2025-06-18

18 June 2025 10:18

→ Data Analytics

Descriptive → statistics

Predictive → Probability (Axiomatic) Prediction

⇒ Predictive Analytics

↓

Machine Learning

⇒ [Drive a bicycle] → 7134.... γ

- → Learn Balance
- → Learn Roadsense
- → Learn Speed

⇒ Preparing JEE

Task

360/360

Python Dev ~ Solve more Ques

Parameter

Training

→ University Exam

→ Whole Learning

↳ [Σ] → Whatever you learn will be asked

↳ Questionnaire / Answers

Parameters / Labels

Supervised Learning

UPSC / GATE / CAT

→ Logical

↳ [∫] → You learn something but diff guy is asked

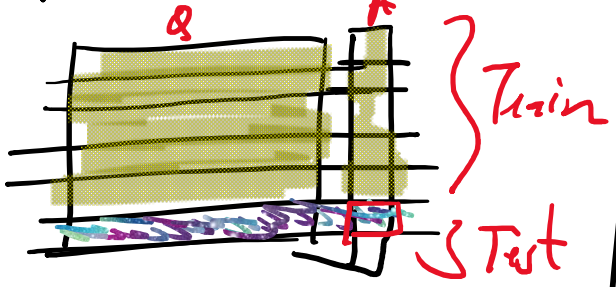
↳ Question → Solve

↳ Get close to ans

Unsupervised Learning

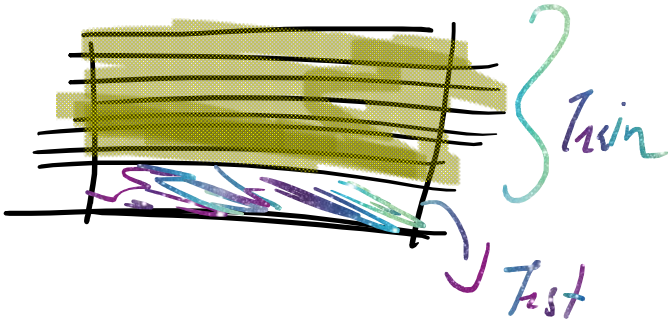
⇒ Training & Testing

Supervised



Supervised

Unsupervised



Supervised

- Regression → Continuous Values
- Classification → Discrete Values

Egs:

Regression
- Parametric → Linear Reg, LDA
- Non-Parametric → Random Forest, K-Neighbors, SVR, GDA

=> REGRESSION
⇓
Linear Regression => Parametric

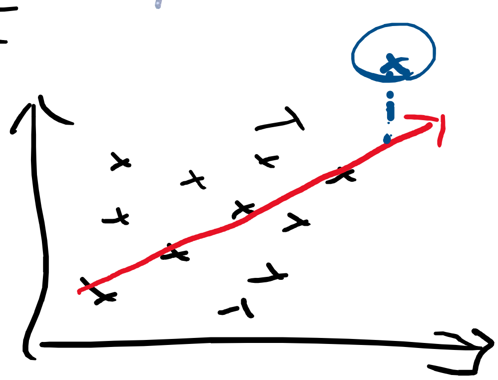
=> Studying For Test = One Method (Traceable) — Parametric

=> Lifting Weights = Many Methods (Non-Traceable) — Non-Parametric

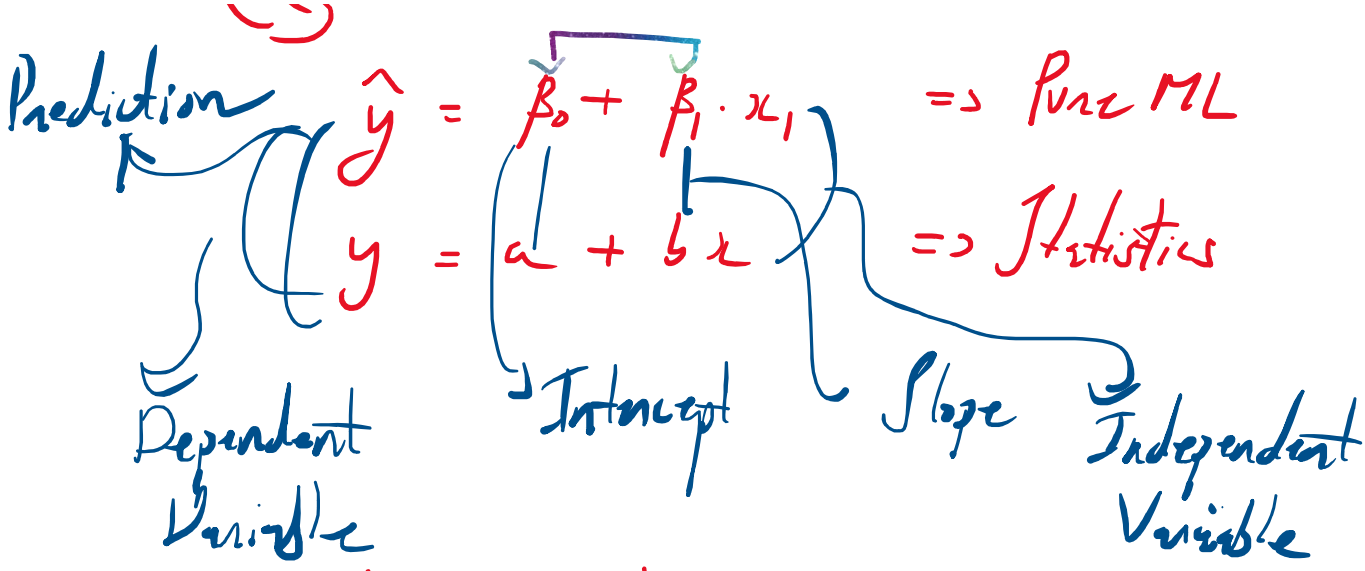
---

=> **Linear Regression:** (Supervised & Parametric)

$y = mx + b$ (m = Slope, b = Intercept)



Prediction: $\hat{y} = \beta_0 + \beta_1 x$ => Pure ML



=> Representation formula

(1) $\hat{y} = \beta_0 + \beta_1 \cdot x + \varepsilon$ ← Error

Intercept $c\beta_0 = \bar{y} - \beta_1 \times \bar{x}$ (2)

Slope $c = \beta_1 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \dfrac{SSXy}{SSX}$ (3)

=> Predict weight of the person based their height. (176 cms)

or

Heights (x)	Weights (y)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
160	72	-5.4	-4.8	88.36	45.12
171	76	1.6	-0.8	2.56	-1.28

160					
171	74	1.6	−0.8	2.56	−1.28
182	77	12.6	0.2	158.76	2.52
180	83	10.6	6.2	112.36	65.72
154	76	−15.4	−0.8	237.16	12.32

$\bar{x} = 169.4$ $\bar{y} = 78.8$ $\Sigma(x_i - \bar{x}) = 0$ $\Sigma = 0$ $\Sigma = 579.2$ $\Sigma = 124.4$

$$\beta_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{124.4}{579.2}$$

$$\beta = 0.2076$$

$$\beta_0 = \bar{y} - \beta_1 \cdot \bar{x}$$

$$= 78.8 - (0.2076 \times 169.4)$$

$$\beta_0 = 41.63$$

$$\Rightarrow \hat{y} = \beta_0 + \beta_1 x_i$$

$$\hat{y} = 41.63 + (0.2076 \times x_i)$$

$$\hat{y} = 41.63 + (0.2076 \times 176)$$

$$\hat{y} = 78.14$$

$\hat{y} = 78.16$ Multiple Regression

=> Multiple-Parameters (Multi-linear Regression)

Least Squares Estimate

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

$$\hat{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = \underbrace{(X^T \cdot X)^{-1} \cdot X^T \cdot y}$$

x_1 Product Sales 1	x_2 Prod Sales 2	y Weekly Sales
1	4	1
2	5	6
3	8	8
4	2	12

=> Adding Arbitrary Bias

$$X = \begin{pmatrix} 1 & 1 & 4 \\ 1 & 2 & 5 \\ 1 & 3 & 8 \\ 1 & 4 & 2 \end{pmatrix}$$

$$X^T \cdot X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 5 & 8 & 2 \end{pmatrix} \times \begin{pmatrix} 1 & 1 & 4 \\ 1 & 2 & 5 \\ 1 & 3 & 8 \\ 1 & 4 & 2 \end{pmatrix}$$

$(3\times4) \quad (4\times3)$

3×3

$$X^T \cdot X = \begin{pmatrix} 4 & 10 & 19 \\ 10 & 30 & 46 \\ 19 & 46 & 109 \end{pmatrix}$$

$$(X^T \cdot X)^{-1} = \begin{pmatrix} 3.15 & -0.59 & -0.30 \\ -0.59 & 0.20 & 0.046 \\ -0.03 & 0.018 & 0.054 \end{pmatrix}$$

$$(X^T \cdot X)^{-1} \cdot X^T = \begin{pmatrix} 3.15 & -0.59 & -0.30 \\ -0.59 & 0.20 & 0.046 \\ -0.03 & 0.018 & 0.054 \end{pmatrix} \times \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 5 & 8 & 2 \end{pmatrix}$$

$3\times3 \qquad 3\times4$

$$= \begin{pmatrix} 1.36 & 0.47 & -1.02 & 0.19 \\ -0.31 & -0.058 & 0.15 & 0.252 \\ -0.085 & 0.005 & 0.185 & -0.12 \end{pmatrix}$$

$$\left((X^T \cdot X)^{-1} \cdot X^T\right) y = \begin{pmatrix} 1.36 & 0.47 & -1.02 & 0.19 \\ -0.31 & -0.058 & 0.15 & 0.252 \\ -0.085 & 0.005 & 0.185 & -0.12 \end{pmatrix} \times \begin{pmatrix} 1 \\ 6 \\ 8 \\ 12 \end{pmatrix}$$

$3\times4 \qquad 4\times1$

$3 \times 4 \quad 4 \times 1$

3×1

$$\hat{\beta} = \begin{pmatrix} -1.7 \\ 3.486 \\ -0.005 \end{pmatrix}$$

β_0 - Intercept

β_1 - Prod Sales 1

β_2 - Prod Sales 2

$$\hat{y} = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2}$$

$$\hat{y} = -1.7 + 3.486 \times x_{i1} + (-0.005) \times x_{i2}$$

$P1 = 5 \qquad P2 = 3$

$$\hat{y} = -1.7 + 3.486 \times 5 + (-0.005) \times 3$$

$$= 15.79$$

X_test $(1 \quad 5 \quad 3)$ $\times$ $\begin{pmatrix} -1.7 \\ 3.486 \\ -0.005 \end{pmatrix}$

1×3 $\qquad 3 \times 1$

$(1,1)$

$[15. \ldots]$