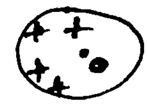
LEE JONES UMASS LOWELL

LOCAL FINITE SAMPLE MINIMAX ESTIMATION

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

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

 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

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

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 $x_i = X_i$; (X_i, Y_i) 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)

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)

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

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

Ex. Reprod. K. H. S.

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

$(k+\ell) \cdot (k+1)$

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

$x_j = X_j \quad (X_j, Y_j)$ 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

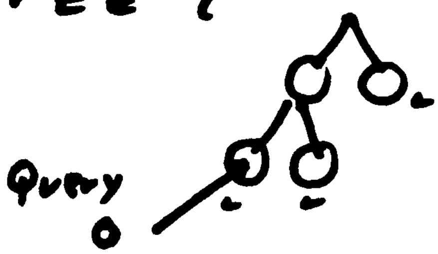
3) RANDOM FOREST : TREE i

DATA PARSED

→ TERMINAL node

corr. to subspace U_i

& data subset $\{X_j^i\}$



1) & 2) could also use subset to localize

EXPERT i

$$\left(\begin{array}{c}\text{loc. lin}\\ \text{case}\end{array}\right) \quad Y_j = f^i(0) + a^i \cdot x_j^i + \xi_j^i + N_j^i$$

$$F_i(w^i) = N^{*i} + \sum_j w_j^i Y_j \qquad \text{only for } x_j \text{ in } \{X_j^i\}$$

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

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

$I(A)$ = information measure e.g. $\mathrm{codim}\, A_i + 1$

$= \dim U_i + 1$

$\mathcal{P}$ prob. m. on Ω (σ-alg)

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

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

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

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

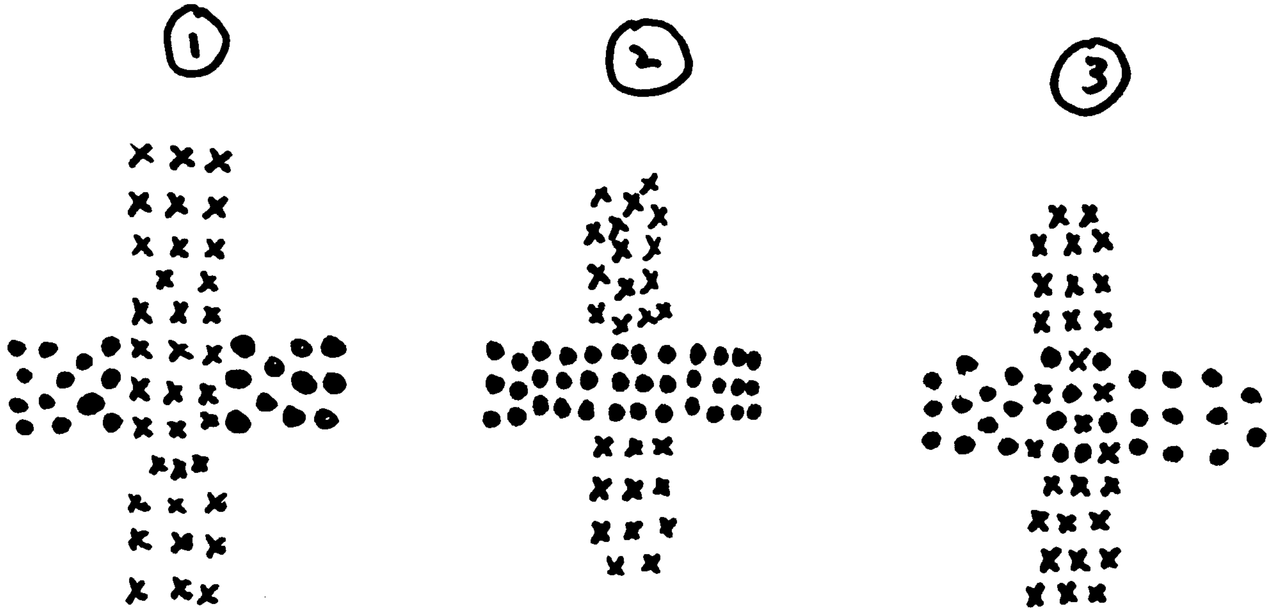
Appl. pb.

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

classification

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

Fusion



	$E(\mathbf{Y} \mid \mathbf{X}_1 = \mathbf{X}_2 = 0)$	$E(\mathbf{Y} \mid \mathbf{X}_1 = 0)$	$E(\mathbf{Y} \mid \mathbf{X}_2 = 0)$	$\mathcal{E}(\mathbf{Y} \mid P)$
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	$1/2$	$1/6 \quad \left(\frac{1}{3}\right)$	$5/6 \quad \left(\frac{1}{3}\right)$	$\frac{1}{2} \quad (0)$

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

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.