Bootstrap

Introduction

- Goal: introduced in 1979 as a computer-based method for estimating the standard error of estimated parameters like $\hat{\theta}$;
- Rationale: having observed a random sample of size n, $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ from a probability distribution F,

$$F \to (x_1, x_2, \dots, x_n),$$

define the empirical distribution function $\hat{F}$ to be the discrete distribution that puts probability $\frac{1}{n}$ on each x_i, $i = 1,2,\dots,n$. A bootstrap sample is defined as a random sample of size n, say $\boldsymbol{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$, drawn from $\hat{F}$,

$$\hat{F} \to (x_1^*, x_2^*, \dots, x_n^*),$$

Then different bootstrap samples are used to estimate standard errors/ construct CI for parameters of interest when the true distribution unknown.

- Fun fact: for a random sample $\boldsymbol{x}$ of size n, it has $\binom{2n-1}{n}$ distinct bootstrap samples.

Procedure

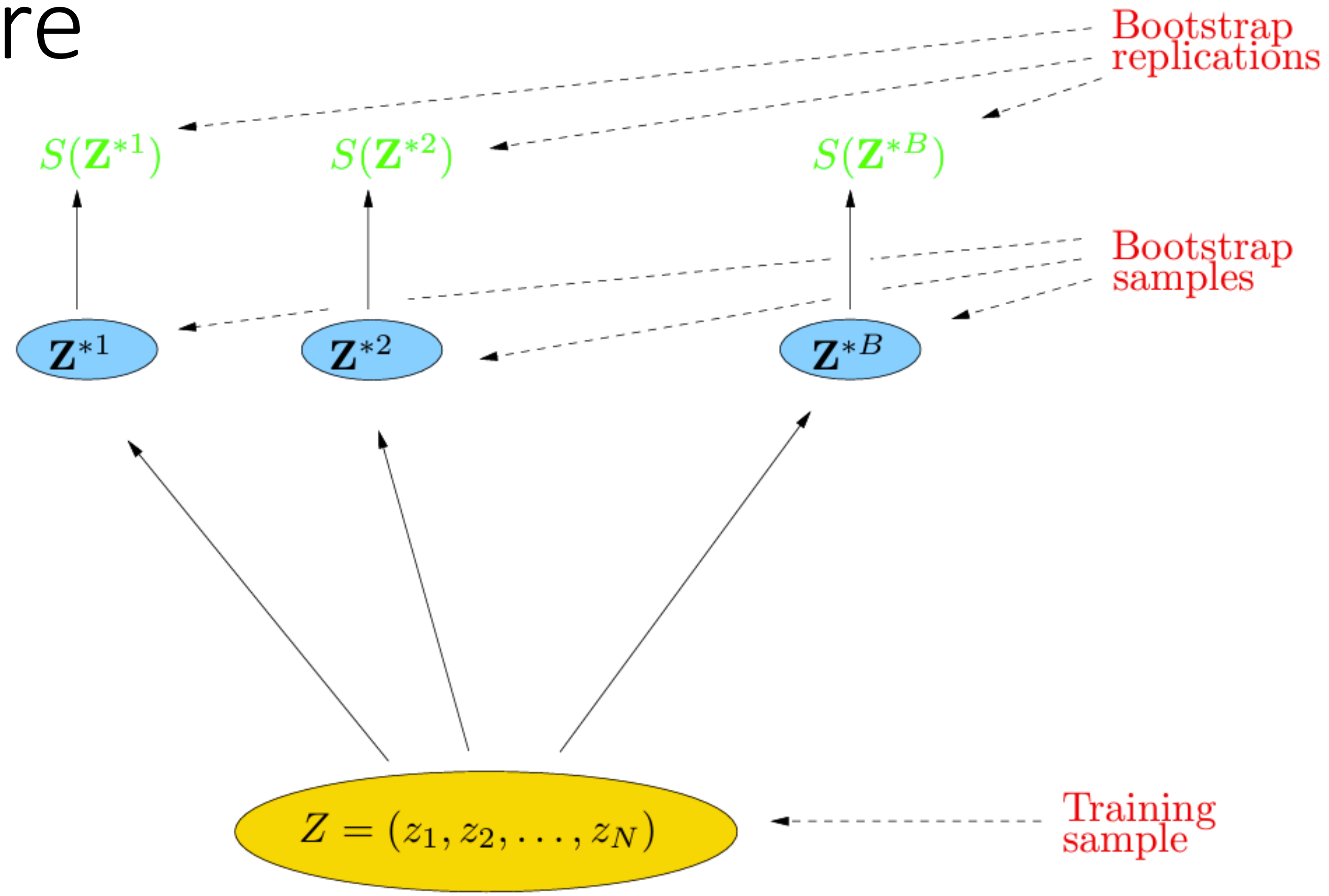

FIGURE 7.12. *Schematic of the bootstrap process. We wish to assess the statistical accuracy of a quantity $S(\mathbf{Z})$ computed from our dataset. B training sets $\mathbf{Z}^{*b}$, $b = 1, \ldots, B$ each of size N are drawn with replacement from the original dataset. The quantity of interest $S(\mathbf{Z})$ is computed from each bootstrap training set, and the values $S(\mathbf{Z}^{*1}), \ldots, S(\mathbf{Z}^{*B})$ are used to assess the statistical accuracy of $S(\mathbf{Z})$.*

Implication of CI

A 90% CI for θ implies:

90% of the time, a random interval constructed in this way will contain the true value θ.

Crude/Reflected Confidence Interval

- Crude CI is straightforward and easy to understand, it has certain problem when the underlying distribution is not that normal;

- Reflected CI comes from the percentile interval of $\hat{\theta} - \theta$: we estimate the distribution of $\hat{\theta} - \theta$ by the bootstrap distribution of $\hat{\theta}^* - \hat{\theta}$. Denote the α-percentile of $\hat{\theta}^* - \hat{\theta}$ by $\widehat{H}^{-1}(\alpha)$ and the α-percentile of $\hat{\theta}^*$ by $\hat{G}^{-1}(\alpha)$. You can show that the interval given by
$$\widehat{H}^{-1}(\alpha) \leq \hat{\theta} - \theta \leq \widehat{H}^{-1}(1 - \alpha)$$
will then give
$$\hat{\theta} + (\hat{\theta} - \hat{G}^{-1}(1 - \alpha)) \leq \theta \leq \hat{\theta} + (\hat{\theta} - \hat{G}^{-1}(\alpha))$$
which is the reflected CI.

Penalized Regression

- Implementation: use sklearn.linear_model.ridge_regression/sklearn.linear_model.Lasso;

- Lasso can restrain parameters to zero compared to ridge due to its constraint functions.

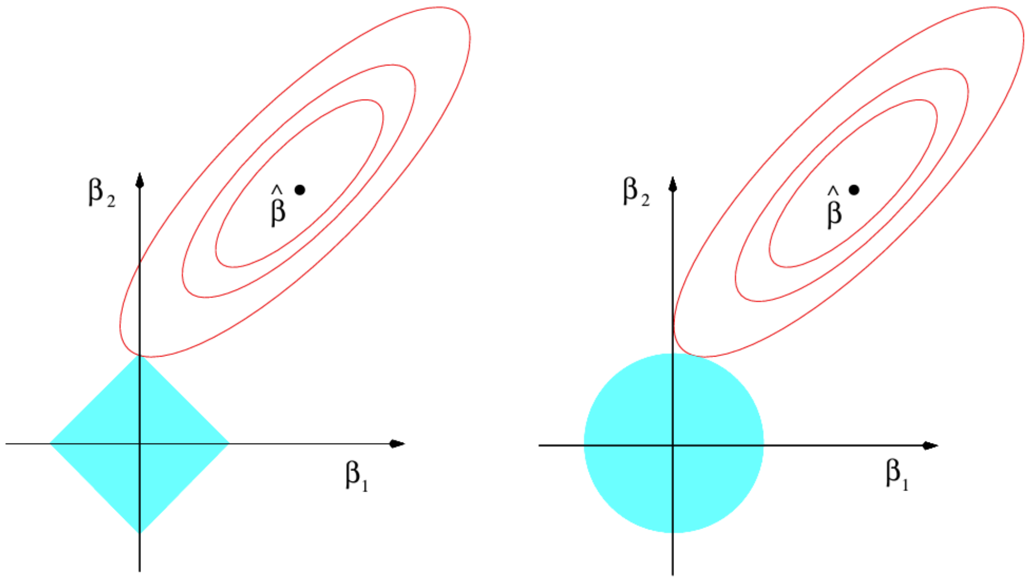

FIGURE 3.11. *Estimation picture for the lasso (left) and ridge regression (right). Shown are contours of the error and constraint functions. The solid blue areas are the constraint regions* $|\beta_1| + |\beta_2| \le t$ *and* $\beta_1^2 + \beta_2^2 \le t^2$, *respectively, while the red ellipses are the contours of the least squares error function.*