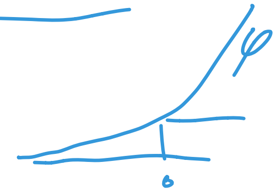
Finishing up kernel machines

Surrogate empirical risk:



$$\min_f A_{\varphi,n}(f) = \frac{1}{n}\sum_{i=1}^{n} \varphi(-Y_i f(X_i))$$

minimize over $\|f\|_K^2 \le \lambda^2$

or add penalty term $+\tau \|f\|_K^2$

If we use representer theorem, becomes

$$f(x) = \sum_{j=1}^{n} c_j K(x, x_j) \qquad \|f\|^2$$

$$\langle K_{x_i}, K_{x_j}\rangle = K(x_i, x_j)$$

$$\min_{c \in \mathbb{R}^n} \frac{1}{n}\sum_{i=1}^{n} \varphi\Big(-Y_i \sum_{j=1}^{n} c_j K(x_i, x_j)\Big)$$

subject to $\displaystyle\sum_{i=1, j=1}^{n} c_i c_j K(x_i, x_j) \le \lambda^2$

or add $\displaystyle\tau \sum_{i,j} c_i c_j K(X_i, X_j)$

Or, if $\psi_1, \psi_2, \ldots$ is an o.n. basis for $\mathcal{F}$

$$f(x) = \sum_j a_j \psi_j(x) \qquad \Big(K(x, x') = \sum_j \psi_j(x)\psi_j(x') \Big)$$

$$\min_{a} \frac{1}{n}\sum_{i=1}^{n} \varphi\Big(-Y_i \sum_j a_j \psi_j(X_i)\Big)$$

subject to $\displaystyle\sum_j a_j^2 \le \lambda^2$

or add $\displaystyle+\tau \sum_j a_j^2$

Chapter 9 Regression with quadratic loss

$$\left(\mathcal{X} \times \mathbb{R}, \mathcal{F}, P, \ell(y,u) = (y-u)^2 \right)$$

subset $\mathbb{R}^d$

observe $Z^n = ((X_1, Y_1), \dots (X_n, Y_n))$ - independent P

$A(Z^n) =$ some f use $f(X)$ to predict Y.

$$\underset{\underline{\text{all}}\ f}{\arg\min}\ E_P[(Y - f(X))^2] = E(Y|X]$$

$f^*(x) = E(Y|X=x]$

$$L^* \triangleq E[(Y - f^*(X))^2]$$

$$L^*(\mathcal{F}) = \min_f L(f) = \min_f E[(Y - f(X))^2] \geq L^*$$

$L^*(\mathcal{F}) - L^*$ = approximation error due regularization.

$$= \inf_{f \in \mathcal{F}} \|f - f^*\|^2_{P_X} = E[(f(X) - f^*(X))^2]$$

(Note $E[(Y-f(X))^2] = E[(Y - f^*(X) + f^*(X) - f(X))^2]$

$$= L^* + \|f^* - f\|^2_{P_X}$$

9.1 Hard constraint | 9.2 Additive penalty constraint

Let K be a kernel on $X \times X$

$$\mathcal{F} = \mathcal{F}_\lambda = \{ f \in \mathcal{H}_K : \|f\|_K \le \lambda \}$$

ERM

$$\hat{f}_n = \underset{f \in \mathcal{F}_\lambda}{\arg\min}\ L_n(f) = P(\ell_f) \quad \ell_f(z) = (y - f(x))^2$$

$z = (x, y)$

$$= \underset{f \in \mathcal{F}_\lambda}{\arg\min}\ \frac{1}{n} \sum_{i=1}^{n} (Y_i - f(X_i))^2$$

$\varphi(-y f(x))$

Gather our forces

Use UCEM – want $L_n(f) \approx L(f)$ for all $f \in \mathcal{F}_\lambda$

$$\Delta_n(Z^n) = \sup_f |L_n(f) - L(f)|$$

$$E\,\Delta_n(Z^n) \le 2 E R_n(\mathcal{L}_\lambda(Z^n))$$ sort of

$$\mathcal{L}_\lambda = \{ \ell_f : f \in \mathcal{F}_\lambda \} \quad \ell_f(z) = (y - f(x))^2$$

$z = (x, y)$

Use McDiarmid inequality

contraction priciple.

Assume $|Y| \le M$ for some $M \ge 0$.

RKHS theory
$$\|f\|_\infty \le C_k \|f\|_K \qquad C_k = \sup_x \sqrt{K(x,x)}$$

Theorem 9.1 With probability $1-\delta$, the ERM algorithm outputs $\hat{f}_n$ satisfying

$$L(\hat{f}_n) \le L^*(f_\lambda) + \frac{16(M+C_k\lambda)^2}{\sqrt{n}} + (M^2 + C_k^2\lambda^2) \cdot \sqrt{\frac{8\log(1/\delta)}{n}}$$

$$\Delta_n(Z^n) = \sup_f L_n(f) - L(f)$$
$$= \sup_f \left| \frac{1}{n} \sum (Y_i - f(X_i))^2 - L(f) \right|$$

If for some i, change (X_i, Y_i) to (X_i', Y_i') to get $Z^n_{(i)}$

$$\left|\Delta_n(Z^n) - \Delta_n(Z^n_{(i)})\right| \le \sup_f \frac{1}{n}\left|\underbrace{(Y_i - f(X_i))^2}_{\in [0,\, 2M^2 + 2(\lambda C_k)^2]} - (Y_i' - f(X_i'))^2\right|$$

$Y_i \in [-M, M]$

$|f(X_i)| \le \|f\|_K \cdot C_k \le \lambda C_k$

$(a+b)^2 \le 2a^2 + 2b^2$

$\therefore$ $\Delta_n(Z^n)$ has Bounded difference property with

$$c_i = \frac{2(M^2 + C_k^2\lambda^2)}{n}$$

$\therefore$ McDiarmid:

$$P\left(\Delta_n(Z^n) - E[\Delta^n(Z^n)] \geq t \right) \leq e^{-\frac{nt^2}{(M^2 + C_\lambda^2 \lambda^2)^2}}$$

$$L(\hat{f}_n) - L^*(\mathcal{F}) \leq 2\Delta_n(Z_n)$$ by the mismatch minimization lemma.

$$E\Delta_n(Z^n) \leq 2ER_n(\underline{\mathcal{L}_\lambda}(Z^n))$$

$$\mathcal{L}_\lambda = \{ l_f : f \in \mathcal{F}_\lambda \}$$

$$l_f(x, y) = (y - f(x))^2$$

$$R_n(\mathcal{L}_\lambda(Z^n))$$

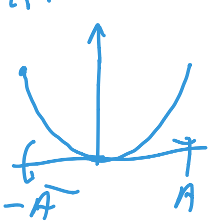
$$= R_n\left(\left\{ \left((Y_1 - f(X_1))^2, (Y_2 - f(X_2))^2, \ldots, (Y_n - f(X_n))^2 \right) : f \in \mathcal{F}_\lambda \right\} \right)$$

$$= E_{\varepsilon^n}\left[\sup_{f \in \mathcal{F}_\lambda} \frac{1}{n} \left| \sum_{i=1}^{n} (Y_i - f(X_i))^2 \varepsilon_i \right| \right]$$

$$|Y_i - f(X_i)| \leq |Y_i| + |f(X_i)| \leq M + \lambda C_K = A$$

$\varphi(u) = u^2$

for [redacted] $[-A, A]$



$\varphi'(u) = 2u$

$\varphi(0) = 0$

$\therefore \varphi$ is $2(M + \lambda C_K)$ Lipschitz continuous

$$R_n(\mathcal{L}_\lambda(Z^n))$$

$$\leq 2\cdot 2(M+\lambda C_k) R_n\left(\left\{\left((Y_1 - f(X_1)), (Y_2 - f(X_2)), \ldots, (Y_n - f(X_n))\right) : f \in \mathcal{F}_\lambda\right\}\right)$$

$$= \quad " \quad E_{\varepsilon^n}\left[\sup_{f \in \mathcal{F}_\lambda} \frac{1}{n}\left|\sum_{i=1}^{n} (Y_i - f(X_i))\,\varepsilon_i\right|\right]$$

$$\leq \quad " \quad E_{\varepsilon_n}\left[\frac{1}{n}\left|\sum_{i=1}^{n} Y_i \varepsilon_i\right|\right] + E_{\varepsilon_n}\left[\frac{1}{n}\sum |f(X_i)\,\varepsilon_i\right]$$

$$\lambda\sqrt{\frac{E[K(X,X)]}{n}}$$

$$\vdots$$

(out of time — almost done with proof of theorem.)