basque center for applied mathematics

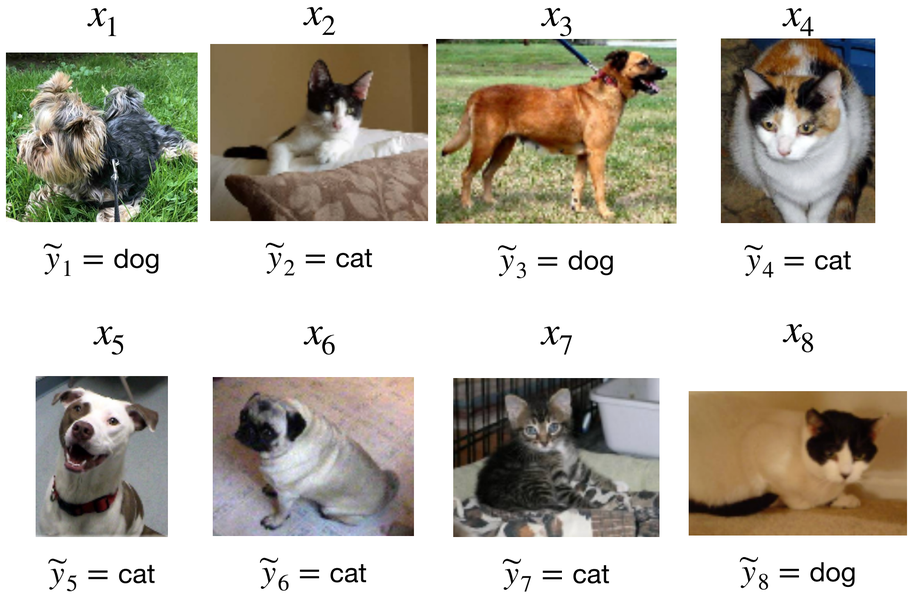
UNIVERSITY OF MICHIGAN

NEURAL INFORMATION PROCESSING SYSTEMS

Estimation of the Label-Noise Transition Matrix with Performance Guarantees via Selective Classification

Xabier de Juan Santiago Mazuelas Yilun Zhu Clayton Scott

Goal: Estimate the label-noise transition matrix given $\{(x_k, \widetilde{y}_k)\}_{k=1}^{n}$ noisy i.i.d. data

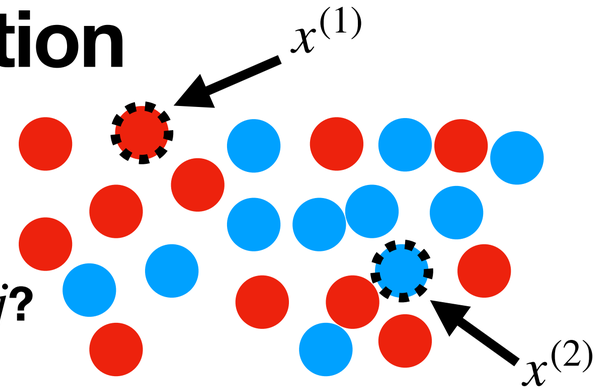


$$\longrightarrow \widehat{\mathbf{T}} \approx \mathbf{T} = \begin{bmatrix} \mathbb{P}(\widetilde{Y} = 1 \mid Y = 1) & \cdots & \mathbb{P}(\widetilde{Y} = 1 \mid Y = |\mathcal{Y}|) \\ \vdots & \ddots & \vdots \\ \mathbb{P}(\widetilde{Y} = |\mathcal{Y}| \mid Y = 1) & \cdots & \mathbb{P}(\widetilde{Y} = |\mathcal{Y}| \mid Y = |\mathcal{Y}|) \end{bmatrix}$$

Prior work: Anchor-based transition matrix estimation

For $x^{(j)}$ an anchor-point for class j (i.e., $\mathbb{P}(Y = j \mid X = x^{(j)}) = 1$)

$$T_{i,j} = \mathbb{P}(\widetilde{Y} = i \mid X = x^{(j)})$$



Anchor-based transition matrix estimation

$$\widehat{T}_{i,j}^{\text{anchor}} := \widehat{\mathbb{P}}(\widetilde{Y} = i \mid X = x^{(j)})$$

How to find anchor-point for class j?

$$x^{(j)} \in \arg\max_{x \in \mathcal{X}} \mathbb{P}(\widetilde{Y} = j \mid X = x)$$

Problem: Estimating $\mathbb{P}(\widetilde{Y} \mid X = x)$ is provably hard (suffers curse of dimensionality)

Existing knowledge gaps

Existing estimators of $\mathbf{T}$:

- Require pointwise class-posterior estimation
- Do not come with finite sample performance guarantees

Our contributions

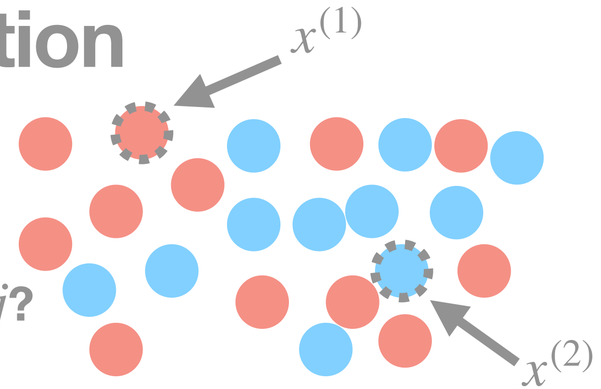
Methodology to estimate $\mathbf{T}$:

- Bypass pointwise class-posterior estimation
- Provides finite-sample performance guarantees

Prior work: Anchor-based transition matrix estimation

For $x^{(j)}$ an anchor-point for class j (i.e., $\mathbb{P}(Y = j \mid X = x^{(j)}) = 1$)

$$T_{i,j} = \mathbb{P}(\widetilde{Y} = i \mid X = x^{(j)})$$



Anchor-based transition matrix estimation

$$\widehat{T}^{\text{anchor}}_{i,j} := \widehat{\mathbb{P}}(\widetilde{Y} = i \mid X = x^{(j)})$$

How to find anchor-point for class j?

$$x^{(j)} \in \arg\max_{x \in \mathcal{X}} \mathbb{P}(\widetilde{Y} = j \mid X = x)$$

Problem: Estimating $\mathbb{P}(\widetilde{Y} \mid X = x)$ is provably hard (suffers curse of dimensionality)

Ours: Transition matrix estimation via one-sided selective classification

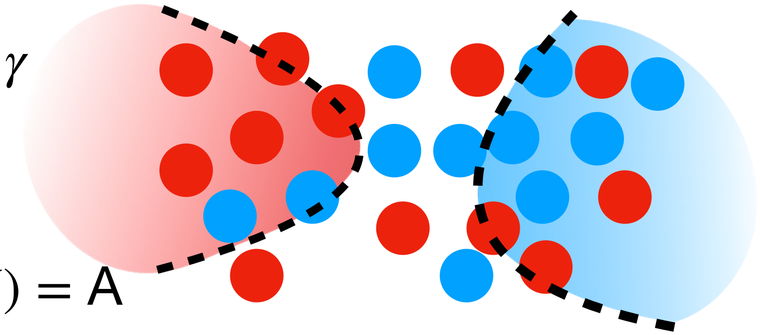
$$\widehat{T}_{i,j} := \widehat{\mathbb{P}}(\widetilde{Y} = i \mid h^{(j)}(X) = \mathsf{A})$$

one-sided selection function for class j

$$h^{(j)} : \mathcal{X} \to \{\mathsf{A}, \mathsf{R}\}$$

How to obtain selection function for class j ?

$$h^{(j)} \in \arg \min_{h : \mathcal{X} \to \{\mathsf{A},\mathsf{R}\}} \mathbb{P}(\widetilde{Y} \neq j \mid h(X) = \mathsf{A}) \quad \text{s.t.} \quad \mathbb{P}(h(X) = \mathsf{A}) \geq \gamma$$



Finite sample performance guarantees: With high probability

$$|T_{i,j} - \widehat{T}_{i,j}| \lesssim R^{(j)}_{\gamma} + \frac{1}{\sqrt{\gamma n}}$$

smallest one-sided error with no noise

Choose your favourite binary classifier, turn it into a transition matrix estimator

One-sided selective classification

$$h^{(j)} \leftarrow \min_{h:\mathcal{X}\to\{\mathsf{A},\mathsf{R}\}} \mathbb{P}(\widetilde{Y} \neq j \mid h(X) = \mathsf{A})$$

$$\text{s.t.} \quad \mathbb{P}(h(X) = \mathsf{A}) \geq \gamma$$

$\approx$

Binary classification

$$f^{(j)} \leftarrow \min_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \ell^{(j)}(\tilde{y}_k, f(x_k))$$

$$\text{s.t.} \quad \widehat{\mathbb{P}}\,(f(X) = 1) \geq \gamma$$

$\mathcal{F}$ = {binary classifiers}

$\ell^{(j)}$ = loss that penalizes misclassification of $\widetilde{Y} = j$ (e.g., cost-sensitive losses)

Theorem: Define $h^{(j)}(x) = \mathsf{A}$ if $f^{(j)}(x) = 1$ and $h^{(j)}(x) = \mathsf{R}$ otherwise. Then, for $\widehat{T}_{i,j} := \widehat{\mathbb{P}}\,(\widetilde{Y} = i \mid h^{(j)}(X) = \mathsf{A})$

$$|T_{i,j} - \widehat{T}_{i,j}| \lesssim R_\gamma^{(j)} + \frac{1}{\gamma}\left(\mathcal{A}_{\ell^{(j)}}(\mathcal{F}) + \sqrt{\frac{\mathrm{VC}(\mathcal{F})\log(n)}{n}}\right) + \frac{1}{\sqrt{\gamma n}}$$

approximation error of the hypothesis class $\mathcal{F}$ for the loss $\ell^{(j)}$:

$\mathcal{A}_{\ell^{(j)}}(\mathcal{F})$ = “Bayes risk”-“best in class risk

Proposed methods perform comparably to an oracle ✅

Anchor-based method's error stagnates with n 🚨

