STA 561: Homeworks 9 (Due April 26 at midnight)

Reminder: work together! Share ideas, brainstorm, explain/verify your answers but write up your own work. Your homework should be submitted as pdf file generated using either latex or an python notebook.

1. (Failure.) Suppose we have data $(\boldsymbol{X}, A, Y) \sim P$ where $\boldsymbol{X} \in \mathbb{R}^p$ are decision contexts, $A \in \{-1, 1\}$ is the assigned decision, and $Y \in \mathbb{R}$ is the observed utility (outcome). You can assume the standard causality assumptions hold. Let $Q(\boldsymbol{x}, a) = \mathbb{E}(Y|\boldsymbol{X} = \boldsymbol{x}, A = a)$ and suppose that we posit a linear model of the form

$$Q(\boldsymbol{x}, a; \boldsymbol{\beta}) = \boldsymbol{x}^\intercal \boldsymbol{\beta}_0 + a\boldsymbol{x}^\intercal \boldsymbol{\beta}_1,$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_0^\intercal, \boldsymbol{\beta}_1^\intercal)^\intercal$ is a vector of unknown coefficients (we are NOT assuming this model is correctly specified). Construct a generate model P such that the following hold (simultaneously):

   i. $\pi^{\text{opt}}(\boldsymbol{x}) = \text{sign}(\boldsymbol{x}^\intercal \theta^*)$ for some $\theta^* \in \mathbb{R}^p$, i.e., a linear decision rule is optimal;

   ii. define $\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}} P \left\{Y - Q(\boldsymbol{X}, A; \boldsymbol{\beta})\right\}^2$ and $\pi^*(\boldsymbol{x}) = \arg\max_a Q(\boldsymbol{x}, a; \boldsymbol{\beta}^*) = \text{sign}(\boldsymbol{x}^\intercal \boldsymbol{\beta}_1^*)$; however $V(\pi^*) < V(\pi^{\text{opt}}) - 1$.

   That is, even though the true optimal rule is linear and we estimate a linear decision rule using Q-learning, the resulting decision rule is suboptimal.