CORRECTION

Correction to: Efficient feature selection using shrinkage estimators

Konstantinos Sechidis[1] · Laura Azzimonti[2] · Adam Pocock[3] · Giorgio Corani[2] · James Weatherall[4] · Gavin Brown[1]

Correction to: Machine Learning (2019) 108:1261–1286
https://doi.org/10.1007/s10994-019-05795-1

There was a mistake in the proof of the optimal shrinkage intensity for our estimator presented in Section 3.1. The main theorem still holds, and the shrinkage intensity presented in the corrected version is the optimal in the sense of minimizing the mean squared error (MSE). In this document, apart from correcting the proof for the optimal shrinkage intensity, we provide empirical verification on the correctness via simulations. The third term of Theorem 1 needs to be corrected as follows:

✉ Konstantinos Sechidis
konstantinos.sechidis@manchester.ac.uk

Laura Azzimonti
laura@idsia.ch

Adam Pocock
adam.pocock@oracle.com

Giorgio Corani
giorgio@idsia.ch

James Weatherall
James.Weatherall@astrazeneca.com

Gavin Brown
gavin.brown@manchester.ac.uk

[1] School of Computer Science, University of Manchester, Manchester, UK

[2] Istituto Dalle Molle di studi sull' Intelligenza Artificiale (IDSIA), Manno, Switzerland

[3] Oracle Labs, Burlington, MA, USA

[4] Advanced Analytics Centre, Global Medicines Development, AstraZeneca, Cambridge, UK

$$\widehat{\mathbb{E}}\big[(\hat{p}^{\mathrm{Ind}}(xy))^2\big] = \frac{1}{N^3}\bigg((N-1)(N-2)(N-3)\big(\hat{p}^{\mathrm{ML}}(x)\hat{p}^{\mathrm{ML}}(y)\big)^2$$

$$+ (N-1)(N-2)\hat{p}^{\mathrm{ML}}(x)\hat{p}^{\mathrm{ML}}(y)\big((\hat{p}^{\mathrm{ML}}(x) + \hat{p}^{\mathrm{ML}}(y) + 4\hat{p}^{\mathrm{ML}}(xy))\big)$$

$$+ (N-1)\big(2\hat{p}^{\mathrm{ML}}(xy)(\hat{p}^{\mathrm{ML}}(x) + \hat{p}^{\mathrm{ML}}(y)) + 2(\hat{p}^{\mathrm{ML}}(xy))^2$$

$$+ \hat{p}^{\mathrm{ML}}(x)\hat{p}^{\mathrm{ML}}(y)\big) + \hat{p}^{\mathrm{ML}}(xy)\bigg).$$

$$(1)$$

Parts of supplementary material's pages 4–6, where the above term is derived, need the following corrections. In page 4 the term $A(xy)$ needs to be corrected as follows:

$$A(xy) = \sum_{\substack{x',x''\in\mathcal{X} \\ x'\neq x''\neq x}} \sum_{\substack{y',y''\in\mathcal{Y} \\ y'\neq y''}} \mathbb{E}\big[N_{xy'}N_{xy''}N_{x'y}N_{x''y}\big] + 2\sum_{\substack{x'\in\mathcal{X} \\ x'\neq x}} \sum_{\substack{y',y''\in\mathcal{Y} \\ y'\neq y''\neq y}} \mathbb{E}\big[N_{xy'}N_{xy''}N_{x'y}N_{xy}\big].$$

As a consequence in page 5 the same term needs correction:

$$A(xy) = N^{(4)}\bigg[\bigg(p(x)^2 - \sum_{y'\in\mathcal{Y}} p(xy')^2\bigg)\bigg(p(y)^2 - \sum_{x'\in\mathcal{X}} p(x'y)^2\bigg)$$

$$-4\big(p(x) - p(xy)\big)p(xy)^2\big(p(y) - p(xy)\big)\bigg].$$

Finally, the first equation in page 6 needs the following correction:

$$\sum_{x',x''\in\mathcal{X}} \sum_{y',y''\in\mathcal{Y}} \mathbb{E}\big[N_{xy'}N_{xy''}N_{x'y}N_{x''y}\big] = N^{(4)}p(x)^2p(y)^2$$

$$+ N^{(3)}p(x)p(y)(p(x) + p(y) + 4p(xy))$$

$$+ N^{(2)}\big[2p(xy)(p(x) + p(y)) + 2p(xy)^2 + p(x)p(y)\big]$$

$$+ Np(xy),$$

which will result in the estimate for $\widehat{\mathbb{E}}\big[(\hat{p}^{\mathrm{Ind}}(xy))^2\big]$ presented in Eq. (1).

Apart from correcting the proof, we also provide some simulation results that validate the correctness of the optimal shrinkage intensity. To this end we followed the procedure described in the main paper's Section 3.2, to generate probabilities that lead to different types of effect size, i.e. different population values for the mutual information $I(X; Y)$. The squared error of our shrinkage estimator for the probabilities is defined as $\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\big(p(xy) - \hat{p}^{\mathrm{Ind\text{-}JS}}(xy)\big)^2$. We estimated the MSE by averaging over 1000 simulation runs. In Fig. 1 we present the results for three different effect sizes: $I(X;Y) = 0.01, 0.05$ and 0.15. In each graph we plot the MSE for all possible values of the shrinkage intensity $[0, 1]$ and we also point out the optimal intensity using the corrected value λ^* and the value we erroneously used in the previous version of the paper λ_e^*. As we see, the corrected value leads to the minimum MSE.
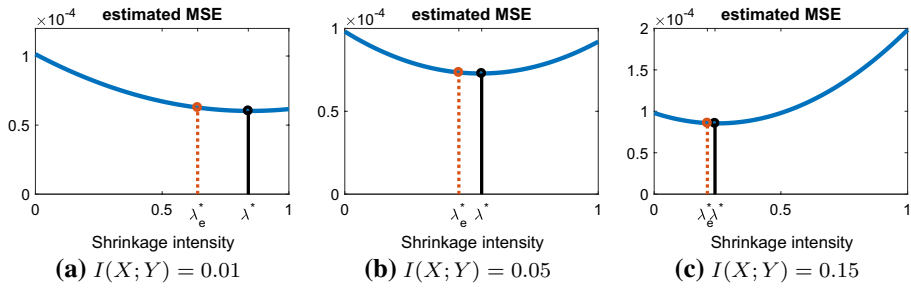
Fig. 1 Comparing the performance of our shrinkage estimator for different values of the shrinkage intensity in terms of MSE for three different effect sizes: **a** *I(X;Y)* = 0.01, **b** *I(X;Y)* = 0.05 and **c** *I(X;Y)* = 0.15. The vertical lines show the optimal shrinkage intensity presented in this correction (solid black line), and the intensity erroneously presented in the initial version of the paper (dashed red line)