Literature Review

Zheng Yuan

November 2019

1 Yang's Paper

Consider the regression model

$$Y = f_k(x, \theta_k) + \epsilon,$$

where for each k, $F_k = \{f_k(x, \theta_k), \theta_k \in \Theta_k\}$ is a linear family of regression functions with θ_k being the parameter of a finite dimension m_k.

- Average squared error: for a model selection criterion δ that selects model $\hat{k}$, $ASE(f_{\hat{k}}) = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - f_{\hat{k}}(x_i, \hat{\theta}_{\hat{k}}))^2$
- Risk function: $R(f; \delta; n) = \frac{1}{n}\sum_{i=1}^{n} E(f(x_i) - f_{\hat{k}}(x_i, \hat{\theta}_{\hat{k}}))^2$

Theorem 1: Suppose that model $k* \in \Gamma$ is the true model. Then

$$sup_{f \in F_{k^*}} R(f; \delta_{AIC}; n) \leq \frac{Cm_{k^*}}{n},$$

$m_k^* : dimension\ of\ model\ k^*$. Thus the worst-case risk of δ_{AIC} under the true model k^* is at the parametric rate $\frac{1}{n}$.

Assumption 1: There exists two models $k_1, k_2 \in \Gamma$ such that

- $F_{k_1} = \{f_{k_1}(x, \theta_k), \theta_{k_1} \in \Theta_{k_1}\}$ is a sub-linear space of $F_{k_2} = \{f_{k_2}(x, \theta_k), \theta_{k_2} \in \Theta_{k_2}\}$
- There exists a function $\phi(x)$ in F_{k_2} orthogonal to F_{k_1} with $\frac{1}{n}\sum_{i=1}^{n}\phi^2(x_i)$ being bounded between two positive constants;
- There exists a function $f_0 \in F_{k_1}$ such that f_0 is not in any family F_k that does not contain F_{k_1}

Theorem 2: Under Assumption 1, if any model selection method δ is consistent in selection, then we must have

$$n \ \ sup_{f \in F_{k_2}} R(f; \delta; n) \to \infty$$

The theorem says that in the parametric case, if one is to pursue consistency in selection, one must pay a somewhat high price for estimating the regression function.

2 Leave One Out Error and RSS

In Yang's notation, $R(f;\delta;n) = \frac{1}{n}\sum_{i=1}^{n} E(f(x_i) - f_{\hat{k}}(x_i,\hat{\theta}_{\hat{k}}))^2$. This can be written as

$$R(f;\delta;n) = \frac{1}{n}E[RSS_{M_{\hat{k}}}]$$

We have showed before that

$$\lim_{n\to\infty}[log\hat{\Gamma}^{CV}_{\alpha,n} - log\frac{RSS_{\alpha}}{n}] = 0 \tag{1}$$

which implies

$$\lim_{n\to\infty}[log\hat{\Gamma}^{CV}_{\alpha,n} - log\frac{RSS_{\alpha}}{n}] = 0 \tag{2}$$

which implies that $\lim_{n\to\infty} E[\hat{\Gamma}^{CV}_{\alpha,n}] - \frac{1}{n}E[RSS_{M_{\hat{k}}}] = 0$ and thus

$$\lim_{n\to\infty} R(f;\delta;n) - E[\hat{\Gamma}^{CV}_{\alpha,n}] = 0$$