PS5841

Data Science in Finance & Insurance

Method of Least Squares

Yubo Wang

Spring 2022

Columbia University

Squared Error Loss for Prediction

- Find $f(\boldsymbol{X})$ for predicting Y given values of $\boldsymbol{X}$, where
  - Random input vector $\boldsymbol{X} \in \mathcal{R}^p$
  - Random output variable $Y \in \mathcal{R}$
- Squared error loss

$$Loss = [Y - f(\boldsymbol{X})]^2$$

- Expected (squared) prediction error

$$\begin{aligned} EPE(f) &= E([Y - f(\boldsymbol{X})]^2) \\ &= E_{\boldsymbol{X}} E_{Y|\boldsymbol{X}}([Y - f(\boldsymbol{X})]^2 | \boldsymbol{X}) \end{aligned}$$

COLUMBIA UNIVERSITY

Regression Function

- Expected (squared) prediction error

$$EPE(f) = E([Y - f(\boldsymbol{X})]^2)$$
$$= E_{\boldsymbol{X}} E_{Y|\boldsymbol{X}}([Y - f(\boldsymbol{X})]^2 | \boldsymbol{X})$$

- Minimize EPE pointwise

$$f(\boldsymbol{x}) = \underset{c}{\operatorname{argmin}}\, E_{Y|\boldsymbol{X}}([Y - c]^2 | \boldsymbol{X} = \boldsymbol{x})$$

- Regression function

$$f(\boldsymbol{x}) = E(Y | X = \boldsymbol{x})$$

COLUMBIA UNIVERSITY

Basic/Simple Linear (Regression) Model

- Training set of size n: $\{(x_i, y_i)\}$
  - y_i is the observed value of Y_i
  - The Y_is are independent
- Regression function

$$E(Y|X = x) = \beta_0 + \beta_1 x$$

COLUMBIA UNIVERSITY

Fitted by the Method of Least Squares

- Assume $Var(Y_i) = \sigma^2 \; \forall \, i$ for now
- Minimize loss $\sum_i (y_i - \beta_0 - \beta_1 x_i)^2$
- Estimated parameters

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \sum_i \omega_i y_i$$

$$\omega_i = \frac{x_i - \bar{x}}{(n-1)s_X^2} \rightarrow \sum_i \omega_i = 0$$

$$s_X^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$$

- Fitted model $\hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$

COLUMBIA UNIVERSITY

Useful Results

- $\bar{\hat{y}} = \bar{y}$
- $\sum_i \hat{e}_i = 0, \;\; \hat{e}_i = \hat{y}_i - y_i$
- $\sum_i x_i \hat{e}_i = 0$
- TSS = ESS + RSS

$$\sum_i (y_i - \bar{y})^2 = \sum_i (\hat{y}_i - y_i)^2 + \sum_i (\hat{y}_i - \bar{y})^2$$

- Estimating σ^2: $s^2 = \frac{1}{n-2} \sum_i \hat{e}_i^2$
- Coefficient of Determination: $R^2 = \frac{RSS}{TSS}$
- ANOVA Table: keep track of variability

COLUMBIA UNIVERSITY

Multiple Linear (Regression) Model

- Training set of size n: $\{(\boldsymbol{x}_i, y_i)\}$
  - y_i is the observed value of Y_i
  - The Y_is are independent
  - $Var(\boldsymbol{Y}) = V$
- Regression function

$$E(Y|\boldsymbol{X} = \boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}$$

to be continued

Columbia University

Fitted by the Method of Least Squares (1)

- Minimize loss

$$\sum_i \frac{1}{\sigma_i^2}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)^2$$

$$= (\boldsymbol{y} - \boldsymbol{\mu})^T V^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$$

  – Matrix of 2nd derivatives positive definite

  – Solutions of the normal equations vs local minima at parameter space boundaries

COLUMBIA UNIVERSITY

Fitted by the Method of Least Squares (2)

- Estimated parameters (if has intercept X_0)

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

$$X = \begin{pmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{pmatrix} = (X_0 \quad X_1 \quad \cdots \quad X_k)$$

- If $V = Var(\boldsymbol{Y}) = \sigma^2 I$

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \boldsymbol{y} = \sum_i \boldsymbol{\omega}_i y_i$$

$$\boldsymbol{\omega}_i = (X^T X)^{-1} (1, x_{i1}, \dots, x_{ik})^T$$

- Fitted model $\hat{\boldsymbol{y}}(\boldsymbol{x}) = \boldsymbol{x}^T \hat{\beta}$

COLUMBIA UNIVERSITY

Useful Results (1)

When $E(\boldsymbol{Y}) = X\boldsymbol{\beta}, Var(\boldsymbol{Y}) = \sigma^2 I, Y_i$ s are independent RV

- $E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$
- $Var(\widehat{\boldsymbol{\beta}}) = \sigma^2 (X^T X)^{-1}$
- LSE is BLUE (Gauss-Markov)
- $s^2 = \hat{\sigma}^2 = \frac{1}{n-p}(\boldsymbol{y} - X\widehat{\boldsymbol{\beta}})^T(\boldsymbol{y} - X\widehat{\boldsymbol{\beta}})$

COLUMBIA UNIVERSITY

Useful Results (2)

- Hat Matrix $H = X(X^T X)^{-1} X, \;\; \widehat{\boldsymbol{y}} = H(\boldsymbol{y})$
$$H^T = H, \qquad HH = H$$
$$HX = X, \qquad HX_j = X_j$$
- $\widehat{\boldsymbol{e}}^T X_j = 0, \widehat{\boldsymbol{e}} = (I - H)\boldsymbol{y}$
- $\boldsymbol{y}^T \widehat{\boldsymbol{y}} = \widehat{\boldsymbol{y}}^T \widehat{\boldsymbol{y}}$
- TSS = ESS + RSS
$$TSS = \boldsymbol{y}^T \boldsymbol{y} - n(\bar{y})^2$$
$$ESS = \boldsymbol{y}^T \boldsymbol{y} - \boldsymbol{y}^T \widehat{\boldsymbol{y}}$$
$$RSS = \widehat{\boldsymbol{y}}^T \widehat{\boldsymbol{y}} - n(\bar{y})^2$$
- Coefficient of Determination: $R^2 = \frac{RSS}{TSS} = \left(r_{\boldsymbol{y},\widehat{\boldsymbol{y}}}\right)^2$
- ANOVA Table: keep track of variability

COLUMBIA UNIVERSITY

Application

- Equity beta
  - Market model
  - CAPM
- Demand for term life insurance
  - Family characteristics that influence the amount of insurance purchased

Columbia University

That was

COLUMBIA UNIVERSITY