Linear Regression

$$y_{n\times 1} = X_{n\times(p+1)}\beta_{(p+1)\times 1} + \varepsilon_{n\times 1}$$

$$= \begin{pmatrix} \beta_0 + x_{11}\beta_1 + x_{12}\beta_2 + \dots + \varepsilon_1 \\ \dots \\ \dots \end{pmatrix}$$

$$\underset{n\times 1}{\begin{pmatrix} y \end{pmatrix}} = \underset{n\times(p+1)}{\begin{pmatrix} X \end{pmatrix}} \underset{(p+1)\times 1}{\begin{pmatrix} \beta \end{pmatrix}} + \underset{n\times 1}{\begin{pmatrix} \varepsilon \end{pmatrix}}$$

Large n small p

$$\begin{pmatrix} \\ \\ \end{pmatrix} = \begin{pmatrix} \quad & \quad \\ \\ \\ \end{pmatrix}\begin{pmatrix} \\ \\ \end{pmatrix} + \varepsilon$$

or Large p small n

$$\Big(\ \Big) = \Big(\qquad\qquad\Big)\Big(\ \Big) + \varepsilon$$

For Linear regression use LS

$$\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$$

$$= \underbrace{\sum_{i=1}^{n}\overbrace{\left(y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip}\right)^2}^{\text{squared loss for individual}}}_{\text{total loss for all predictions}}$$

$$\min_{\beta} \|y - X\beta\|_2^2$$

$$\frac{\partial \|y - X\beta\|^2}{\partial \beta}$$

$$\min_{\beta} (b - a \cdot \beta)^2$$

$$f(\beta) = (b - a \cdot \beta)^2$$

$$0 = f'(\beta) = 2(b - a\beta)(-a)$$

$$\Rightarrow b = a\beta$$

$$\Rightarrow a^{-}\beta = \beta^{*} \longrightarrow \text{Least Squares solution}$$

for vectors

$$f(\beta) = \|\underbrace{y}_{n \times 1} - \underbrace{X}_{n \times (p+1)} \underbrace{\beta}_{(p+1) \times 1}\|^2_{L2\ Norm}$$

$$\begin{matrix} \frac{\partial f}{\partial \beta_1} \\ \frac{\partial f}{\partial \beta_2} \\ \vdots \\ \frac{\partial f}{\partial \beta_{p+1}} \end{matrix} \quad \vec{0} = \frac{\partial}{\partial \beta} f(\beta) = 2(\underbrace{y - X\beta}_{n \times 1})(\underbrace{-X}_{n \times (p+1)})$$

$$0 = -2X^T(y - X\beta)$$ Now dims match

$$0 = X^T(y - X\beta)$$ Divide out -2

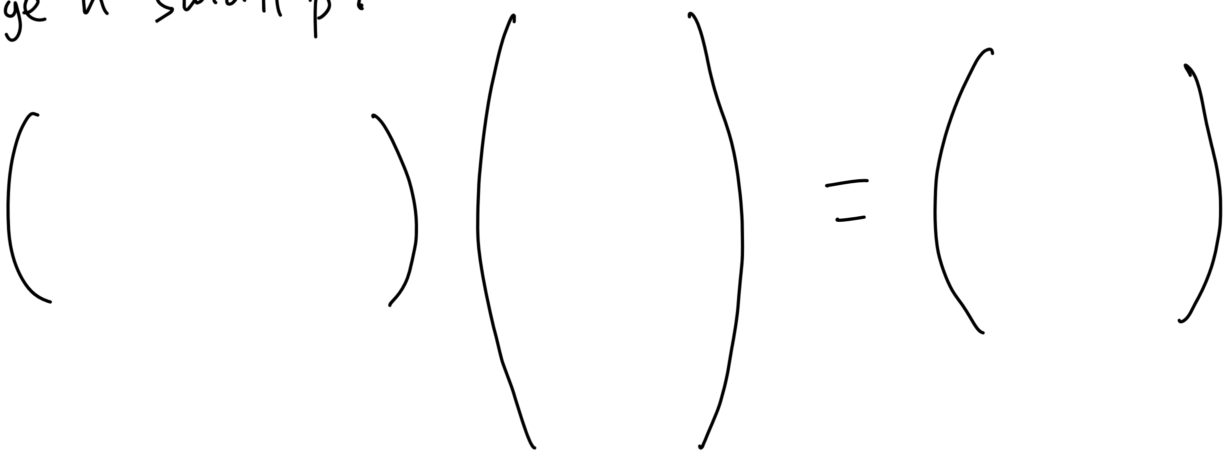
$$= X^T y - (X^T X)\beta$$ LS solution for matrices

$$X^T y = X^T X \beta \Longrightarrow \beta^{*} = (X^T X)^{-1} X^T y$$

assuming we can take inverse of $X^T X$

$$X_{n\times(p+1)} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & & & \\ 1 & & & \\ \vdots & & & \\ 1 & & & \end{pmatrix}$$

Large n small p:



Square matrix

df of residuals is $n-(p+1) \Rightarrow$ df(residua) = (sample size) − (# of linear coefficients)

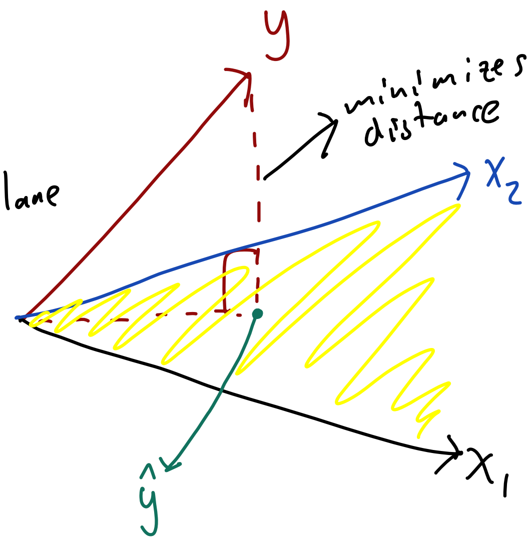
$(y - X\hat{\beta})$ is residual vector $= \gamma_{n\times 1}$

so $$0 = X^T(y - X\hat{\beta}) = \underset{(p+1)\times n}{X^T}\, \gamma_{n\times 1}$$

so dot product of X^T & residuals is equal to 0 & sum of residuals must be 0 for this to work

$$\begin{pmatrix} y \end{pmatrix} = \begin{pmatrix} \vdots & \vdots \\ \vdots & \vdots \end{pmatrix} + \varepsilon$$
$$\quad\quad\quad\quad X_1 \quad X_2$$

projection of y onto $X_1 X_2$ plane



Find β_1 & β_2 such that

$$\beta_1 X_1 + \beta_2 X_2 = \hat{y}$$

equivalent to finding vector v from $C(X)$ that minimizes $\|y - v\|^2$

R^2 is how well model fits data, R^2 is also correlation between y & $\hat{y}$

For Rank deficiency:

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

in R:

$$\min \|\beta\|_0 \quad \text{subject to} \quad \hat{y} = X\beta \qquad \left\| \begin{pmatrix} \beta_1 \\ \cdots \\ \beta_p \end{pmatrix} \right\|_0$$

Find β with most predictors as 0s.
So finding the simplest model that fits the LS solution. Labels redudant columns as N/a coeff

in Python:

$$\min \|\beta\|_2^2 \quad \text{subject to} \quad \hat{y} = X\beta$$

Python takes L_2 Norm instead so it doesn't look for simplest model

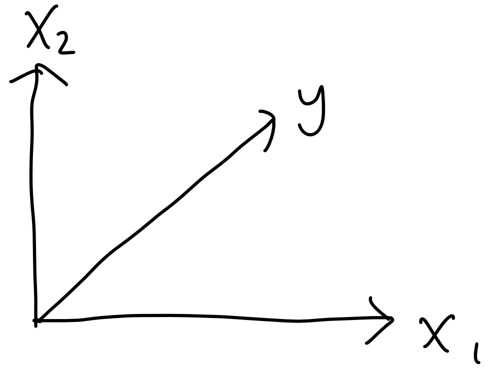
Partial Regression Coefficients:

M_1: $y \approx \beta_0 + \beta_1 X_1$ $\rightarrow$ marginal regression between X_1 & y

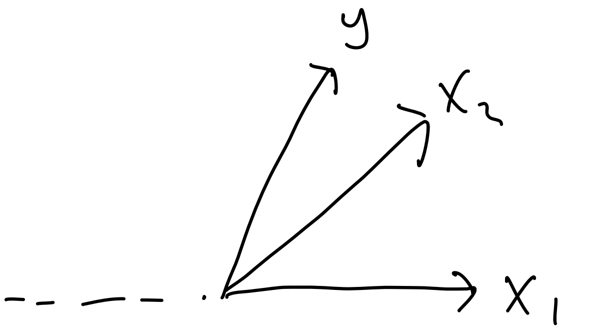
M_2: $y \approx \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2$

$$(\ ) \underset{\sim}{\approx} \left(\begin{matrix} 1 \\ \vdots \\ 1 \end{matrix} \ X_1 \ X_2 \right)$$

consider no intercept

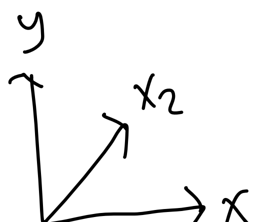


Both coeff positive



Need X_1 coeff to be negative because X_2 changed

it is used to offset X_2



Here X_1 is perpendicular to y so it is marginally irrelevant to predicting y but when we introduce X_2 it becomes useful