A Theoretical Summary of Cross-Validation Frameworks with Closed-From Expressions

Presenter: Chieh Lin
Andrew ID: chiehl1

Main focus:
leave-p-out (LPO) and leave-one-out (LOO)

Framework	Density Estimation	Regression	Classification
Estimators	LPO: Projection Estimator (ex: histogram) Kernel Estimator	LPO: Projection Estimator Regresogram Estimator (with condition) Kernel Estimator (with condition) LOO: Linear Smoother	✗

Projection Density Estimator LPO Closed-Form

Presenter: Chieh Lin
Andrew ID: chiehl1

Definition: projection estimator

$$\hat{s}_m = \sum_{\lambda \in \Lambda(m)} \hat{\beta}_\lambda \phi_\lambda \text{ with } \hat{\beta}_\lambda = \frac{1}{n}\sum_{i=1}^{n} \phi_\lambda(Z_i)$$

where

$\Lambda(m)$: a set of indices

$\{\phi_\lambda\}_{\lambda \in \Lambda(m)}$: an orthonormal family of functions

Lemma1: simple combinatorial results

$$\sum_{e \in e_p} 1_{(j \in \bar{e})} = \binom{n-1}{p} \qquad \sum_{e \in e_p} 1_{(j \in \bar{e})} 1_{(k \in \bar{e})} = \binom{n-2}{p} \qquad \sum_{e \in e_p} 1_{(i \in e)} 1_{(j \in \bar{e})} = \binom{n-2}{p-1}$$

Lemma2: follows from definition and Lemma1

$$\sum_{e \in e_p} ||\hat{s}_m^{\bar{e}}||_2^2 = \frac{1}{(n-p)^2}\Big[\binom{n-1}{p}\sum_{k=1}^{n} ||H_m(X_k, .)||_2^2 + \binom{n-2}{p}\sum_{k \neq l} < H_m(X_k, .), H_m(X_l, .) >_2\Big]$$

$$\sum_{e \in e_p}\sum_{i \in e} \hat{s}_m^{\bar{e}}(X_i) = \frac{1}{n-p}\binom{n-2}{p-1}\sum_{i \neq j} H_m(X_i, X_j)$$

where

$$H_m(X_i, X_j) = \sum_{\lambda \in \Lambda(m)} \phi_\lambda(X_j)\phi_\lambda(X_i)$$

Theorem: LPO closed-from CV of projection density estimator

$$\hat{L}_p(\hat{s}_m) = \frac{1}{n(n-p)}\sum_{\lambda \in \Lambda(m)}\Big[\sum_j \phi_\lambda^2(X_j) - \frac{n-p+1}{n-1}\sum_{j \neq k}\phi_\lambda(X_j)\phi_\lambda(X_k)\Big]$$