Problemset (1), Foundations of Machine learning, HT 2022

Maximilian Kasy

In this problemset you are asked to implement some simulations and estimators in R. Please make sure that your solutions satisfy the following conditions:

- The code has to run from start to end on the grader's machine, producing all the output.
- Output and discussion of findings have to be integrated in a report generated in R-Markdown.
- Figures and tables have to be clearly labeled and interpretable.
- The findings need to be discussed in the context of the theoretical results that we derived in class.

1. In this problem, you will calculate the risk functions (MSE) of various estimators in the normal means setting, using simulations. To do so,
   (a) Pick some vector $\boldsymbol{\theta}_1$ of length 1 (it does not matter which one),
   (b) take $\boldsymbol{\theta} = r \cdot \boldsymbol{\theta}_1$ for $r \in [0, 6]$,
   (c) repeatedly (say, 10,000 times) draw $\boldsymbol{X} \sim N(\boldsymbol{\theta}, I)$,
   (d) calculate estimates $\widehat{\boldsymbol{\theta}}$,
   (e) evaluate loss $\frac{1}{k}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2$,
   (f) average loss over simulation draws,
   (g) and plot average loss as function of r.

Do this separately for $k = \dim(\boldsymbol{\theta}) = 2, 3, 10$ and for the following estimators:

(a) The MLE,

(b) the estimator $\widehat{\boldsymbol{\theta}} = (1 - 1/\overline{X^2}) \cdot \boldsymbol{X}$,

(c) the James-Stein estimator,

(d) the positive part James-Stein estimator,

(e) the estimator shrinking to the grand mean using the optimal shrinkage factor $1 - \frac{(k-3)/k}{s_X^2}$.

For a given dimension, plot the risk functions of all these estimators in one figure. Discuss your results.

2. In this problem, you are asked to implement optimal experimental designs using Gaussian process priors, as in Kasy (2016). Consider a Gaussian process prior with mean 0 and covariance kernel

$$C((x_1, d_1), (x_2, d_2)) = 10 \cdot \sigma^2 \cdot \exp\left(-\left(\|x_1 - x_2\|^2 - (d_1 - d_2)^2\right)/10\right),$$

where we assume that the variance of covariates has been standardized. Here x_i is a vector of covariates, and d_i is a binary treatment indicator.

(a) Write a function that takes as its input the (not-yet normalized) covariate matrix $\boldsymbol{X}$ as well as a vector of treatments $\boldsymbol{D}$ and provides as its output the expected MSE (normalized by σ) (i) for the Bayes estimator of the ATE, and (ii) for the difference in means estimator.

(b) Write a routine that re-randomizes treatment assignment a given number of times, evaluates expected mean squared error (i) or (ii), and provides as its output the treatment assignment with minimal expected MSE among these random draws.

(c) Find data on covariates for some field experiment, and find an optimal treatment assignment using this procedure.

References

Kasy, M. (2016). Why experimenters might not always want to randomize, and what they could do instead. *Political Analysis*, 24(3):324–338.