Lahiru Dulanjana Chaman Heera Gamage

915519699

E1

$x_i \in \mathbb{R}^p \quad y_i \in \{0, 1\} \; i = 1 \ldots n. \quad g : \mathbb{R}^p \longrightarrow \{0, 1\}$

1. True risk $R(g) = E\, l(Y, g(X))$

$$= E_X\left[E_{Y/X}\left[l(Y, g(X)) \mid X\right]\right]$$

$$l(y, g(x)) \begin{cases} 1 & \text{if } y \neq g(x) \\ 0 & y = g(x) \end{cases}$$

$$g^* = \arg\min_g E_X\left[E_{Y/X}\left[1\{y \neq g(x)\} \mid X\right]\right]$$

$$= \arg\min_g E_X\left[\Pr(y \neq g(x) \mid x)\right]$$

$$= \arg\min_g E_X\left[1 - \Pr(y = g(x) \mid x)\right]$$

we can select g^* such that $\Pr(y = g^*(x) \mid x)$ is the maximum for each $y \in \{0, 1\}$

This is Bayes rule.

2. $h(x) = 1\{x_j > 0\}$

empirical risk $= R_n(g) = \frac{1}{n} \sum_{i=1}^{n} l(y_i, g(x_i))$

$$= \frac{1}{n} \sum_{i=1}^{n} l(y_i, h(x_i))$$

$$g^* = \arg\min_g \frac{1}{n} \sum_{i=1}^{n} l(y_i, h(x_i))$$

where $h(x_i) = 1\{x_{ij} > 0\}$

fit: we can calculate $\frac{1}{n} \sum_{i=1}^{n} \ell(y_i, h(x_i))$

for each j. loss can be hamming loss.

$$j^* = \arg\min_j \left\{ \frac{1}{n} \sum_{i=1}^{n} \ell(y_i, h(x_i)) \right\}$$

predict : for x_{n+1}
$$\begin{cases} \hat{y}_{n+1} = 1 & \text{if } x_{n+1, j^*} > 0 \\ \hat{y}_{n+1} = 0 & x_{n+1, j^*} \le 0 \end{cases}$$

3. $P\{ |R_n(g) - R(g)| > t \} \le 2e^{-2nt^2}$

$p = 10$, let gatern $\hat{g}$,

$$P\{ |R_n(\hat{g}) - R(\hat{g})| > t \} \le 2e^{-2nt^2}$$

$$= \Pr\{ R_n(\hat{g}) - R_{\hat{g}} > t \cup R(\hat{g}) - R_n(\hat{g}) > t \}$$

$$1 - \Pr\{ |R_n(\hat{g}) - R(\hat{g})| \le t \} \le 2e^{-2nt^2}$$

$$\Pr\{ |R_n(\hat{g}) - R(\hat{g})| \le t \} \ge 1 - 2e^{-2nt^2}$$

$$\Rightarrow \Pr\{ -t \le R_n(\hat{g}) - R(\hat{g}) \le t \} \ge 1 - 2e^{-2nt^2}$$

union bund

$$\Pr(R_n(\hat{g}) - R(\hat{g}) \le t) \ge \Pr\{ -t \le R_n(\hat{g}) - R(\hat{g}) \le t \} \ge 1 - 2e^{-2nt^2}$$

let $t = 0.1$,

$$\Rightarrow \Pr(R_n(\hat{g}) - R(\hat{g}) \le 0.1) \ge 1 - 2e^{-0.02n}$$

$$\Pr(R_n(\hat{g}) \le R(\hat{g}) + 0.1) \ge 1 - 2e^{-0.02n} \ge 0.95$$

$\Rightarrow$ 184 samples //

Ex 2

OLS

$$\hat{y} = X\hat{\beta} \quad \text{where } \hat{\beta} = (X^T X)^{-1} X^T y$$

$$\Rightarrow \hat{y} = \underbrace{X(X^T X)^{-1} X^T}_{H} y$$

$$\hat{y} = Hy \qquad H \text{ is idempotent.}$$

kNN

$$\hat{y}_j = \frac{1}{k} \sum_{x_i \in N_k(x_j)} y_j$$

$$= \frac{1}{k} \overbrace{\begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & & & \\ & & & \\ h_{n1} & \cdots & & h_{nn} \end{bmatrix}}^{H(k)} \cdot y$$

here

$$h_{ij} \begin{cases} 1 & \text{if } x_j \in N_k(x_i) \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad h_{ii} = 1$$

$$\hat{y} = H(k) y \quad //$$

2. If sample doesn't use it's own point

$$H_{mod}(k-1) = H(k) - I_{n \times n}$$

$$\Rightarrow H_{mod}(k) = H(k+1) - I_{n \times n} \quad //$$

3. cross validation error $= \frac{1}{n} \sum_{l=1}^{n} (y_l - \hat{y}_l)^2$

$$= \frac{1}{n} \sum_{l=1}^{n} (y_l - H_{mod}^T(k) y)^2$$

$$= \frac{1}{n} \| y - H_{mod}(k) y \|^2 \quad //$$

3. $X = UDV^T$

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

$$= (V D \underbrace{U^T U}_{I} D V^T)^{-1} V D U^T y$$

$$(V D^2 V^T)^{-1} V D U^T y$$

$\Downarrow$

(reciprocals of eigenvals)

$$V D^{-2} \underbrace{V^T V}_{I} D U^T y$$

$$= \underbrace{V D^{-1}}_{\substack{A \\ p \times p}} \underbrace{U^T y}_{b \; p \times 1}$$

fit: take SVD of $X = U, D, V^T$

take U ~~as~~ the first p columns of U.

D^{-1} is the reciprocated diagonal of D.

$$A = V D^{-1} \quad b = U^T y$$

$$\hat{\beta} = Ab.$$

predict. $y_{n+1} = x_{n+1}^T \hat{\beta}$ //

4. changerank (V, D, r)

{ $D_r^{-1} \Leftarrow D^{-1}[r:p, r:p] = 0$

$A_r \Leftarrow (V D_r^{-1})[r:p, r:p] = 0$

return A_r

complexity $\Rightarrow O(r^2)$ //