Stability Selection

Statistical Learning
CLAMSES - University of Milano-Bicocca

Aldo Solari

References

- Meinshausen, Buhlmann (2010). Stability selection. JRSS-B, 72:417–473
- Shah, Samworth (2013). Variable selection with error control: another look at stability selection. JRSS-B, 75:55–80.
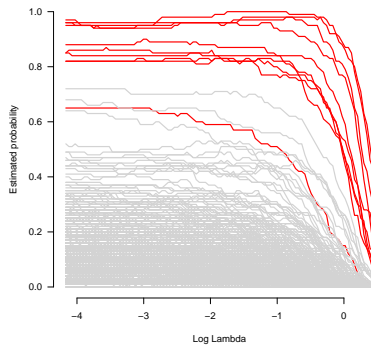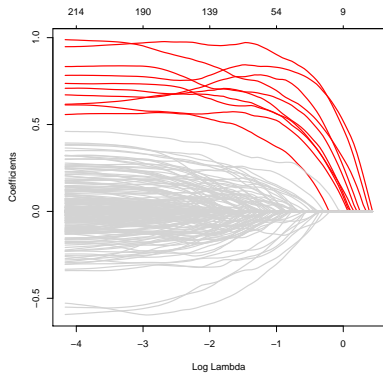
Stability path

- The *regularisation path* of the lasso is

$$\{\hat{\beta}_j(\lambda), j = 1, \ldots, p, \lambda \in \Lambda\}$$

- The *stability path* is

$$\{\hat{\pi}_j(\lambda), j = 1, \ldots, p, \lambda \in \Lambda\}$$

where $\hat{\pi}_j(\lambda)$ is the estimated probability for the jth predictor to be selected by the lasso(λ) when randomly resampling from the data

---

Algorithm 1 Stability Path Algorithm with the Lasso

Require: $B \in \mathbb{N}$, Λ grid, $\tau \in (0.5, 1)$

1: **for** $b = 1, \ldots, B$ **do**
2:      Randomly select $n/2$ indices from $\{1, \ldots, n\}$;
3:      Perform the lasso on the $n/2$ observations to obtain

$$\hat{S}_{n/2}(\lambda) = \{j : \hat{\beta}_j(\lambda) \neq 0\} \qquad \forall \ \lambda \in \Lambda$$

4: **end for**
5: Compute the relative selection frequencies:

$$\hat{\pi}_j(\lambda) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\{j \in \hat{S}_{n/2}(\lambda)\} \qquad \forall \ \lambda \in \Lambda$$

6: The set of *stable predictors* is given by

$$\hat{S}_{\text{stab}} = \{j : \max_{\lambda \in \Lambda} \hat{\pi}_j(\lambda) \geq \tau\}$$

Algorithm 2 (Complementary Pairs) Stability Selection

Require: A variable selection procedure $\hat{S}_n$, $B \in \mathbb{N}$, $\tau \in (0.5, 1)$

1: **for** $b = 1, \ldots, B$ **do**

2:     Split $\{1, \ldots, n\}$ into (I^{2b-1}, I^{2b}) of size $n/2$, and for each get

$$\hat{S}_{n/2}^{2b-1} \subseteq \{1, \ldots, p\}, \qquad \hat{S}_{n/2}^{2b} \subseteq \{1, \ldots, p\}$$

3: **end for**

4: Compute the relative selection frequencies:

$$\hat{\pi}_j = \frac{1}{2B} \sum_{b=1}^{B} \left(\mathbb{1}\{j \in \hat{S}_{n/2}^{2b-1}\} + \mathbb{1}\{j \in \hat{S}_{n/2}^{2b}\} \right)$$

5: The set of *stable predictors* is given by

$$\hat{S}_{\text{stab}} = \{j : \hat{\pi}_j \geq \tau\}$$

– The relative selection frequency $\hat{\pi}_j$ is an unbiased estimator of

$$\pi_j^{n/2} = P(j \in \hat{S}_{n/2})$$

but, in general, a biased estimator of

$$\pi_j^n = P(j \in \hat{S}_n) = \mathbb{E}(\mathbb{1}\{j \in \hat{S}_n\})$$

– The key idea of stability selection is to improve on the simple estimator $\mathbb{1}\{j \in \hat{S}_n\}$ of π_j^n through subsampling.

– By means of averaging involved in $\hat{S}_{\text{stab}}$, we hope that $\hat{\pi}_j$ will have reduced variance compared to $\mathbb{1}\{j \in \hat{S}_n\}$ and this increased stability will more than compensate for the bias incurred.

Theorem

Assume that

1. $\{\mathbb{1}\{j \in \hat{S}_{n/2}\}, j \in N\}$ *is exchangeable;*

2. *The variable selection procedure is not worse than random guessing, i.e.*

$$\frac{\mathbb{E}(|\hat{S}_{n/2} \cap S|)}{\mathbb{E}(|\hat{S}_{n/2} \cap N|)} \geq \frac{|S|}{|N|}.$$

Then, for $\tau \in (1/2, 1]$

$$\mathbb{E}(|\hat{S}_{\text{stab}} \cap N|) \leq \frac{1}{(2\tau - 1)} \frac{q^2}{p}$$

where $q = \mathbb{E}(|\hat{S}_{n/2}|)$

- The choice of the number of subsamples B is of minor importance
- It is possible to fix $q = \mathbb{E}(|\hat{S}_{n/2}|)$ and run the variable selection procedure until it selects q variables. However, if q is too small, one would select only a subset of the signal variables as

$$|\hat{S}_{\text{stab}}| \leq |\hat{S}_{n/2}| = q$$

- For example, with $p = 1000$, $q = 50$ and $\tau = 0.6$ then

$$\mathbb{E}(|\hat{S}_{\text{stab}} \cap N|) \leq 12.5$$