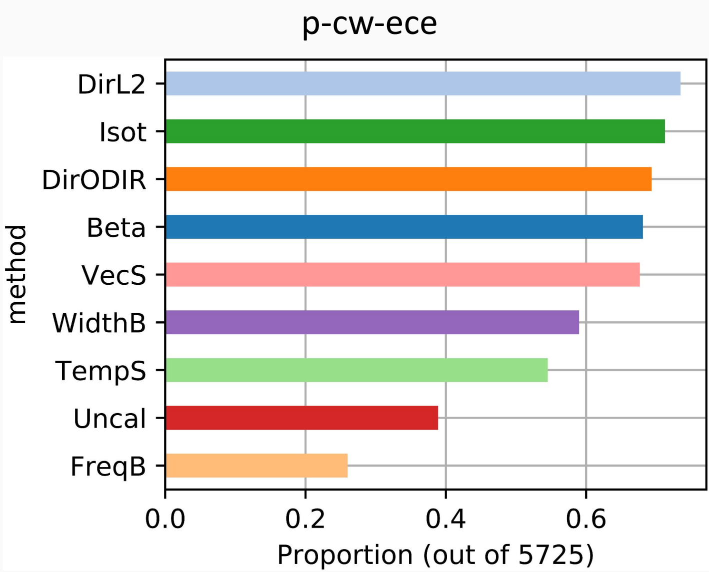
- 5 times 5-fold-crossval.



5. Calibration methods

- 21 datasets, 11 classifiers = 231 settings to compare 8 calibration methods
- 5 times 5-fold-crossval and inner 3-fold-crossval

Binary	Multiclass
• Empirical binning [1] • Platt scaling [3] • Isotonic regression [2] • Beta calibration [4] • Bayesian binning into quantiles [7]	• One-vs-rest counterparts • Temperature scaling [5] • Vector scaling [5] • Matrix scaling [5] • Dirichlet calibration [6]



University of BRISTOL

MaVi Machine Learning and Computer Vision

Classifier Calibration

Conclusion

1. Classifier calibration adjusts the probabilities output by a classifier to be more precise.
2. Optimal decision making under changing operating conditions.
3. Multiple measures with their caveats.
4. MLP, decision trees, SVMs with Platt scaling and ensembles are among the best calibrated classifiers.
5. Multiple calibration methods can still improve their estimations.

University of BRISTOL

MaVi Machine Learning and Computer Vision

References

[1] Zadrozny, B., & Elkan, C. (2001). Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers. *18th International Conference on Machine Learning (ICML'01)*, 609–616.

[2] Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. *22nd International Conference on Machine Learning (ICML'05)*, 625–632.

[3] Platt, J. (2000). Probabilities for SV Machines. In A. J. Smola, P. Bartlett, B. Schölkopf, & D. Schuurmans (Eds.), *Advances in Large-Margin Classifiers* (pp. 61--74). MIT Press.

[4] Kull, M., Silva Filho, T. D. Menezes., & Flach, P. (2017). Beyond Sigmoids: How to obtain well-calibrated probabilities from binary classifiers with beta calibration. *Electronic Journal of Statistics*, *11*(2), 5052–5080.

[5] Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017, June 14). On Calibration of Modern Neural Networks. *Thirty-Fourth International Conference on Machine Learning*.

[6] Kull, M., Perello-Nieto, M., Kängsepp, M., Filho, T. S., Song, H., & Flach, P. (2019). Beyond temperature scaling: Obtaining well-calibrated multiclass probabilities with Dirichlet calibration. In H. Wallach, H. Larochelle, A. Beygelzimer, F. Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (pp. 12316–12326). Curran Associates, Inc.

[7] Naeini, M. P., Cooper, G. F., Hauskrecht, M., Naeini, P., Cooper, G. F., & Hauskrecht, M. (2015). Obtaining Well Calibrated Probabilities Using Bayesian Binning. *29th AAAI Conference on Artificial Intelligence*, *2015*, 2901–2907

University of BRISTOL

MaVi Machine Learning and Computer Vision