1. $\hat{y} = \beta_0 + \beta_1 x$

$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$\hat{y} = \beta_0 + \beta_1 x^2$

$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$

5. $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^3$

Linear regression (simple)

Model: $\underbrace{\hat{y}}_{\text{cont. value}} = \beta_0 + \beta_1 x$

Params: β_0, β_1

Cost func: $E(\beta_0, \beta_1) = \sum_{i=1}^{N} \left[(\beta_0 + \beta_1 x_i) - y_i \right]^2$

no. of samples

Goal: $\min E(\beta_0, \beta_1)$

find $\frac{dE}{d\beta_0}$ & $\frac{dE}{d\beta_1}$ expression

Solve :

① Analytical : $\left.\begin{array}{l} \frac{dE}{\beta_0} = 0 \\ \frac{dE}{d\beta_1} = 0 \end{array}\right\}$ solve for β_0, β_1

The normal eqn

② gradient desc : find local min

$$\beta_{new} = \beta_{old} - \alpha \frac{dE}{d\beta}$$

constant

column vec

$$\beta = \beta - \alpha c X^T (X\beta - y)$$

sample by # feat

rows columns

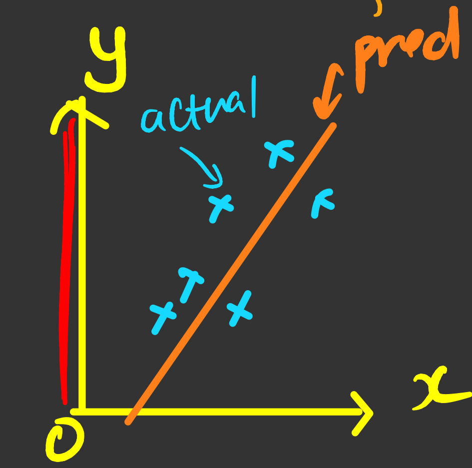
* since sq. E is a convex fn, then local min $\equiv$ global min

Logistic Reg : Classification

logistic function : $\frac{1}{1+e^{-(\beta_0+\beta_1 x)}} = \hat{p}$
("map")

$\underbrace{\beta_0+\beta_1 x}_{\text{for SLR : } \hat{y}}$

map cont & unbounded

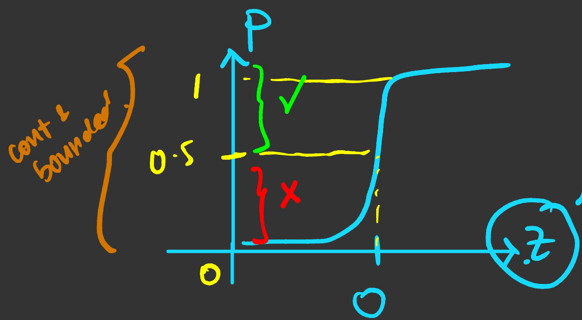


pred range : unbounded
continuous.

map

"discrete"
bounded : $0 \le \hat{y} \le 1$

if $\hat{y} < 0.5$: ✓
if $\hat{y} \geq 0.5$: ✗



pred. target (for 1 sample)
feature

model :
(prediction)

$$\hat{p}(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}} , \quad \underbrace{0 \le p(x) \le 1}_{\text{bounded}}$$

params : β_0, β_1

cost function :

samples index

$$E(\beta_0,\beta_1) = \sum_{i=1}^{N} \left[y^i \log(p(x)) + (1-y^i)\log(1-p(x)) \right]$$

$$= \sum_{i=1}^{N} \left[y_i \log \overbrace{\left(\frac{1}{1+e^{-(\beta_0+\beta_1 x_i)}}\right)}^{\text{for } y_i == 1} + (1-y_i)\log \overbrace{\left(1 - \frac{1}{1+e^{-(\beta_0+\beta_1 x_i)}}\right)}^{\text{for } y_i == 0} \right]$$

if our guess → correct, then $p(x) \to 1$, $\log(1) \to 0$

as we set → correct, $p(x) \to 0$, $\log(1-0) \to 0$

$y_i \nearrow 0$ eg: benign

$\searrow 1$ eg: malignant

E is no longer a simple convex func.

$\frac{dE}{d\beta} \rightarrow$ for GD update func.

matrix notation (SLR)

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \Big\} \text{\# of samples}$$

$\hat{y} = X\beta$

col. vector

apply element-wise to $X\beta$

$$\star \quad P = \frac{1}{1+e^{\underbrace{-X\beta}_{\text{col. vector}}}}$$

col. vector containing mapped pred vals

$$\begin{bmatrix} P(x_1) \\ P(x_2) \\ \vdots \end{bmatrix} \Big\} \text{\# samples}$$

$$A = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad -A = \begin{bmatrix} -1 \\ -2 \\ -3 \end{bmatrix}$$

$$A^{**}2 = \begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix}$$

$$e^A \rightarrow \begin{bmatrix} e^1 \\ e^2 \\ e^3 \end{bmatrix} ?$$

compare w SLR

$\rightarrow \beta = \beta - \alpha c X^T(X\beta - y)$

gradient desc update func:

$$\beta_{new} = \beta_{old} - \alpha c X^T\Big(\overbrace{\underbrace{\frac{1}{1+e^{-X\beta}}}_{\text{col. vector}}}^{\text{pred. target}} - y\Big)$$

col. vector of actual target

constant

X^T is # feat by # samples.