Tue → 15/06/21

Classification: → Classes / categories (Y/N, 1/0, T/F)

Some Concepts: Supervised (Labels)

Common Algorithm

1) DT (ID3, CART, CH.5, C5.0)
2) KNN
3) SVM → Hyperplane
4) Logistic Regression → Logits ⇒ $\frac{e^x}{1+e^x} \equiv \frac{e^x}{e^x+1} \Rightarrow \frac{p}{1-p}$

$e^x \frac{d(y)}{dx} \left(\frac{1}{1+e^{-x}}\right)$

REGRESSION ⇒ (Continuous variable)

(1) Simple Linear Regression
(2) Multiple " " aka polynomial Regression
(3) Support vector for regression (SVR)
(4) Decision Tree Regression
(5) Random forest "
(6) Ridge
(7) Polynomial

Some Intuitions

Cartesian plane (2D plane → x and y variables)

x → Independent variable (IV)

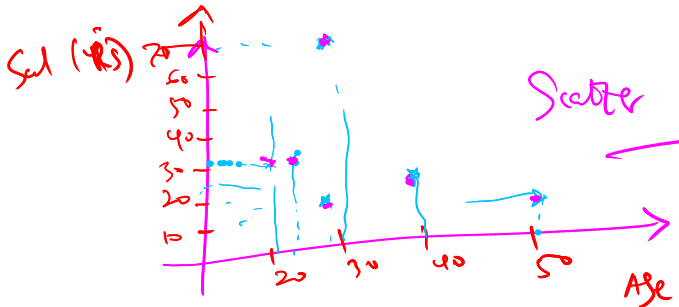
y → Dependent variable (DV)

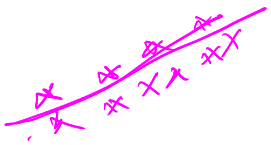
Relationship:
① → Strong Corr
② → weak corr
③ → No correlation relationship

aka Corelation

Consider the dataset below

Age	Salary
22	35,000
25	35,000
29	29,000
30	70,000
40	25,000
50	10,000





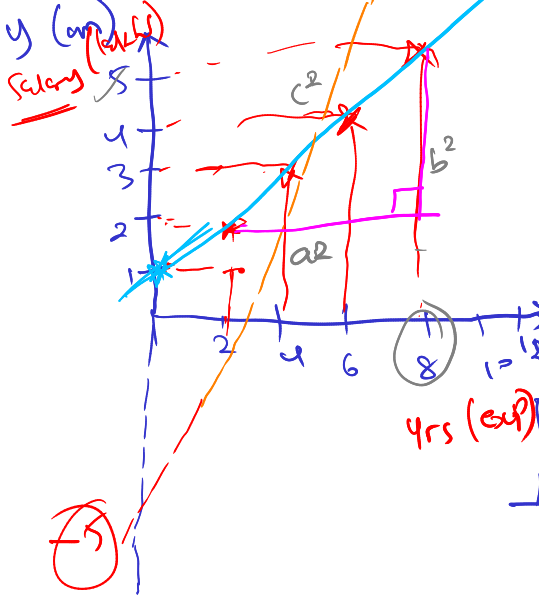
dV ↑ slope (m) ↑ intercept

$y = mx + c$ → maths (stats)

$y = wx + b + \Delta f$

$y = b_0 + b_1 x_i \Rightarrow y = b_0 + b_1 x_1$ → Linear Regression

Constant — Coeff/multiplier



A(2nd) Can you find the unknowns

Given $y = b_0 + b_1 x_1$

(1) (2)

$\text{Slope} = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$

$\frac{3}{6} = \frac{5-2}{8-2} = \frac{3}{6} = 0.5$

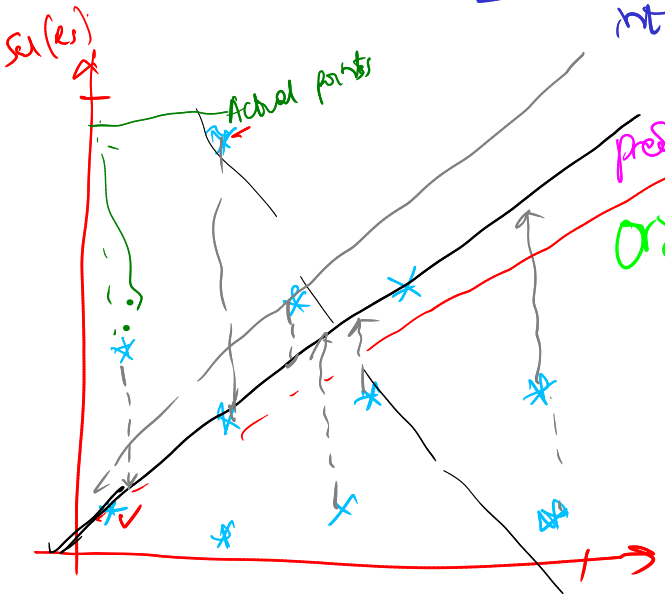
$y = 1 + 0.5x$

$\equiv 0.5x + 1$

$= 1/2 \times 12 + 1 \Rightarrow 6 + 1 = 7$

$= 1/2 \times 20 + 1 = 10 + 1 = 11$

Real world is messi x messy ie unstructure & difficult to interpret.



predicted line (points on a plane)

Regression line = line of best fit

Ordinary least squares

Actual (A), predicted = P

Error ⇒ A − P

$y - \hat{y}$

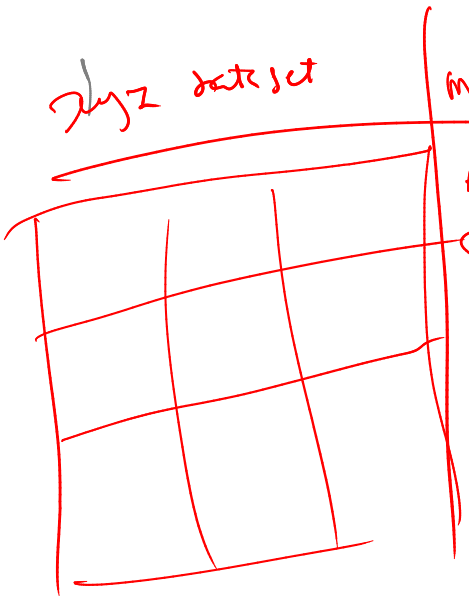
Error aka Loss function

Aim: To minimize $\Rightarrow$ OLS $= \sum |y - \hat{y}|^2$

$6 \quad -7 \quad = |-1|^2$

$OLS = \sum_{i=1} |y - \hat{y}|^2$ $\quad |-2| = 2$

$6 - 5.8 = |0.2|^2 = 0.04$



Analyze dataset | Understanding model | Inference from model ?

Age = 30
Sal = 65k

Age = 30
= ?] Same or change/diff.