AI-HW

Nivedita Daiya.

1) In a linear regression model using OLS identify the effect of correlated features on the matrix computation. (through Derivation)

→ Multiple linear regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

$\varepsilon \rightarrow$ residual terms of the model.

$\beta_0, \beta_1, \beta_2, \dots, \beta_k \rightarrow$ regression coefficients.

Interaction with 2 or more variables.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon \rightarrow (1)$$

This is called linear regression model as they can be called as linear combination of the β-parameters in the model.

$x \rightarrow$ weights and they can be non-linear

So the matrix form for linear regression is:-

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{21} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{2n} & \cdots & x_{kn} \end{bmatrix}_{n\times k} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}_{k\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}$$

Criteria for Estimates

→ Select such $\hat{\beta}$ that it minimizes the sum of squared residuals. (

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

Sum of squared residual (RSS) $= e'e^2$.

$$[e_1 \; \varepsilon_2 \; \ldots \ldots \; \varepsilon_n]_{1\times n} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1} = [\varepsilon_1 \times \varepsilon_1 + \varepsilon_2 \times \varepsilon_2 + \cdots + \varepsilon_n \times \varepsilon_n]_{1\times 1} \rightarrow (3)$$

Sum of squared residuals are

$$e'e = (y - X\hat{\beta})'(y - X\hat{\beta})$$
$$= y'y - \hat{\beta}'X'y - y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$$
$$= y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} \rightarrow (4)$$

Derivate of eq(4) with respect to $\hat{\beta}$.

$$\frac{\partial e'\varepsilon}{\partial \hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0 \rightarrow (5)$$

$$\frac{\partial \varepsilon'\varepsilon}{\partial \hat{\beta}} = +2X'X \rightarrow (6)$$

classmate

Date

Page

Matrix Differentiation

$$\frac{\partial a'b}{\partial b} = \frac{\partial b'a}{\partial b} = a. \rightarrow (6)$$

a and b are $K\times1$ vectors.

$$\frac{\partial b'Ab}{\partial b} = 2Ab = 2b'A \rightarrow (7).$$

A - symmetric matrix.

$$\frac{\partial 2\beta'x'y}{\partial b} = \frac{\partial 2\beta'(x'y)}{\partial b} = 2x'y \rightarrow (8).$$

$$\frac{\partial \beta'x'x\beta}{\partial b} = \frac{\partial \beta'A\beta}{\partial b} = 2A\beta = 2x'\cancel{A}\beta - (9)$$

$x'\cancel{x}X = K\times K$ matrix

Normal equation from eqn 5

$$(x'x)\hat{\beta} = x'y. \rightarrow (10)$$

Multiplying both the sides by $(x'x)^{-1}$.

$$(x'x)^{-1}(x'x)\hat{\beta} = x'y.(x'x)^{-1} \rightarrow (11).$$

$$I.\hat{\beta} = (x'x)^{-1}.x'.y.$$ $\because (x'x)^{-1}(x'x) = I.$

$\because I =$ Identity matrix

u

$$\hat{\beta} = (x'x)^{-1}x'y \rightarrow (12)$$

* Properties of the OLS Estimators

(1) Normal form eqn from Eq (10)

$$(x'x)\hat{\beta} = x'y \rightarrow (13)$$

$$(x'x)\hat{\beta} = x'(x\hat{\beta} + \varepsilon)$$

∴ Substitute $y = x\hat{\beta} + \varepsilon$.

$$(x'x)\hat{\beta} = (x'x)\hat{\beta} + x'\varepsilon$$

$$\boxed{x'\varepsilon = 0} \rightarrow (14)$$

Matrix.

$$\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ X_{k1} & X_{k2} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} X_{11} \times \varepsilon_1 + X_{12}\varepsilon_2 + \cdots + X_{1n} \times \varepsilon_n \\ X_{21} \times \varepsilon_1 + X_{22}\varepsilon_2 + \cdots + \varepsilon_{2n} \times \varepsilon_n \\ \vdots \\ X_{k1} \times \varepsilon_1 + X_{k2}\varepsilon_2 + \cdots + X_{kn}\varepsilon_n \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \rightarrow (15)$$

(1) Observed value of x are uncorrelated with the residual.

(2) The sum of the residuals are zero
(3) Sample mean of the residual are zero.
(4) The regression hyperplane passes through the means of the observed values.

(5) The predicted values of y are uncorrelated with the residuals.

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

$$\hat{y}'\varepsilon = (X\hat{\beta})'\varepsilon = b'X'e = 0 \rightarrow (16)$$

$y = X\beta + \varepsilon$
↳ Assumption states that there is a linear relationship between y and x.

$X = n \times k$
X - linear independent

$E[\varepsilon | X] = 0$.

$$E\begin{bmatrix} \varepsilon_1 | X \\ \varepsilon_2 | X \\ \vdots \\ \varepsilon_n | X \end{bmatrix} = \begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \vdots \\ E(\varepsilon_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \rightarrow (17)$$

$E(\varepsilon\varepsilon'|X) = \sigma^2 I$

$$E(\varepsilon\varepsilon'|X) = E\begin{bmatrix} \varepsilon_1|X \\ \varepsilon_2|X \\ \vdots \\ \varepsilon_n|X \end{bmatrix} [\varepsilon_1|X \quad \varepsilon_2|X \cdots \varepsilon_n|X] \rightarrow (19)$$

$$E(\varepsilon\varepsilon'|X) = E\begin{bmatrix} \varepsilon_1^2|X & \varepsilon_1\varepsilon_2|X & \cdots & \varepsilon_1\varepsilon_n|X \\ \varepsilon_2\varepsilon_1|X & \varepsilon_2^2|X & \cdots & \varepsilon_2\varepsilon_n|X \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \varepsilon_k\varepsilon_1|X & \varepsilon_k\varepsilon_2|X & \cdots & \varepsilon_k\varepsilon_n|X \end{bmatrix}$$

$$E(\varepsilon\varepsilon'|X) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \rightarrow 21.$$

$$E(\varepsilon\varepsilon'|X) = \sigma^2\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \sigma^2 I \rightarrow (22)$$

$\Omega = \sigma^2 I$

↳ Variance-Covariance matrix.