D048

11:30 am
Thu, Apr 24

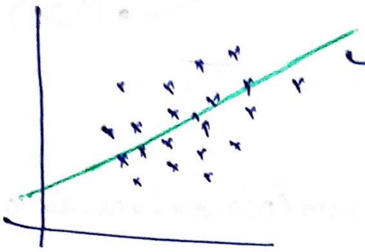
Linear Regression

Simple | multiple | polynomial | Regulization

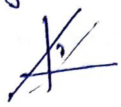
$Y = mx + b$

Intercept (b), slope (m)

m is the weightage of x on y



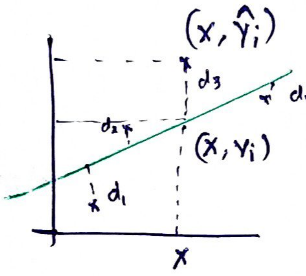
m ↑



D049

(m, b)

- closed form solution — Direct formula — Ordinal least square OLS
- Non-closed form — Gradient Desent



prediction in ^ (Hat)

$J = d_1 + d_2 + d_3 \cdots + d_n$

$J = d_1^2 + d_2^2 + \cdots + d_n^2$ (all dist +ve)

$$J = \sum_{i=1}^{n} d_i^2$$

Error func, loss func

$b = \bar{Y} - m\bar{X}$ $\quad \bar{X}, \bar{Y}$ mean

$$m = \frac{\sum_{i=1}^{n} (x_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$d_i = (Y_i - \hat{Y}_i)^2$

Total error

$$J = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

we know $\hat{Y}_i = mX_i + b$

$$J(m,b) = \sum_{i=1}^{n} (Y_i - mx_i - b)^2$$

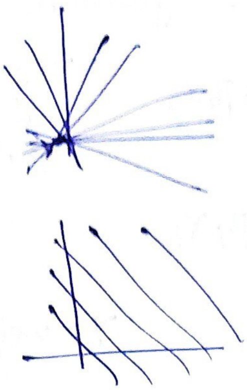
if $b =$

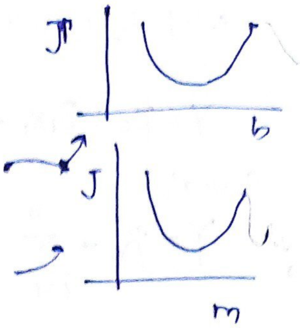
if $b=0$

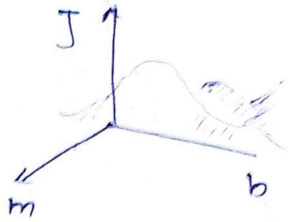
$$J(m) = \sum_{i=1}^{n} (y_i - m x_i)^2$$

if $m=1$

$$J(b) = \sum_{i=1}^{n} (y_i - x_i - b)^2$$







$$J(m,b) = 0$$

$$\frac{\partial J}{\partial m} = 0 \ \& \ \frac{\partial J}{\partial b} = 0$$

$$\frac{\partial J}{\partial b} = \frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - m x_i - b)^2 = 0$$

$$\Rightarrow \sum \frac{\partial}{\partial b} (y_i - m x_i - b)^2 = 0$$

chain Rule

$$\Rightarrow \sum -2(y_i - m x_i - b) = 0$$

(divide by -2)

$$\Rightarrow \sum (y_i - m x_i - b) = 0$$

divide by n

$$\Rightarrow \frac{\sum y_i}{n} - \frac{\sum m x_i}{n} - \frac{\sum b}{n} = 0$$

$$\Rightarrow \bar{y} - m\bar{x} - \frac{nb}{n} = 0$$

$$\Rightarrow \bar{y} - m\bar{x} = b$$

$$\Rightarrow b = \bar{y} - m\bar{x}$$

$J \cancel{= \Sigma} (v$

$$J = \sum (y_i - m x_i - \bar{y} + m\bar{x})^2$$

$$\Rightarrow \frac{\partial E}{\partial m} = \sum \frac{\partial}{\partial m} (y_i - m x_i - \bar{y} + m\bar{x})^2 = 0$$

$$\Rightarrow \sum 2(y_i - m x_i - \bar{y} + m\bar{x})(-x_i + \bar{x}) = 0$$

$$\Rightarrow \sum -2(y_i - m x_i - \bar{y} + m\bar{x})(x_i - \bar{x}) = 0$$

$$\Rightarrow \sum (y_i - m x_i - \bar{y} + m\bar{x})(x_i - \bar{x}) = 0$$

$$\Rightarrow \sum \left[(y_i - \bar{y}) - m(x_i - \bar{x})\right](x_i - \bar{x}) = 0$$

$$\Rightarrow \sum \left[(y_i - \bar{y})(x_i - \bar{x}) - m(x_i - \bar{x})^2\right] = 0$$

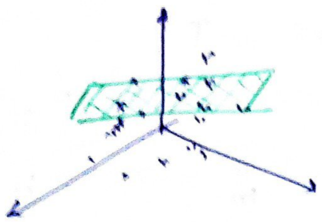
$$\Rightarrow \sum (y_i - \bar{y})(x_i - \bar{x}) = m \sum (x_i - \bar{x})^2$$

$$\Rightarrow m = \cancel{\sum_{i=1}^{n}} \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{(x_i - \bar{x})^2}$$

12:33 pm

Multiple Linear Regression

D050-2



$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

$y = (m)x + b$

$m \to \beta_1, \beta_2$; $b \to \beta_0$

General

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i$$

$X_1 \to cgpa$

$X_2 - iq$

$Y = lpa$

$Y = weight + \beta_1 lpa + \beta_2 iq$

Math

X_1	X_2	X_3	Y
x_{11}	x_{12}	x_{13}	x
x_{21}	x_{22}	x_{23}	

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

$$\hat{Y} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_{100} \end{bmatrix} = \begin{bmatrix} \beta_0 & \beta_1 X_{11} & \beta_2 X_{12} & \beta_3 X_{13} \\ \beta_0 & \beta_1 X_{21} & \beta_2 X_{22} & \beta_3 X_{23} \\ \vdots & & \vdots & \\ \beta_0 & \beta_1 X_{m1} & \beta_2 X_{1002} & \beta_{1003} \end{bmatrix}$$

100 rows = n rows

3 cols = m cols

$$\hat{Y} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} \beta_0 & \beta_1 X_{11} & \beta_1 X_{12} & \beta_1 X_{13} & \cdots & \beta_m X_{1m} \\ \beta_0 & \beta_1 X_{21} & \beta_1 X_{22} & \beta_1 X_{23} & \cdots & \beta_m X_{2m} \\ \vdots & & & & & \\ \beta_0 & \beta_1 X_{n1} & \beta_1 X_{n2} & \beta_1 X_{n3} & \cdots & \beta_m X_{nm} \end{bmatrix}$$

$$= \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & x_{23} & \cdots & x_{2m} \\ \vdots & & & & & \\ 1 & x_{n1} & x_{n2} & x_{n3} & \cdots & x_{nm} \end{bmatrix}$$

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

2D $\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix}$

3D $\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$

$$\begin{array}{c|c|c|c} & X_1 & X_2 & X_3 \\ 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ 1 & x_{31} & x_{32} & x_{33} \end{array}$$

(1 I/p dataset)

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

already have

$$e = Y - \hat{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} - \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} Y_1 - \hat{Y}_1 \\ Y_2 - \hat{Y}_2 \\ \vdots \\ Y_n - \hat{Y}_n \end{bmatrix}$$

in single LR

$$J = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$J = e^T e$$

$$J = \underset{1\times n}{[(Y_1-\hat{Y}_1)(Y_2-\hat{Y}_2)\cdots(Y_n-\hat{Y}_n)]} \underset{n\times 1}{\begin{bmatrix} (Y_1-\hat{Y}_1) \\ (Y_2-\hat{Y}_2) \\ \vdots \\ (Y_n-\hat{Y}_n) \end{bmatrix}}$$

$$= \underset{1\times 1}{\left[(Y_1-\hat{Y}_1)^2 + (Y_2-\hat{Y}_2)^2 + (Y_3-\hat{Y}_3)^2 \cdots (Y_n-\hat{Y}_n)^2\right]}$$

$$= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$(A+B)^T = A^T + B^T$

$(A-B)^T = A^T - B^T$

$$J = e^T e$$
$$= (Y-\hat{Y})^T (Y-\hat{Y})$$
$$= (Y^T - \hat{Y}^T)(Y-\hat{Y}) \qquad [\because \hat{Y} = X\beta]$$
$$= (Y^T - (X\beta)^T)(Y - X\beta)$$
$$= Y^TY - Y^TX\beta - (X\beta)^TY + (X\beta)^TX\beta$$
$$= Y^TY - 2Y^TX\beta + \beta^TX^T\beta X$$

let $Y = A$, $X\beta = B$

$Y^TX\beta$, $(X\beta)^TY$

$A^TB = B^TA$ — (i)

$(A^TB)^T = B^TA$ — (ii)

$(AB)^T = B^TA^T$

$(A^T)^T = A$

let $A^TB = C$

from i, ii

$A^TB = (A^TB)^T$

$C = C^T$

row x col

$$\underset{(1\times n)}{Y^T}\ \underset{(n\times(m+1))}{X}\ \underset{((m+1)\times 1)}{\beta}$$

$$(Y^TX\beta)^T = Y^TX\beta$$

$$\frac{dJ}{d\beta} = \frac{d}{d\beta}\left[Y^TY - 2Y^TX\beta + \beta^TX^TX\beta\right]$$

$$\Rightarrow = 0 - 2Y^TX\beta + \frac{d}{d\beta}\left[\beta^TX^TX\beta\right] = 0$$

$$= -2Y^TX + 2X^TX\beta^T = 0$$

$$\Rightarrow \cancel{2}Y^TX = \cancel{2}X^TX\beta^T$$

$$\Rightarrow \beta^T = Y^TX(X^TX)^{-1}$$

$$\Rightarrow (\beta^T)^T = \left[Y^TX\,(X^TX)^{-1}\right]^T$$

$$\Rightarrow \beta = \left[(X^TX)^{-1}\right]^T \left[Y^TX\right]^T$$

$$= (X^TX)^{-1}\, X^TY$$

<u>matrix diff.</u>

$$\omega = A^TXA$$

$$\frac{d}{dA}\omega = 2XA^T$$

$$X^TX$$

$$(m+1)\times \cancel{n} \times \cancel{n}\times(m+1)$$

$$(m+1)\cdot(m+1)$$

sq. matrix inverse also sq

$$\beta = (X^T X)^{-1} X^T Y$$

$$[(m+1)(m+1)]\,[(m+1)\times n]\,[n\times 1]$$

$$[(m+1)\times n]\,[n\times 1]$$

$$(m+1)\times 1$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}$$

$(m+1)\times 1$

why Gradient Desent

Computational complexity of mathematical operations (wiki)

Inverse matrix $(n\times n)$

Gaus-Jordhan $O(n^3)$

if we have 100 colums then $(X^T X)^{-1}$ 100×100 10000

n^{10000}

Polynomial Regression

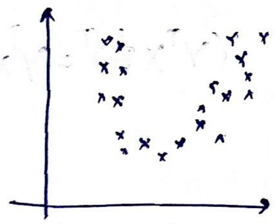
10:50 am
FRi, Apr 25

simple LR

$$Y = \beta_0 + \beta_1 X$$

multiple LR

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_m$$



How mush (Hyper parameter)

degree = 3 → x^0, x^1, x^2, x^3, y

X	Y
5	2

(1, 2)

degree = 2 polynomial

Y	x^0	x^1	x^2
2	5^0	5^1	5^2

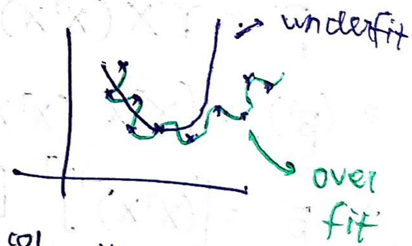
(1, 4)

Train

$$\hat{Y} = \beta_0 + \beta_1 X + \beta_2 X^2$$

simple polynomial linear regression

we found β so it called LR



if we have two I/P col x_1 x_2 Y

degree = 2

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2^1 + \beta_4 X_2^2$$

d = 3

$$\hat{Y} = \beta_0 + \beta_1 X_1^1 + \beta_2 X_1^2 + \beta_3 X_1^3 + \beta_4 X_2^1 + \beta_5 X_2^2 + \beta_6 X_2^3 + \beta_7 X_3^1 + \beta_8 X_3^2 + \beta_9 X_3^3$$