Wednesday, January 10, 2024 1:50 AM

$MSE_f = E[(f(x) - Y)^2]$ f: estimator $\begin{cases} \text{unlinear} \\ \text{linear} \end{cases}$

$X \in \mathbb{R}^d$

$Y \in \mathbb{R}$

$\boxed{f \in C^\infty(\mathbb{R}^d)}$ $f \in C(\mathbb{R})$

$\boxed{f \in C^1(\mathbb{R})}$

Taylor expansion

$$f(x) = f(x_0) + \nabla f(x_0)(x - x_0) + o(x - x_0)$$

$x \in \mathbb{R}^d$ $\lim_{x \to x_0} \frac{o(x-x_0)}{x - x_0} = 0$

model class

$f \in \mathcal{F}$ $\mathcal{F} = \{ f ; f = ax^2 + bx + c,\ a, b, c \in \mathbb{R} \}$

linear model $\rightarrow \mathcal{F} = \{ f : f = a^T x + b,\ a \in \mathbb{R}^d,\ x \in \mathbb{R}^d, b \in \mathbb{R} \}$

$$\boxed{f^* = \underset{f \in \mathcal{F}}{\operatorname{argmin}}\ E[(f(x) - y)^2]}$$

□ Known X and Y distribution .

What if we don't know the distribution?

If we observe (x_i, y_i) from $P(X, Y)$
$i = 1, 2, \cdots, n$

$\boxed{\frac{1}{n}\sum_{i=1}^{n} (f(x_i) - y_i)^2} \rightarrow$ Empirical MSE

$\downarrow$ in probability

$E[(f(x) - y)^2]$

* Law of Large numbers and continuous mapping $f \in C(\mathbb{R}^d)$, $x_n \xrightarrow{p.d} x$, $f(x_n) \xrightarrow{p.d} f(x)$

$\frac{1}{n}\sum_{i=1}^{n} (f(x_i)^2 - 2f(x_i) y_i + y_i^2)$

$= \frac{1}{n}\sum_{i=1}^{n} f(x_i)^2 - \frac{1}{n}\sum_{i=1}^{n} 2f(x_i)y_i + \frac{1}{n}\sum_{i=1}^{n} y_i^2$

$= E[f(x)^2] - 2E[f(x)Y] + E[Y^2]$

$= E[(f(x) - y)^2]$

$\boxed{f^*} = \underset{f \in \mathcal{F}}{\operatorname{argmin}}\ E[\underbrace{(f(x) - y)^2}_{MSE}]$

$\boxed{\hat{f}} = \underset{f \in \mathcal{F}}{\operatorname{argmin}}\ \frac{1}{n}\sum_{i=1}^{n} (f(x_i) - y_i)^2$

} best predictor in $\mathcal{F}$

$E[(f^*(x) - \hat{f}(x))^2] \longrightarrow 0\ ? \Rightarrow \hat{f} \rightarrow f^*$

Least squared error $\rightarrow$ minimize MSE

$L(f, (x, y)) = g(f(x), y)$ $g(x, y) = (x - y)^2$

loss function, prediction, response

$= f(x)$

MLE. $g(f(x), y) = \exp\{ \sim \}$

log - loss

Example : $\mathcal{F} = \{ f : f(x) = \alpha + x^T\beta,\ \alpha \in \mathbb{R}, \beta \in \mathbb{R}^d \}$ parametric method

$MSE_f = R(\alpha, \beta) = E[(\alpha + x^T\beta - y)^2]$

$= E[\alpha^2 + \boxed{(x^T\beta)^2} + y^2 + 2\alpha x^T\beta - 2\alpha y - 2x^T\beta y]$

$(E[x^T]\beta)^2$

$$\begin{cases} \frac{\partial R(\alpha, \beta)}{\partial \alpha} = 2E[\alpha] + 2E[x^T\beta] - 2E[y] = 0 \\ \frac{\partial R(\alpha, \beta)}{\partial \beta} = 2E[(x^T\beta) \cdot x] + E[2\alpha x] - E[2yx] = 0 \end{cases}$$

$\begin{cases} 1+d \text{ " = "} \\ 1+d \text{ param.} \end{cases}$

solve the linear system

$\Rightarrow \gamma = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \in \mathbb{R}^{1+d}$, $\beta \in \mathbb{R}^d$ $\beta = \begin{pmatrix} 1 \\ 2 \\ \vdots \end{pmatrix}$

Gamma

$Z = \begin{pmatrix} 1 \\ x \end{pmatrix} \in \mathbb{R}^{1+d}$ $\boxed{\alpha + \beta^T x}$ $\boxed{\gamma^T z}$

differentiating

$\Rightarrow E[zz^T]\gamma = E[zy]$

invertible $zz^T \in \mathbb{R}^{1+d \times 1+d}$ $zy \in \mathbb{R}^{1+d}$

$\Rightarrow \gamma = E[zz^T]^{-1} E[zy]$

$\sum_{i=1}^{n} (z_i^T\gamma - y_i)^2 = \| Z\gamma - Y \|_2^2$

$Z = \begin{pmatrix} z_1^T \\ z_2^T \\ \vdots \\ z_n^T \end{pmatrix} \in \mathbb{R}^{n \times 1+d}$

$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$

$\| Y \|_2 = \sqrt{y_1^2 + \cdots + y_n^2}$

$\frac{\partial \| Z\gamma - Y \|_2^2}{\partial \gamma} = 2 \cdot (Z\gamma - Y) \cdot Z$

$\| $

$(Z\gamma - Y)^T (Z\gamma - Y)$

$\downarrow$

$\mathbb{R}^n$

$= \boxed{(Z\gamma)^T \cdot (Z\gamma)} - 2(Z\gamma)^T \cdot Y + Y^T Y$

$\boxed{\gamma^T} \cdot Z^T Z \cdot \boxed{\gamma}$

$2Z^T Z\gamma - 2Z^T Y$

$= 2Z^T(Z\gamma - Y) = 0$

$Z^T Z\gamma = Z^T Y$ Wednesday, January 10, 2024 1:51 AM

$N_d(0, I_d)$ $x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$ x_i is a R.V. $\sim N(0,1)$
x_i i.i.d.

$\begin{pmatrix} 1 & & 0 \\ & 1 & \\ 0 & & 1 \end{pmatrix}$

$X \sim N_d(0, I_d)$ Covariance $\frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2} x_i^2\}$

pdf $f(x) = \left(\frac{1}{2\pi}\right)^{\frac{d}{2}} \cdot \exp\{ -\frac{1}{2} \sum_{i=1}^{d} x_i^2 \}$

$-\frac{1}{2} x^T I_d x = -\frac{1}{2} x^T x$

$x \in \mathbb{R}^d$ $Cov(X_1, X_1)$

$\Sigma = Cov(X, X) = \begin{bmatrix} \square_{(1,1)} & \square_{(i,j)} \\ & \end{bmatrix}$ $d \times d$

$Cov(X_i, X_j)$

$N_d(0, \Sigma)$

pdf $f(x) = \left(\frac{1}{2\pi \det(\Sigma)}\right)^{\frac{d}{2}} \exp\{ -\frac{1}{2\det(\Sigma)} x^T \Sigma x \}$

zero mean and Σ exists.

$z = \begin{pmatrix} 1 \\ x \end{pmatrix} \in \mathbb{R}^{1+d}$

$E[zz^T]$

$Cov(z, z) = \begin{bmatrix} 1 & 0 \\ 0 & \Sigma \end{bmatrix}$ $d \times d$ $Cov(1, x_i)$ $i = 1, \ldots, d$ $= 0$ $Cov(1, 1)$

$Cov(x, x)$

$Cov(z, z) = \boxed{\begin{bmatrix} 1 & 0 \\ 0 & \Sigma \end{bmatrix}}$

$E[zz^T] = \boxed{Cov(z, z)} + E[z]E[z^T]$

zero-mean

$\det(Cov(z, z)) = |\Sigma|$