BAYESIAN OUTCOME WEIGHTED LEARNING

PRÉ-JOURNÉE STATISTIQUES DU SUD 18/06/24

SOPHIA YAZZOURH, INSTITUT DE MATHÉMATIQUES DE TOULOUSE

NIKKI L.B. FREEMAN, DEPARTMENT OF BIOSTATISTICS AND BIOINFORMATIC, DUKE CLINICAL RESEARCH INSTITUTE

OUTLINE

- Contextualization : precision medicine & individualized treatment regimes
- Method : Bayesian Outcome Weighted Learning
- Results : classification & uncertainty quantification

Precision medicine : « *The right treatment for the right patient (at the right time)* » [1]

- Develop models for personalized decision-making (policy) :
  - Input: patient's unique characteristics
  - Output: a treatment recommendation

- Application example, determine which of two treatments is more suitable for losing weight according to your proper features?
  - patients features : gender, age, parentBMI, baselineBMI
  - patient response : reduction in BMI

Determine a policy : $\pi(A, S) = \mathbb{P}[A \in \{-1,1\} \,|\, S = \{s_0, s_1, \ldots, s_p\}]$

Tailored to align with a specified objective : $R \in \mathbb{R}$

One way to solve this problem with machine learning :

- think of it as a two-class classification problem
- with treatment as label
- and a weighting of individuals by response and propensity of treatments administered.

OWL

[2] OWL method determines an optimal ITR, π^*, by formulating the policy such that : $\pi^* \in \text{argmax}_{\pi} \mathbb{E}[\frac{I(A = \pi(X))}{A\rho + (1 - A)/2} R]$

[3] showed that maximizing the exception part in OWL is equivalent to weighted classification problems where we minimizes the objective function :

$$Q_n^{\text{OWL}}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{R_i}{A_i\rho + (1 - = A_i)/2} (1 - A_i h(S_i, \boldsymbol{\beta}))_+$$

where $(z)_+ = max(z,0)$ denotes the hinge loss function and $h(\cdot)$ is the ITR parameterized by $\boldsymbol{\beta}$.

[3] showed that maximizing the exception part in OWL is equivalent to weighted classification problems where we minimizes the objective function :

$$Q_n^{\text{OWL}}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{A_i\rho + (1 - = A_i)/2}(1 - A_i h(S_i, \boldsymbol{\beta}))_+$$

where $(z)_+ = max(z,0)$ denotes the hinge loss function and $h(\cdot)$ is the ITR parameterized by $\boldsymbol{\beta}$.

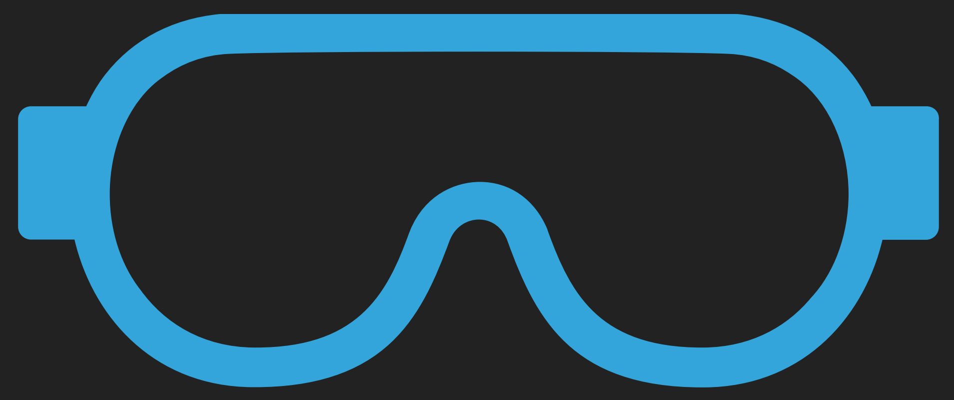
[4] introduced a penalized variant of OWL that minimizes the objective function:

$$Q_n^{\text{POWL}}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{A_i\rho + (1 - = A_i)/2}(1 - A_i h(S_i, \boldsymbol{\beta}))_+ + \sum_{j=1}^{p} p_\lambda(|\beta_j|)$$

where $p_\lambda(\beta)$ is a penalty function and λ is a tuning parameter.

BAYESIAN OWL

- It is possible to cast SVM into Bayesian framework [4]
- Why coming back to a statical method?
  - Because Bayesian framework is able to capture and model uncertainty
  - **Uncertainty quantification is a power tool in treatment recommandation for medical experts**



BAYESIAN FRAMEWORK

- Parametric estimation of β from observed data
- Combines prior information about the parameters with observed data to produce a posterior distribution of the parameters
  - Prior distribution: reflects the initial knowledge or beliefs about the parameter
  - Likelihood: probability of observing the data given the parameter
  - Posterior distribution: Pior + Data $\longrightarrow$ Posterior

BAYESIAN OWL

- Pseudo-posterior distribution :

$$
\begin{aligned}
p(x \mid a_i, \nu, \alpha) &\propto \exp(-Q_n(\boldsymbol{\beta}, \nu, \alpha)) \\
&\propto \exp\left\{ \sum_{n=1}^{n} \frac{r_i}{a_i\rho + (1-a_i)/2} (1 - a_i h(s_i, \beta))_+ \right\} \prod_{j=1}^{p} p(\beta_j \mid \mu_0, \sigma_0^2) \\
&\propto C(\nu, \alpha) L(a \mid \beta) p(\beta \mid \mu_0, \sigma_0^2)
\end{aligned}
$$

- Pseudo-likelihood :

$$
\begin{aligned}
L_i(a_i \mid r_i, s_i, \beta) &= \exp\left\{ -2\frac{r_i}{a_i\rho + (1-a_i)/2} \max(1 - a_i s_i^T \beta, 0) \right\} \\
&= I(a_i = 1) \int_0^\infty \frac{1}{\sqrt{2\pi\lambda_i}} \exp\left\{ -\frac{1}{2\lambda_i} \left(\frac{r_i}{\rho} + \lambda_i - \frac{r_i}{\rho} a_i s_i^T \beta \right)^2 \right\} d\lambda_i \\
&\quad + I(a_i = -1) \int_0^\infty \frac{1}{\sqrt{2\pi\lambda_i}} \exp\left\{ -\frac{1}{2\lambda_i} \left(\frac{r_i}{1-\rho} + \lambda_i - \frac{r_i}{1-\rho} a_i s_i^T \beta \right)^2 \right\} d\lambda_i
\end{aligned}
$$

BAYESIAN OWL

Prior distribution for β

- **Normal distribution prior**

- **Exponential power prior distribution** :

$$p(\beta_j|\nu,\alpha=1)=\int_0^\infty \phi(\beta_j|0,\nu^2\omega_j\sigma_j^2)\frac{1}{2}e^{-\frac{\omega_j}{2}}d\omega_j$$

where $p(\omega_j|\alpha)\propto\omega_j^{-\frac{3}{2}}St^{+}_{\alpha/2}(\omega_j^{-1})$ and $St^{+}_{\alpha/2}$ is the density function of a positive stable random variable of index $\alpha/2$. In particular, when $\alpha=1, p(\omega_j|\alpha)\sim Exponential(2)$.

- **Spike-and-slab prior distribution** :

$$p(\beta_j|\gamma_j,\nu^2)=\gamma_jN(0,\nu^2\sigma_j^2)+(1-\gamma_j)\delta_0(\beta_j)$$

where $\delta_0(\cdot)$is the Dirac measure (point mass at 0). The prior on γ_j is given by $p(\gamma_j|\pi)=\pi^{\gamma_j}(1-\pi)^{1-\gamma_j}$.

BAYESIAN OWL

Box 2. Gibbs sampling algorithm for exponential power distribution prior on $\boldsymbol{\beta}$

Initialize $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$.
Step 1: Draw $\boldsymbol{\beta}^{(g+1)}|\nu, \boldsymbol{\Lambda}^{(g)}, \boldsymbol{\Omega}^{(g)}, \mathbf{r}, \mathbf{a}, \mathbf{x} \sim \mathcal{N}(B_2^{(g)} b_2^{(g)}, B_2^{(g)})$.
Step 2: Draw $\boldsymbol{\lambda}^{-1(g+1)}|\boldsymbol{\beta}^{(g+1)}, \mathbf{r}, \mathbf{a}, \mathbf{x}$ where

$$\lambda_i^{(g+1)}|\boldsymbol{\beta}^{(g+1)}, \nu, r_i, x_i \sim \mathbb{1}(a_i = 1)\mathcal{GIG}\left(\frac{1}{2}, 1, \left(\frac{r_i}{\rho}\right)^2 (1 - a_i \mathbf{x}_i^\top \boldsymbol{\beta}^{(g+1)})^2\right) + \mathbb{1}(a_i = -1)\mathcal{GIG}\left(\frac{1}{2}, 1, \left(\frac{r_i}{1-\rho}\right)^2 (1 - a_i \mathbf{x}_i^\top \boldsymbol{\beta}^{(g+1)})^2\right).$$

Step 3: Draw $\omega_j^{-1(g+1)}|\beta_j^{(g+1)}, \nu \sim \mathcal{IG}(\nu\sigma_j|\beta_j|^{-1}, 1)$
Repeat Steps 1, 2, and 3 until the chains converge.

SIMULATIONS

- **Patient features:** $X_1, \ldots, X_{10} \sim U([-1,1])$
- **Treatment:** $A \sim U([-1,1])$ independently of the prognostic variables with $\mathbb{P}(A = 1) = 1/2$
- **Patient response:** $R \sim N(1 + 2X_1 + X_2 + 0.5X_3 + (X_1 + X_2)A, 1)$
- **True optimale value:** $I(X_1 + X_2 > 0)$

SIMULATIONS RESULTS

n	OWL	Bayesian OWL Normal Prior	Bayesian OWL Exponential Power Prior	Bayesian OWL Spike and Slab
100	0.24	0.38	0.38	0.39
200	0.18	0.34	0.34	0.34
400	0.13	0.29	0.29	0.30
800	0.10	0.24	0.24	0.26

Table 1: Misclassification rates for different methods and sample sizes for scenario 1.

UNCERTAINTY QUANTIFICATION WITH BAYESIAN OWL EXPONENTIAL PRIOR

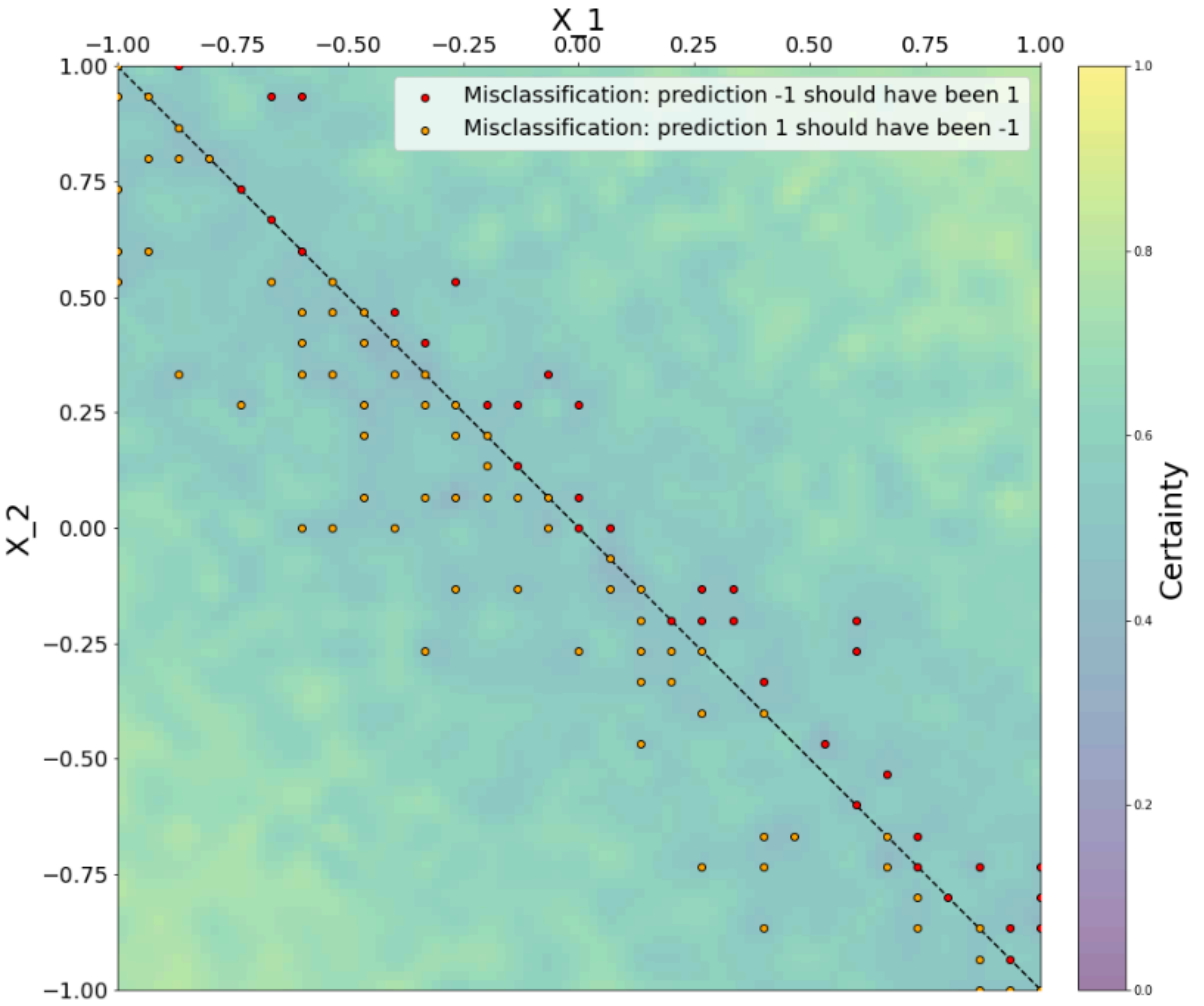


Figure 2: Heatmap of uncertainty quantification

CONCLUSION AND PERSPECTIVES

- We have introduced Bayesian formulation for OWL and demonstrated an approach with uncertainty quantification
  arXiv:2406.11573
- But only examined linear decision rules
- without variable selection

REFERENCES

- [1] Michael R Kosorok and Eric B Laber. Precision medicine. Annu Rev Stat Appl, 6:263–286, March 2019.
- [2] Yingqi Zhao, Donglin Zeng, A John Rush, and Michael R Kosorok. Estimating individualized treatment rules using outcome weighted learning. J. Am. Stat. Assoc., 107(449):1106–1118, September 2012.
- [3] Nicholas G Polson and Steven L Scott. Data augmentation for support vector machines. ba, 6(1):1–23, March 2011.
- [4] Rui Song, Michael Kosorok, Donglin Zeng, Yingqi Zhao, Eric Laber, and Ming Yuan. On sparse representation for optimal individualized treatment selection with penalized outcome weighted learning. Stat, 4(1):59–68, 2015.

CONCLUSION AND PERSPECTIVES

- We have introduced Bayesian formulation for OWL and demonstrated an approach with uncertainty quantification
  arXiv:2406.11573
- But only examined linear decision rules
- Variable selection

THANK YOU