Linear Regression

$f(x) = \hat{y} = w_1 x_1 + w_2 x_2 \dots w_d x_d + w_0$

↳ weights

look at weight prediction ↑↓ with increase in feature

compare magnitude for feature imp. (scale the features)

normalise → faster GD

Gradients

d features

$w_1 \rightarrow -2(y^i - \hat{y}^i) x^i_1$

w_2

⋮

$w_d \rightarrow -2(y^i - \hat{y}) x^i_d$

$w_0 \rightarrow (-2)(y - \hat{y})$

f_0: 1, 1, 1, …, 1 | d | y

$f(x)$

$= w_0 x_0 + w_1 x_1 + w_2 x_2 \dots w_d x_d$

$x_0 \rightarrow 1$

$f_d \rightarrow i \rightarrow (-2)(y - \hat{y}) x^i_d$

$f_{d=0} \rightarrow i \rightarrow \underbrace{(-2)(y - \hat{y})}_{w_0}$ $\quad x^i_0 \rightarrow 1$

R2 score $= 1 - \dfrac{SSR_{model}}{SSR_{total/avg}}$

$SSR_{model} \rightarrow \Sigma(y^i - \hat{y}^i)^2$

$SSR_{total/avg} \rightarrow \Sigma(y^i - \bar{y}^i)^2$

$-\infty$ ……… 0 ……… 1

worse than avg model | avg model | Best

$d \rightarrow d+1$ weights

$w_0 x_0 + w_1 x_1 \dots\dots w_d x_d + w_{d+1} x_{d+1}$

R2 score ↑↑ ✓ | ↓↓ ✗ (very unlikely)

$d \rightarrow d+1$

$w\ d+1 \rightarrow d+2$

$(-\infty, 1]$

Adj R2 score $= 1 - \left[\dfrac{(1 - R2)(m - 1)}{(m - d - 1)}\right]$

$R2 = 1 - \dfrac{SSR_{model}}{SSR_{avg}}$

m → # of data points

d → dimensions

Stats model

↳ LR → OLS (ordinary least squares)

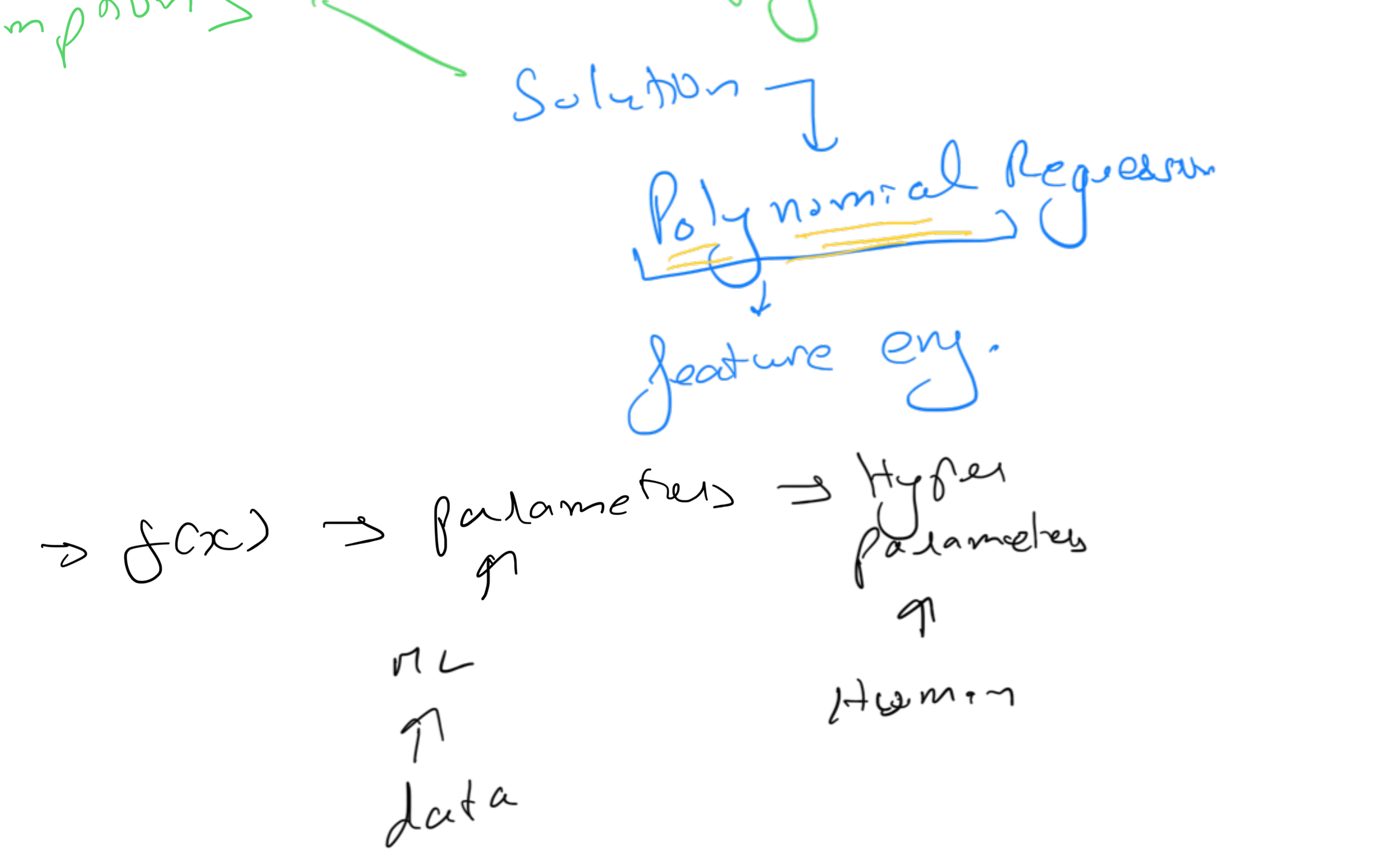
$w^* \rightarrow \min$ (sq of residual)

↳ MSE

OLS	Sklearn
summary.	model at producting grade efficiency.
too much details	gives an overview
add a const for intercept	more extra possible functions ↳ sklearn.metrics

Assumption in Linear Regression

→ Assumption of Linearity



Assumptions

- Performance doesnot directly go to 0
- Solution → Polynomial Regression → feature eng.

Model → $f(x)$ → parameters → Hyper parameters

parameters ↑ ML ↑ data

Hyper parameters ↑ Human

$y = mx + c$

① Assumption of linearity

② No multi-colinearity

$f_1\ f_2\ f_3 \dots \rightarrow T$

$T \rightarrow af_1 + bf_2 + cf_3 + d$

$f_1 \neq bf_2 + cf_3 + d$

③ Residuals should be normal.

④ No heteroskedacity

⑤ No auto-corelation.

① linear.