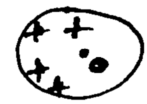
LEE JONES UMASS Lowell

Local Finite Sample Minimax Estimation

$$Y_j = f(x_j) + N_j \qquad \operatorname{cov}\{N_j\} = \mathcal{N} < \sigma$$

$$\hat{f}(0) = w^* + \sum_1^k w_j Y_j = F(w)$$

 V

[1] local approximating parametric (nonparametric) family

$\{ f(x;a) \}$

[2] local approximating error function $\epsilon(x)$ $(\epsilon(0)=0)$

[3] structural conditions on f - $\mathcal{C}$ $(f \in \mathcal{C})$

$$\boxed{W = \operatorname{Arg} \min_{w} \max_{\substack{f \in \mathcal{C} \\ |f(x) - f(x;a_f)| \le \epsilon(x)}} E\left[F(w) - f(0)\right]^2}$$

compare to global learning $\min_{\hat{f}} \max_{\mathcal{C}} \|\hat{f}(x) - f(x)\|_{P(x)}$

$\| \|_p = \| \|$ wrt. predictor measure $P(x)$ (Cucker, Smale, ... Temlyakov)

We are attempting to give a "window" for $\uparrow$ (in singular case

EXS $f(x;a) = a_0 + a \cdot x$ linear

parametric $= a_0 + a \cdot x + x^t B x$ quadratic

non parametric $= \sum_{x'} a_{x'} K(x', x)$ with $|||f(x;a)||| \le M$

Reproducing kernel Hilbert sp.

Support vector machine

$\mathcal{C}$: $0 \le f(x) \le 1$

$f(x) = Pr(z / x)$

classification

$|f(x) - f(y)| \le 2\overline{M}$

bnd. oscillation

(Lipschitz 0)

Global Finite Sample Minimax Estimation

$$e_k(\mathcal{C}) = \inf_{F} \sup_{f \in \mathcal{C}} E\left(\|f - F\|_{L_2(P)}\right)$$

ASSUME $x_j = X_j$; (X_j, Y_j) i.i.d. $|Y - f(X)| \le M_0$; $X \in \mathcal{X}$

CONSIDER A ball $b(W^s(L_\infty(\mathcal{X}))) = \mathcal{C}$ and $f \in \mathcal{C}$

Then

① $$E\left(\|f - \bar{F}\|_{L_2(P)}\right) \le C\, k^{-s/2s+2d}$$

$\bar{F}$ = element of $\mathcal{C}$ with min empirical squared error

Cucker - Smale

② $$e_k(\mathcal{C}) \ge C'\, k^{-s/2s+d}$$

STONE; DeVore, Kerkyacharian, Picard, Temlyakov

③ $$e_k(\mathcal{C}) \le C''\, k^{-\frac{s}{2s+d}}$$

Konyagin, Temlyakov

minimizing $\bar{F}$ is an ϵ-net

(computationally infeasible)

(1')

Solutions and Approximate Solutions

Old and New Algor.

Ridge Regression

Tikhonov Reg.

identify ridge, reg. γ

$$\min \frac{1}{k}(g-Y)^t\sigma^{-1}(g-Y) + \gamma |||g|||^2$$

$$\gamma = k^{-1}M^{-2}$$

Classification

1. Quad. prog.
2. choice of M
3. locally approx Quad. sol. to Finite (Fourier) mom. p.

Bounds (New)

$$\max_{f \in \mathcal{C}} E\left(F(w) - f(0)\right)^2 \le \underline{\underline{L(w) = B(w) + w^t\sigma w}}$$

$$|f(x) - f(x; a_f)| \le \epsilon(x)$$

Ex. Reprod. K. H. S.

$$\epsilon(x) = 0 \quad \min_w L(w) = \left(\sum_i \sum_j \left(\sigma + M^2 K\right)^{-1}\right)^{-1}$$

$(k+1)\times(k+1)$

ENSEMBLING, BAGGING, AVERAGING ← $\boxed{\text{Fusion}}$

$x_j = \mathbb{X}_j$ $\quad (\mathbb{X}_j, \mathbb{Y}_j)$ i.i.d.

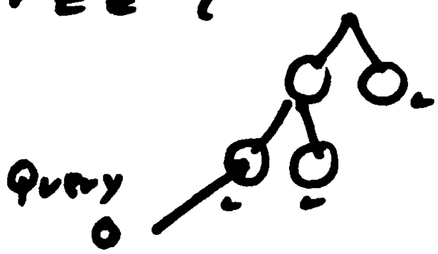
EMPIRICAL EVIDENCE: AVERAGING (RANDOM) alg. improves PERFORMANCE (over any single alg.)

Ex. 1) Alg. i. : proj. onto random subspace U_i

2) Alg i. : use kernel $K_i(X_j', x)$ with shape (bandwidth) only in dir. in U_i

3) RANDOM FOREST : TREE i

DATA parsed to terminal node corr. to subspace U_i & data subset $\{X_j^i\}$



1) & 2) could also use subset to localize

EXPERT i

(loc. lin. case) $\mathbb{Y}_j = f^i(0) + a^i \cdot x_j^i + \xi_j^i + N_j^i$

$F_i(w^i) = N^{*i} + \sum_j w_j^i \mathbb{Y}_j$ only for x_j in $\{X_j^i\}$

$\boxed{f^i(x) = E(\mathbb{Y} \mid \mathrm{proj}_{U_i} = x)}$ $\quad N^i < \sigma^i$ (min. $L_i(w^i)$)

$$f^i(0) = E(Y \mid X \in A_i) \qquad A_i = \text{"}\operatorname{proj}_{U_i} X = 0\text{"}$$

$I(A)$ = information measure e.g. $\operatorname{codim} A_i + 1$
$= \dim U_i + 1$

P prob. m. on Ω (σ-alg)

$$\boxed{\begin{aligned} \mathcal{E}(Y \mid P) &= \int_\Omega E(Y \mid X \in A)\, dP(A) \\ \mathcal{I}(P) &= \int_\Omega I(A)\, dP(A) \end{aligned}}$$

Experts i: $F = \sum \alpha_i F_i \qquad \mathcal{E}(Y \mid \alpha) = \sum \alpha_i E(\cdot \mid \cdot)$

$\mathcal{I}(\alpha) = \sum \alpha_i I(A_i)$

$$E(F - \mathcal{E}(Y \mid \alpha))^2 \le \underline{\underline{G(w, \alpha)}}$$

Appl. pb.

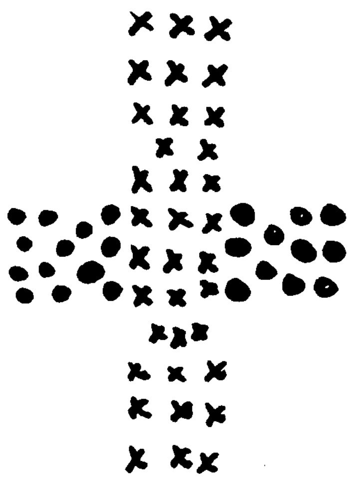
$$\boxed{\min_{w, \alpha} \; G(w, \alpha) + h\left(\frac{1}{\mathcal{I}(\alpha)}, w\right)}$$

classification

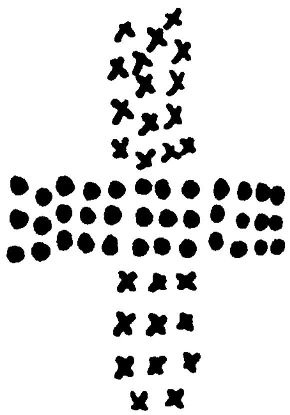
$$\underline{\underline{G(w, \alpha) + \lambda \left(\frac{1}{\mathcal{I}(\alpha)} - \frac{1}{d+1}\right)^{\frac{1}{F^\beta (1-F)^\beta}}}}$$

FUSION

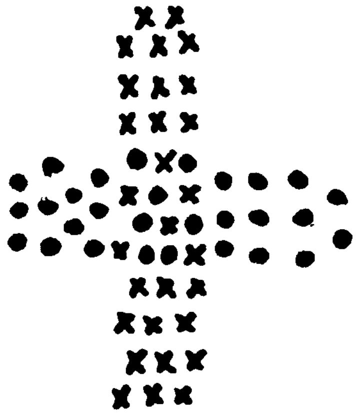
①



②



③



	$E(Y \mid X_1 = X_2 = 0)$	$E(Y \mid X_1 = 0)$	$E(Y \mid X_2 = 0)$	$\mathcal{E}(Y \mid \mathcal{P})$
①	0	$0 \quad (0)$	$2/3 \quad \left(\frac{2}{3}\right)$	$1/3 \quad \left(\frac{1}{3}\right)$
②	1	$1/3 \quad \left(\frac{2}{3}\right)$	$1 \quad (0)$	$2/3 \quad \left(\frac{1}{3}\right)$
③	$1/2$	$1/6 \quad \left(\frac{1}{3}\right)$	$5/6 \quad \left(\frac{1}{3}\right)$	$\frac{1}{2} \quad (0)$

$$\mathcal{E}(Y \mid \mathcal{P}) = \int_{\Omega} E(Y \mid X \in A)\, d\mathcal{P}(A)$$

Generalized Conditional Expectation

CURSE OF DIMENSIONALITY

if expert i corrupted, incorrect with prob. π_0 (or bnd by π_0)

For class. take $\epsilon^{i*}(x) = \begin{cases} 0 & x=0 \\ 1 & x \neq 0 \end{cases}$

obtain bnds L_i^*

get bound $G^*(n, \alpha)$

which is expected bnd. on M.S.E. which is bnd. on expected M.S.E. which is M.S.E.