Polynomial Regression

Friday, 16 August 2019 8:15 AM

In regression model,

$$y_i = \sum_{j=1}^{k} \alpha_i P_j(x_i) + \epsilon_i \qquad i = 1, \dots, n$$

$$\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

$$P_0(x_i) = 1 \quad \forall \ i$$

$$\sum_{i=1}^{n} P_j(x_i) P_k(x_i) = P_j^T P_k = 0 \quad \text{given the data} \quad \forall\, j \neq k$$

$$X_0 = \begin{pmatrix} P_0(x_1) & P_1(x_1) & \cdots & P_k(x_0) \\ P_0(x_2) & \cdots & \cdots & \cdots \\ \vdots & \vdots & & \\ P_0(x_n) & P_1(x_n) & \cdots & P_k(x_n) \end{pmatrix}$$

X_0 matrix for orthogonal polynomial.

$$\underset{\sim}{Y} = X_0 \underset{\sim}{\alpha} + \underset{\sim}{\epsilon} \qquad \underset{\sim}{\alpha} \in \mathbb{R}^{k+1}$$

$$\hat{\alpha} = (X_0^T X_0)^{-1} X_0^T \underset{\sim}{Y}$$

$$= \begin{pmatrix} \sum_1^n P_0^2(x_i) & & 0 \\ & \sum_1^n P_1^2(x_i) & \\ 0 & & \ddots \ \sum_1^n P_k^2(x_i) \end{pmatrix} X_0^T Y$$

$$\hat{\alpha}_j = \frac{P_j^T Y}{\sum_{i=1}^{n} P_j^2(x_i)} \qquad j = 0, \dots, k$$

$$\hat{\alpha}_j = \frac{\sum_{i=1}^{n} P_j(x_i) y_i}{\sum_{i=1}^{n} P_j^2(x_i)} \Rightarrow \hat{\alpha}_0 = \bar{y}$$

(Check)

$$= \frac{\sum_{i=1}^{n} P_0(x_i) y_i}{\sum_{i=1}^{n} P_0^2(x_i)} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad N(\mu, \sigma^2$$

Distribution of $\hat{\alpha}_j$

$$\underset{\sim}{Y} \sim N(X_0 \underset{\sim}{\alpha}, \sigma^2 I_n)$$

$$\hat{\alpha}_j = \frac{P_j^T Y}{\sum_{i=1}^{n} P_j^2(x_i)} = \frac{P_j^T Y}{P_j^T P_j} \sim \frac{N(P_j^T X_0 \alpha, \hat{\sigma}^2 P_j^T P_j)}{P_j^T P_j}$$

$$= N\left(\frac{P_j^T X_0 \alpha}{P_j^T P_j}, \frac{\sigma^2}{P_j^T P_j}\right)$$

$$= N\left(\alpha_j, \frac{\sigma^2}{P_j^T P_j}\right)$$

α_j's are unbiased estimators of α_j.

$$SS\,Res = \underset{\sim}{Y}^T (I_n - P_{X_0}) \underset{\sim}{Y}$$

$$= \underset{\sim}{Y}^T \underset{\sim}{Y} - \underset{\sim}{Y}^T P_{X_0} \underset{\sim}{Y}$$

$$= \sum_{i=1}^{n} Y_i^2 - \sum_{j=0}^{k} \hat{\alpha}_j \left(\sum_{i=1}^{n} P_j(x_i) y_i\right)$$

$$\boxed{Y^T P_j = \sum_{i=1}^{n} P_j(x_i) y_i \qquad \alpha_j = \frac{P_j^T Y}{P_j^T P_j}}$$

$$\boxed{P_{X_0} = X_0 (X_0^T X_0)^{-1} X_0^T}$$

Orthogonal Projection matrix for $C(X)$.

$$= \sum_{i=1}^{n} Y_i^2 - \hat{\alpha}_0 \sum_{i=1}^{n} P_0(x_i) y_i - \sum_{j=1}^{k} \hat{\alpha}_j \left(\sum_{i=1}^{n} P_j(x_i) y_j\right)$$

$$= \underbrace{\sum_{i=1}^{n} Y_i^2 - \hat{\alpha}_0 n \bar{y}}_{SS\,Total} - \underbrace{\sum_{j=1}^{k} \hat{\alpha}_j \left(\sum_{i=1}^{n} P_j(x)\right)}_{SS\,Model}$$

$$\Rightarrow SS\,Res = SS\,Total - SS\,Model$$

$$\Rightarrow \boxed{SS\,Total = SS\,Res + SS\,Model}$$

Test $H_0 = \alpha_j = 0$

vs $H_1 = \alpha_j \neq 0$

$$F = \frac{\hat{\alpha}_j \sum_{i=1}^{n} P_j(x_i) y_i}{\hat{\sigma}^2 / (n-k-1)} \sim F_{1, n-k-1}, \ ncp = 0$$

Under H_0

$$\sigma^2 = \frac{SS\,Error}{n-k-1}$$