STAT 499: Undergraduate Research

Week 8: Tutorial on the Lasso simulation experiment Part II

Facheng Yu

Compiled on: 2024-02-29, 21:42:54

8.1 For misspecified model

Suppose that $(Y_i, X_i) \in \mathbb{R} \times \mathbb{R}^d$ are i.i.d. observed from P_{XY} for $i = 1, \dots, n$, where a constant term are included in X_i. Throughout this lecture, we assume that $\mathbb{E}[X_i X_i^\top]^{-1}$ exists.

We now want to approximate Y_i using linear combination of X_i, i.e., we consider such a model:

$$Y_i = X_i^\top \theta^* + e_i, \text{ for some } \theta^* \in \mathbb{R}^d.$$

In most case, since $\mathbb{E}[e_i|X_i] = \mathbb{E}[Y_i|X_i] - X_i^\top \theta^* \neq 0$, the model above is not correctly specified, i.e., the model is misspecified. However, even in the misspecified case, the linear approximation still exists and is useful in some way.

Without any further assumption but that $\mathbb{E}[X_i X_i^\top]^{-1}$ exists, θ^* is defined as:

$$\theta^* = \arg\min_{\theta \in \mathbb{R}^d} \mathbb{E}[(Y_i - X_i^\top \theta)^2].$$

from the exercise sheet, you proved that θ^* uniquely exists since

$$2\mathbb{E}[-X_i(Y_i - X_i^\top \theta^*)] = 0 \implies \theta^* = \mathbb{E}[X_i X_i^\top]^{-1} \mathbb{E}[X_i Y_i].$$

By the consistency of Lasso, we know that under certain conditions, we will have the lasso estimator $\hat{\theta}_n$ satisfies

$$\|\hat{\theta}_n - \theta^*\|_2 \leq \frac{3}{k}\sqrt{s}\lambda_n,$$

where λ_n is the regularization parameter of Lasso.

In specific, the conditions include:

(A1) The support S of θ^* satisfies $|S| = s$.

(A2) RE condition holds. Here we require the $\frac{1}{n}\|\mathbf{X}v\|_2 \geq k\|v\|_2$ for any $v \in \mathbb{C}(k, 3)$.

Note. This is the case under the high-dimentional setting where $\mathbf{X}^\top \mathbf{X}$ is not invertible. But for d fixed, we can simply compute k as $\lambda_{min}(\frac{\mathbf{X}^\top \mathbf{X}}{n})$ for n large enough.

(A3) $\lambda_n \geq \|\frac{\mathbf{X}^\top e_i}{n}\|_\infty$.

Thus, we can estimate θ^* through Lasso program instead of minimizing the empirical MSE.

8.2 feature selection

Suppose that the feature dimension d of the dataset (Y_i, X_i) is fixed, and θ^* is not sparse. How can do feature selection using Lasso?

We suppose that the selected feature $\tilde{X}_i$ behaves as similar as X_i. That is, there is a sparse θ_s^* satisfing:
(A4) The support S of θ^* satisfies $|S| = s$.

(A5) Here we require the $\lambda_{min}(\frac{\mathbf{X}^\top \mathbf{X}}{n}) \geq k > 0$.

(A6) $\lambda_n \geq \|\frac{\mathbf{X}^\top (Y - \mathbf{X}\theta_s^*)}{n}\|_\infty$.

And it holds that

$$\mathbb{E}[(Y_i - X_i^\top \theta_s^*)^2] \approx \mathbb{E}[(Y_i - X_i^\top \theta^*)^2].$$

In this case, θ_s^* can be approached by Lasso estimator $\hat{\theta}_n$. Let S^* be the support of $\hat{\theta}_n$, and the selected feature is $\tilde{X}_i = X_{i,S^*}$