SLIDS Homework 4

Dave Zachariah

March 3, 2023

- Solution proposals are individual.
- Each solution must be reproducible by your peer. Code should be added to the appendix.
- The solution to each subproblem yield 0 to 2 points. Students are expected to attempt each problem.

1

Let us return to the ship localization problem from HW1 when the sampling distribution is corrupted:

$$\widetilde{p}(\mathbf{z}) = (1-\varepsilon)p(\mathbf{z}) + \varepsilon q(\mathbf{z})$$

We will assume that we can specify the quality of the data using an upper bound $\widetilde{\varepsilon} \geq \varepsilon$. We will localize the ship using the square-error loss $\ell_\theta(\mathbf{z}) = \|\mathbf{z} - \boldsymbol{\theta}\|^2$

Using $n = 100$, compare $\widehat{\boldsymbol{\theta}}_n$ using ERM and RRM (lecture 3) with $\boldsymbol{\theta}_\circ$ as in HW1.

For RRM, we set the bound $\widetilde{\varepsilon}$. To solve the robust risk minimization problem, use the following numerical search method (aka. coordinate descent):

- Outer problem: Fix $\widehat{\boldsymbol{\omega}}$, solve

$$\widehat{\boldsymbol{\theta}} := \operatorname*{arg\,min}_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}; \widetilde{p}_{\widehat{\omega}})$$

- Inner problem: Fix $\widehat{\boldsymbol{\theta}}$, solve

$$\boldsymbol{\omega} := \operatorname*{arg\,min}_{\boldsymbol{\omega} \geq \mathbf{0}, \mathbf{1}^\top \boldsymbol{\omega} = 1, \mathbb{H}(\boldsymbol{\omega}) \geq \ln[(1-\widetilde{\varepsilon})n]} L(\widehat{\boldsymbol{\theta}}; \widetilde{p}_\omega)$$

Repeat until convergence and start with uniform weights $\widehat{\boldsymbol{\omega}} = n^{-1}\mathbf{1}$.

Tip: The inner problem cannot be solved in closed form, but since it is convex you can use a numerical package such as `cvx`, `cvxopt` or `cvxpy` to solve it.

Study what happens when you vary the tolerance bound $\widetilde{\varepsilon} \in \{0.10, 0.20, 0.30, 0.40\}$ (that is tolerance up to 40% outliers).

The sample $\mathbf{z}^n$ from $\widetilde{p}(\mathbf{z})$ can be generated as follows:

Let $c \sim \mathrm{Ber}(c; \varepsilon)$ be an unobserved sample corruption indicator, where $\varepsilon =$ 20%. Given c, draw $\mathbf{z}$ from

$$\begin{cases} p(\mathbf{z}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) & c = 0 \\ q(\mathbf{z}) = t_\nu(\boldsymbol{\mu}, \mathbf{C}) & c = 1 \end{cases} \tag{1}$$

where

$$\boldsymbol{\mu} = \begin{bmatrix} 200 \\ 200 \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 400 & 50 \\ 50 & 400 \end{bmatrix} \quad \mathbf{C} = 2\frac{\nu}{\nu - 2}\boldsymbol{\Sigma}$$

and $\nu = 2.5$.

2

We first familiarize ourselves with the structural causal description of data-generating processes.

Generate $n = 1000$ samples $(x, a, y) \sim p(x, a, y)$ from the processes below, where w is a binary variable. Overlay two scatter plots of the the training data (a, y)

- one plot using all n samples
- the other plot using a subset of n' samples when $w = 1$

a) Mediating structure

$$\mathcal{S} : \begin{cases} a & \sim \mathcal{N}(a; 0, 1) \\ x & \sim \mathrm{Ber}\left(x; s(10(a - \frac{1}{2}))\right) \\ y & \sim \mathcal{N}(y; 2x, 1) \end{cases}$$

b) Confounding structure

$$\mathcal{S} : \begin{cases} x & \sim \mathrm{Ber}\left(x; 0.80\right) \\ a & \sim \mathcal{N}(a; 2x, (x + 1)) \\ y & \sim \mathcal{N}(y; -a + x, 1) \end{cases}$$

c) Colliding structure

$$\mathcal{S} : \begin{cases} a & \sim \mathcal{N}(a; 0, 1) \\ y & \sim \mathcal{N}(y; 0, 1) \\ x & \sim \mathrm{Ber}\left(w; s(-a + 2y - 1)\right) \end{cases}$$

where

$$s(z) = (1 + \exp(-z))^{-1}$$

is the sigmoid function.

Compare our visual results with the theoretical dependencies discussed in Lecture 4.

3

Suppose we want to construct a policy $\pi(\mathbf{x})$ to predict an outcome y given given covariates $\mathbf{x} = (x_1, \dots, x_{15})$. The more covariates we use, the more training data we need to reliably learn a good predictor.

However, if we know or can assert the structure $\mathcal{S}$ of the data generating process, it is possible to find a small subset $\mathbf{x}_s \subset \mathbf{x}$ that contains all association to predict y. That is,

$$p(y|\mathbf{x}) = p(y|\mathbf{x}_s)$$

or, equivalently, $y \perp\!\!\!\perp \mathbf{x}_s^c | \mathbf{x}_s$, where $\mathbf{x}_s^c$ is the complement of $\mathbf{x}_s$.

Show that $\mathbf{x}_s = (x_5, x_6, x_9, x_{10}, x_{11}, x_{12}) \subset \mathbf{x}$ is indeed such a subset for a structure given in Figure 1.

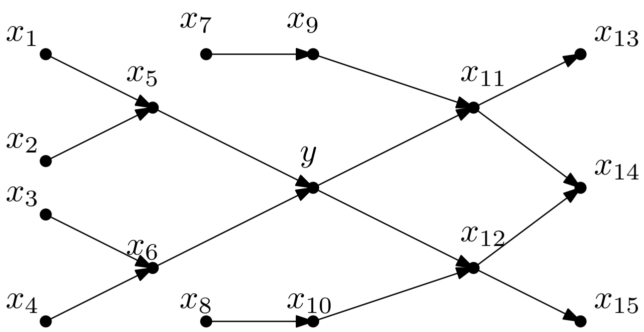

Figure 1: $\mathcal{G}(\mathcal{S})$

4

Suppose we want to study the causal association between a and y by observing a data generating process with structure $\mathcal{S}$.

For the three processes with structures illustrated in Figure 2 show that conditioning on the set of variables $\mathbf{c}$ blocks the noncausal association between a and y:

a) $\mathcal{S}_1 : \mathbf{c} = \varnothing$

b) $\mathcal{S}_2 : \mathbf{c} = w_1$

c) $\mathcal{S}_3 : \mathbf{c} = \varnothing$

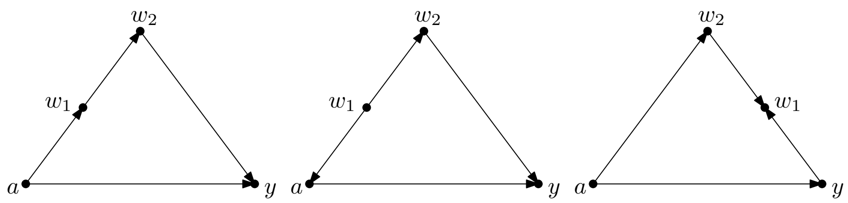

Figure 2: $\mathcal{G}(\mathcal{S}_1)$, $\mathcal{G}(\mathcal{S}_2)$ and $\mathcal{G}(\mathcal{S}_3)$