Regression Analysis
Polynomial Regression

Buddhananda Banerjee

Department of Mathematics
Centre for Excellence in Artificial Intelligence
Indian Institute of Technology Kharagpur

bbanerjee@maths.iitkgp.ac.in

Multiple linear regression

- Consider a data set $D = \{(x_i, y_i) | x_i \in \mathbb{R}^1, y_i \in \mathbb{R}, \ \forall i = 1, 2, \cdots, n\}$
- x_is are non stochastic
- y_is are stochastic and realized values of random variable Y_is
- ϵ_is are iid $N(0, \sigma^2)$ random variables
- Regression parameter $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \cdots, \beta_k)^T$ is unknown
- Error parameter σ^2 is unknown

Problem statement

Considering the linear model

$$y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \cdots + \beta_k x_i^k + \epsilon_i, \ \forall \ i = 1, 2, \ldots n$$

we want to estimate $\boldsymbol{\beta}$ which will minimize least squared condition.

Polynomial regression

- We can extend the idea of multiple linear regression to polynomial regression.
- In polynomial regression we consider **higher degrees of the components x but it is linear in parameters.**
- Hence, it is a linear model too. For example,

$$\begin{aligned} y &= \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon \text{ for single regressor} \\ y &= \beta_0 + \beta_1 x_1 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \epsilon \text{ for multiple regressors} \end{aligned}$$

Matrix representation

In general for k-degree polynomial for single regressor , we can write in matrix notation

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \tag{1}$$

where, $\mathbf{Y} = (y_1, y_2, \cdots y_n)^T$, $\beta = (\beta_0, \beta_1, \beta_2, \cdots, \beta_k)^T$, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)^T$ with $\mathbf{x}_i = (1, x_i, x_i^2, \cdots, x_i^k)^T$ and $\epsilon = (\epsilon_1, \epsilon_2, \cdots \epsilon_n)^T \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$.

Hence there are $k+2$ unknown model parameters, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \cdots, \beta_k)^T$ and $\sigma^2 > 0$.

Least square solution

When $\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I}_n)$ the least square estimate of β will be

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y}.$$

Least square

Prediction

Fitted regression for the used data is

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y} = P_{\mathbf{X}}\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 P_{\mathbf{X}})$$

where $P_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T$ is the orthogonal projection matrix of the column space of $\mathbf{X}$ i.e. $\mathcal{C}(\mathbf{X})$. It means $\hat{\mathbf{y}} \in \mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{X}^T\mathbf{X})$.

Least square

Estimated Error

Hence the estimated error in prediction

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I}_n - P_{\mathbf{X}})\mathbf{y} \sim N(\mathbf{0}, \sigma^2(\mathbf{I}_n - P_{\mathbf{X}}))$$

where $\mathbf{e} \in \mathcal{C}(\mathbf{X})^{\perp} = \mathcal{C}(\mathbf{X}^T\mathbf{X})^{\perp}$.

Note: Hence $\hat{\mathbf{y}}$ and $\mathbf{e}$ are uncorrelated and they are independently distributed when $\epsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$.

Some Remarks

- **Finding order of the model**: For that we can go by either (a) forward selection or (b) backward elimination.
- **Extrapolation:** Beyond the range of the data the prediction may be more erroneous.
- **Ill-conditioning:** The matrix $(\mathbf{X}^T\mathbf{X})$ may be computationally singular, specially when the magnitude of th regressor is closed to zero.
- **Hierarchy:** A model with all lower order terms of the highest degree is called hierarchical model. But regression model need not be so. In design of experiment $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_5 x_1 x_2 + \epsilon$ might be sufficient considering only individual effect and interaction effect.

Orthogonal Polynomial

- For a single variable polynomial regression if we increase one more degree then we need to re-estimate all the coefficients including the new one each time.
- To overcome this problem we introduce the notion of orthogonal polynomial.

Definition

For a given set of input data $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ a set of polynomials $\{P_0, P_1, P_2, \cdots, P_k\}$ are said to be **orthogonal polynomials** if

$$P_0(x_i) = 1 \text{ and } \sum_{i=1}^{n} P_j(x_i) P_k(x_i) = 0 \ \forall j \neq k \tag{2}$$

Orthogonal Polynomial

Model

Now the model will look as follows

$$y_i = \sum_{j=0}^{k} \alpha_j P_j(x_i) + \epsilon_i \ \forall i = 1, 2, \cdots n \tag{3}$$

or denoting $\mathbf{Z} = ((P_j(x_i)))_{n \times (k+1)}$

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\alpha} + \epsilon$$

Estimation

When $\mathbf{Y} \sim N(\mathbf{Z}\boldsymbol{\alpha}, \sigma^2 \mathbf{I}_n)$ the least square estimate of $\boldsymbol{\alpha}$ will be

$$\hat{\boldsymbol{\alpha}} = (\mathbf{Z}^T\mathbf{Z})^{-}\mathbf{Z}^T\mathbf{y}.$$


THANK
YOU