CompSci, Oct 26, 2022

Linear Regression

- Data, input and output
  $D = \{(x_0, y_0), (x_1, y_1), \ldots (x_{n-1}, y_{n-1})\}$
- Model
- Cost/loss etc, the way we estimate the "quality" of the model.

Assumption

noise

$$y(x) = \underbrace{f(x)}_{\text{deterministic continuous function}} + \underset{N(0,\sigma^2)}{\underset{\sim}{\varepsilon}}$$

Model: $f(x) \simeq \tilde{y}(x)$

$$y(x_i) = y_i \simeq \tilde{y}(x_i) = \tilde{y}_i$$

Assume polynomial model

$$\tilde{y}_i = \sum_{j=0}^{p-1} \beta_j x_i^j$$

$$= \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_{p-1} x_i^{p-1}$$

$$p \le n$$

$\tilde{y}_i \rightarrow$ rewrite

$$\tilde{y}_0 = \beta_0 + \beta_1 x_0 + \beta_2 x_0^2 + \dots + \beta_{p-1} x_0^{p-1}$$

$$\tilde{y}_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \dots + \beta_{p-1} x_1^{p-1}$$

$$\vdots$$

$$\tilde{y}_{n-1} = \beta_0 + \beta_1 x_{n-1} + \beta_2 x_{n-1}^2 + \dots \beta_{p-1} x_{n-1}^{p-1}$$

$$\tilde{y}^T = [\tilde{y}_0 \ \tilde{y}_1 \ \dots \ \tilde{y}_{n-1}]$$

$$\tilde{y} \in \mathbb{R}^n$$

$$\beta^T = [\beta_0 \; \beta_1 \; \dots \; \beta_{p-1}]$$

$$\beta \in \mathbb{R}^p$$

$$\tilde{y} = X\beta$$ — linear in the unknown β

$$X = \begin{bmatrix} 1 & x_0 & x_0^2 & \dots & x_0^{p-1} \\ 1 & & & & \\ \vdots & & & & \\ 1 & x_{n-1} & x_{n-1}^2 & \dots & x_{n-1}^{p-1} \end{bmatrix}$$

$$= \begin{bmatrix} x_{00} & x_{01} & \dots & x_{0\,p-1} \\ x_{10} & & & \\ \vdots & & & \\ x_{n-1\,0} & \dots & & x_{n-1\,p-1} \end{bmatrix}$$

Design / Feature matrix

How to assess the model?

Cost function $C(\beta) = ?$

$$C(\beta) = |y - \tilde{y}|$$

$$= |y - X\beta| =$$

$$\|y - X\beta\|_1$$

$$\|x\|_1 = \sum_{i=0}^{n-1} |x_i|$$

$$C(\beta) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \tilde{y}_i)^2$$

$$= \frac{1}{n} \sum_{i=0}^{n-1} \left(y_i - \sum_{j=0}^{p-1} x_{ij} \beta_j\right)^2$$

$$= \frac{1}{n} \|(y - X\beta)\|_2^2$$

$$\|x\|_2 = \sqrt{\sum_{i=0}^{n-1} x_i^2}$$

$$\frac{\partial C(\beta)}{\partial \beta_j} = 0$$

$$\frac{\partial |x|}{\partial x} = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \end{cases}$$

$$MSE = C(\beta)$$

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^P}{\arg\min}\ C(\beta)$$

$$\frac{\partial}{\partial \beta}\left[\frac{1}{n}\sum_{i=0}^{n-1}\Big(y_i - \underbrace{x_{i*}\beta}_{\sum_{j=0}^{p-1} x_{ij}\beta_j}\Big)^2\right]$$

$$X \in \mathbb{R}^{n \times p}$$

$$\frac{\partial}{\partial \beta}\left(\frac{1}{n}\underbrace{(y - X\beta)}_{\in \mathbb{R}^n}{}^T \underbrace{(y - X\beta)}_{\in \mathbb{R}^n}\right)$$

$$\frac{\partial}{\partial \beta} C(\beta) = -\frac{2}{n} X^T (y - X\beta)$$

$$= 0$$

$$\Rightarrow X^T y = X^T X \beta \Rightarrow$$

$$\hat{\beta} = \underbrace{(X^T X)^{-1}}_{\in \mathbb{R}^{p \times p}} \underbrace{X^T}_{\mathbb{R}^{p \times n}} y \in \mathbb{R}^n$$

$$\hat{y} = X \hat{\beta} =$$

$$\underbrace{X (X^T X)^{-1} X^T}\, y$$

$$p \ll n$$