

Recap

Stability Selection

random subsampling of half of the data:

$I^{*1}, \dots, I^{*B}$ independent

I^{*b} random subsample $\subset \{1, \dots, n\}, |I^{*b}| = \lfloor n/2 \rfloor$
without replacement

feature selection algorithm $\hat{S}_\lambda \subseteq \{1, \dots, p\}$

stability of selected single features:

$$\hat{\Pi}_j(\lambda) = \mathbb{P}^*[j \in \hat{S}_\lambda(I^*)] \approx B^{-1} \sum_{b=1}^B I(j \in \hat{S}_\lambda(I^{*b}))$$

Why half-sampling?

I.e., subsampling without replacement with $|I^{*b}| = \lfloor n/2 \rfloor$?

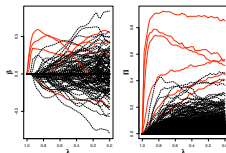
Freedman (1977): sampling without replacement with subsample size $m = \lfloor n/2 \rfloor$ is closest to i.i.d. sampling n individuals with replacement (i.e., bootstrap resampling)

“closest” means w.r.t total variance distance

Connecting to false discoveries

$$\hat{S}_{\text{stable}} = \{j; \max_{\lambda \in \Lambda} \hat{\Pi}_j(\lambda) \geq \pi_{\text{thr}}\}$$

Choice of π_{thr} ?



as a measure for type I error control (against false positives):

$$V = \text{number of false positives} = |\hat{S}_{\text{stable}} \cap S_0^c|$$

where S_0 is the set of the true relevant features

$$\hat{S}_{\Lambda} = \cup_{\lambda \in \Lambda} \hat{S}(\lambda)$$

$$q_{\Lambda} = \mathbb{E}[|\hat{S}_{\Lambda}(\underbrace{\quad}_{\text{random subsample}})|]$$