Tailored Bayes: a risk modelling framework under unequal classification costs

Solon Karapanagiotis[1,2], Umberto Benedetto[3], Sach Mukherjee[4], Paul Kirk[1], Paul Newcombe[1]

[1]MRC Biostatistics Unit, University of Cambridge
[2]The Alan Turing Institute
[3]Bristol Heart Institute, University of Bristol
[4]German Center for Neurodegenerative Diseases (DZNE), Germany

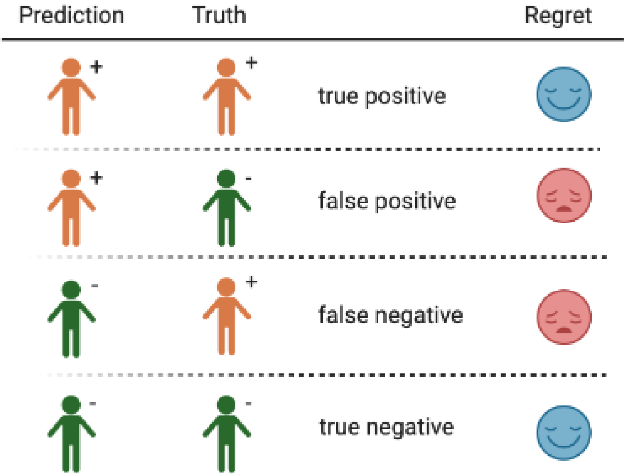
ISCB 24/08/2020

Outline

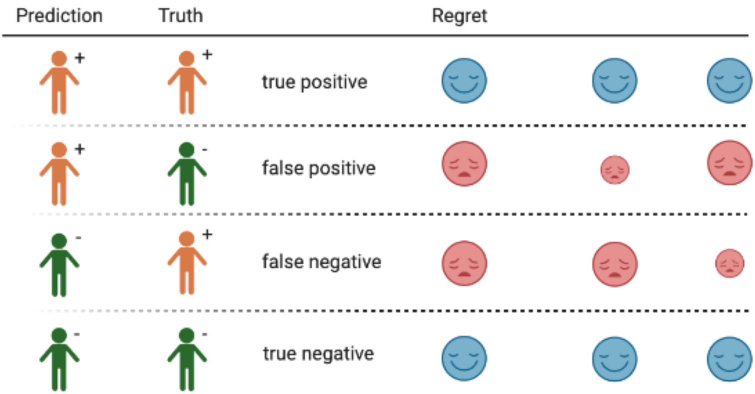
The problem

- Should patient i receive treatment?
- Traditionally, this is answered by
    1. estimate $p(y = 1 \mid \mathbf{x})$
    2. if high -> treat
    3. if low -> no treat

The problem


Prediction
Truth
Regret
true positive
false positive
false negative
true negative

The problem


Prediction
Truth
Regret
true positive
false positive
false negative
true negative

Additional examples ...

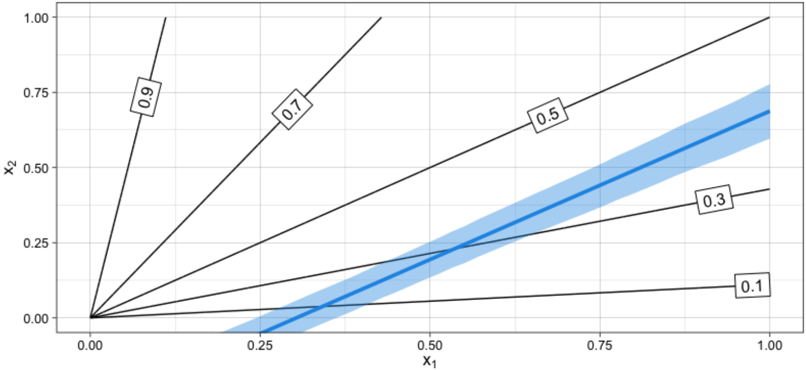
- prognosis, e.g., chemotherapy in breast cancer (later)
- diagnosis, e.g., prostate cancer
- banking/finance, e.g., loan application
- autonomous driving, e.g., misinterpreting road signs
- ...

Decision Theory (Pauker and Kassirer, 1975)

- Step 1: estimate $p(y = 1 \mid \mathbf{x})$
- Step 2: classify $p(y = 1 \mid \mathbf{x}) \geq t$
- $t = \frac{U_{TN} - U_{FP}}{U_{TN} - U_{FP} + U_{TP} - U_{FN}} = \frac{H}{H+B} = \frac{1}{1+\frac{B}{H}}$

	Truth	
Predict	U_{TP}	U_{FP}
	U_{FN}	U_{TN}

Toy example: The recipe can fail



Posterior mean boundaries for standard Bayes logistic model (blue) when targeting $t = 0.3$ (1:2.3 ratio). Shaded regions represent 90% highest predictive density (HPD) intervals. Data simulated from $p(y = 1|x_1, x_2) = \frac{x_2}{x_1+x_2}$, where $x_1, x_2 \sim \mathcal{U}[0, 1]$ and $n = 5000$.

Tailored Bayes I

Data $\{(y_i, \mathbf{x}_i) : i = 1, \ldots, n\}$. The association between y and $\mathbf{x}$ is described through the following generalized logistic loss

$$\ell(y_i, p_{w_i}) = -(p_{w_i})^{y_i}(1 - p_{w_i})^{1-y_i} \tag{1}$$

- $p_{w_i} = p(y_i = 1|\mathbf{x}_i; \beta) = (\exp\{\mathbf{x}_i^T \beta\}/1 + \exp\{\mathbf{x}_i^T \beta\})^{w_i}$
- $w_i \in [0, 1]$ are datapoint-specific weights

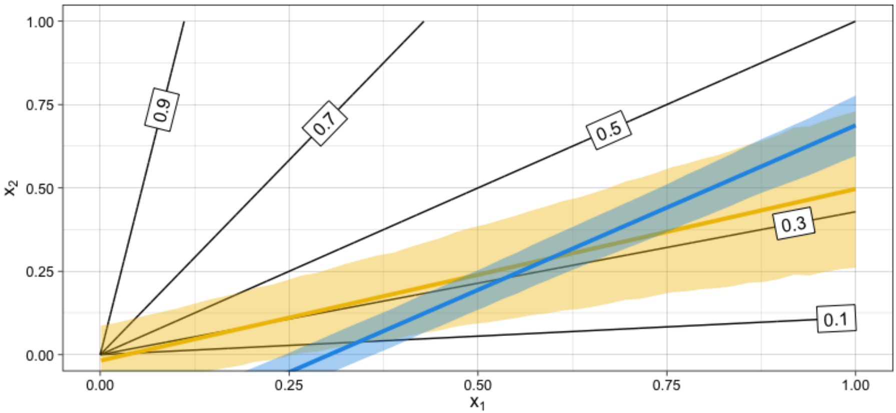
$$\beta_j \sim \mathcal{N}(0, 100^2), (j = 1, \ldots, d = parameters)$$

Tailored Bayes (Hand and Vinciotti, 2003) I

$$w_i = \exp\Big\{ -\lambda(p_u(\mathbf{x}_i) - t)^2 \Big\}$$

- $p_u(\mathbf{x}_i) = p(y_i = 1|\mathbf{x}_i)$
- t is the target threshold. It captures how we weigh the relative harms of false-positive and false-negative results
- $\lambda \geq 0$ is a tuning parameter. For $\lambda = 0$ we recover the standard logistic regression model
- In practice, $p_u(\mathbf{x}_i)$ needs to be estimated

Toy example

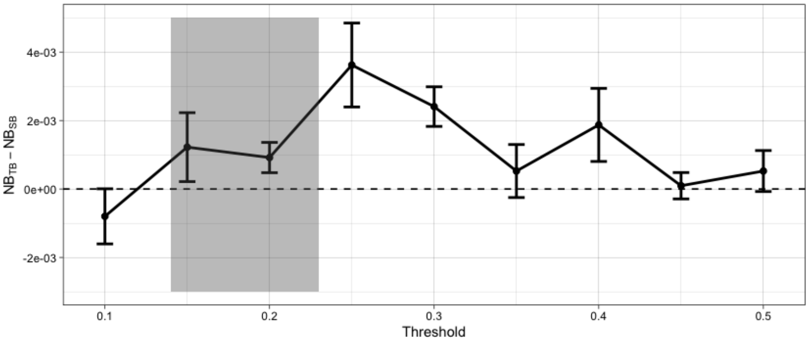


Posterior mean boundaries for standard Bayes (blue) and Tailored Bayes (yellow) when targeting $t = 0.3$.

Breast cancer prognostication

- Data
  - Train: 4718 invasive breast cancer.
  - Test: 3810 subjects from an independent cohort.
- Outcome 10-year breast cancer–specific mortality.
- The covariates are
  - age at diagnosis (years)
  - tumor grade (I, II, III)
  - number of positive lymph nodes
  - presentation (screening vs. clinical)
  - type of adjuvant therapy (chemotherapy, endocrine therapy, or both).

Breast cancer prognostication



Difference in Net Benefit (NB) for various t values. A positive difference means Tailored Bayes (TB) outperforms standard Bayes (SB). $NB = \frac{TP_t}{n} - \frac{FP_t}{n}\frac{t}{1-t}$ (Vickers and Elkin, 2006). The units on the y axis may be interpreted as the difference in benefit associated with one patient who would die without treatment and who receives therapy.

Contributions I

- A key aim of precision medicine is to tailor clinical management.
- Here we present a framework to tailor model development incorporating misclassification costs into Bayesian modelling.
- Attractive features that make it flexible, easy-to-use, and widely applicable:
  - Relies solely on calculating, w_i - robust to different choices, $w_i = \exp\{-h(p_u(\mathbf{x}_i), t)\}$.
  - Bayesian: hierarchical modelling and incorporation of external information.
- Generic:
  1. Implemented in any learning framework (not necessarily Bayesian).

Contributions II

2. Not restricted to logistic loss. The scheme can be used to adapt any loss.

- Current work: Implications for variable selection.

Key References

Hand, D. J. and Vinciotti, V. (2003). Local versus global models for classification problems: Fitting models where it matters. *The American Statistician*, 57(2):124–131.

Pauker, S. G. and Kassirer, J. P. (1975). Therapeutic decision making: a cost-benefit analysis. *New England Journal of Medicine*, 293(5):229–234.

Vickers, A. J. and Elkin, E. B. (2006). Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making*, 26(6):565–574.

'TailoredBayes' -
https://github.com/solonkarapa/TailorBayes

solon.karapanagiotis@mrc-bsu.cam.ac.uk