Bagged indicator function

April 22, 2015

$$\begin{aligned}
\hat{\theta}_n(x) &= \mathbb{1}_{\{\bar{Y_n} \leq x\}} \\
\frac{\sqrt{n}(\bar{Y_n} - \mu)}{\sigma} &\xrightarrow{D} N(0,1)
\end{aligned}$$

Take x in the $n^{-1/2}$ neighborhood of μ

$$\begin{aligned}
x &= x_n(c) = \mu + c\sigma n^{-1/2} \\
\hat{\theta}_n(x_n(c)) &= \mathbb{1}_{\{\bar{Y_n} \leq \mu + c\sigma n^{-1/2}\}} \\
&= \mathbb{1}_{\{\frac{\sqrt{n}(\bar{Y_n} - \mu)}{\sigma} \leq c\}} \\
&\approx \mathbb{1}_{\{Z \leq c\}}
\end{aligned}$$

This gives us

$$E(\hat{\theta}_n(x_n(c))) \to P[Z \leq c] = \Phi(c) \text{ and } Var(\hat{\theta_n}(x_n(c))) \quad \to \Phi(c)(1 - \Phi(c))$$

Now consider the bagged estimator

$$\begin{aligned}
\hat{\theta}_{B;n}(X_n(c)) &= E^{\star}[\mathbb{1}_{\{\bar{Y_n^{\star}} \leq x_n(c)\}}] \\
&= E^{\star}[\mathbb{1}_{\{\frac{\sqrt{n}(\bar{Y_n}^{\star} - \bar{Y_n})}{\sigma} \leq \frac{\sqrt{n}(x_n(c) - \bar{Y_n})}{\sigma}\}}] \\
&= \Phi\left(\frac{\sqrt{n}(x_n(c) - \bar{Y_n})}{\sigma}\right) + o_p(1) \\
&\overset{D}{\approx} \Phi(c - Z) \\
\hat{\theta}_{n;B}(x_n(0)) &\to \Phi(-Z) = U[0,1]
\end{aligned}$$

$$\begin{aligned}
E(\hat{\theta}_{n;B}(x(0)) &= 1/2 \text{So unbiased!} \\
Var(\hat{\theta}_{n;B}(x(0)) &= 1/12
\end{aligned}$$

Reduction in variance by a factor of 3!

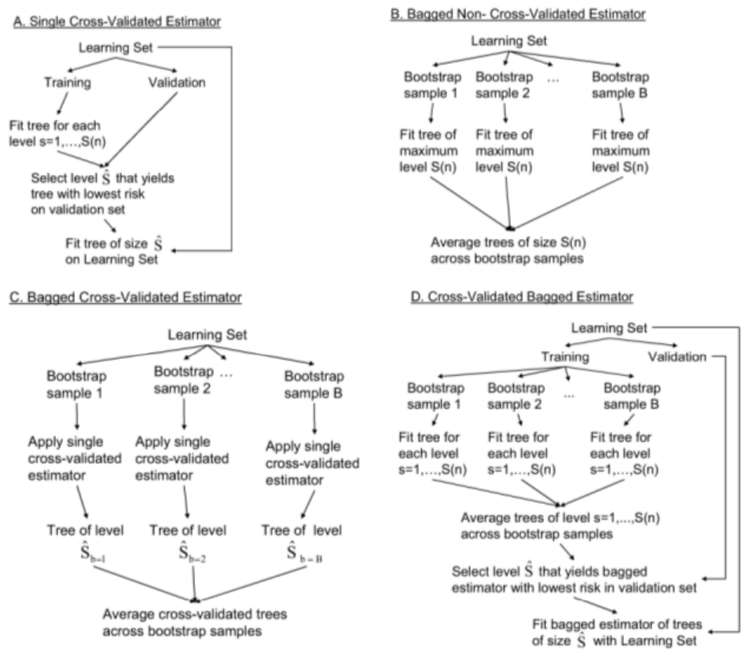

Figure 1: Flowcharts for bootstrap and cross-validation, from [2].

References:

[1] Buhlmann, P. and Yu, B. (2002). Analyzing bagging. *Ann. Statist.* **30** 927-961.

[2] Petersen, M. L., Molinaro, A. M., Sinisi, S. E., van derLaan, M. J. Cross-Validated Bagged Learning. *J Multivar Anal.* (2008). 25(2): 260-266