#1 Probability to Logits

Key concept for understanding logistic regression

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$$

$$p = \frac{1}{1 + e^{-\text{logit}(p)}}$$

Terms:

- Probability (p): Value in $[0, 1]$
- Logit: Log of odds, maps $[0, 1]$ to $(-\infty, +\infty)$
- Odds: Ratio $\frac{p}{1-p}$

@AIinMinutes

#2 Conditional Independence

Fundamental concept for understanding Causal Inference

$$P(A \cap B \mid C) = P(A \mid C)P(B \mid C)$$

Terms:

- Conditional Independence: A and B are independent given C
- $P(A \mid C)$: Probability of A given C
- $P(B \mid C)$: Probability of B given C
- $P(A \cap B \mid C)$: Probability of both A and B given C

@AIinMinutes

#3 Convex Functions

Important concept for understanding gradient descent algorithms

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

Where $0 \leq \lambda \leq 1$

Terms:

- λ: Weight between 0 and 1
- $\lambda x + (1-\lambda)y$: Weighted average of points

@AIinMinutes

#4 Lagrange Multipliers

Used to solve constrained optimization problems; fundamental in SVMs and PCA

$$\nabla f(x) = \lambda \nabla g(x)$$

Subject to: $g(x) = c$

Terms:

- $\nabla f(x)$: Gradient of objective function
- λ: Lagrange multiplier
- $g(x) = c$: Constraint equation
- $\nabla g(x)$: Gradient of constraint

@AIinMinutes

#5 Hessian Matrix

Describes function curvature; important concept for understanding Newton's gradient descent method

$$H(f)_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

For $f(x_1, x_2, ..., x_n)$:

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Terms:

- $\frac{\partial^2 f}{\partial x_i \partial x_j}$: Second partial derivative
- Positive Definite: All positive eigenvalues $\rightarrow$ local minimum
- Curvature: Direction and strength of function bending

@AIinMinutes

#6 Trapezoidal Rule

Approximates integrals with linear segments; used for computing area under curves (AUC)

$$\int_a^b f(x)\,dx \approx \frac{b-a}{2n}\left[f(a) + f(b) + 2\sum_{k=1}^{n-1} f(a + k\Delta x)\right]$$

Where $\Delta x = \frac{b-a}{n}$

Terms:

- $\int_a^b f(x)\,dx$: Definite integral from a to b
- Δx: Width of each subinterval
- n: Number of subintervals
- $\sum_{k=1}^{n-1}$: Sum over interior points

@AIinMinutes

#7 Condition Number

Used to detect multicollinearity in regression

$$\kappa(A) = \frac{\sigma_{\text{max}}(A)}{\sigma_{\text{min}}(A)}$$

Terms:

- $\kappa(A)$: Condition number
- $\sigma_{\text{max}}, \sigma_{\text{min}}$: Largest/smallest singular values
- High κ: Unstable solutions, ill-conditioned system

@AIinMinutes

#8 Vector Norms

Measures vector magnitude; essential for understanding regularization

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

Common Norms:

- ℓ_1: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ (Manhattan)
- ℓ_2: $\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$ (Euclidean)
- ℓ_∞: $\|x\|_\infty = \max_i |x_i|$ (Chebyshev)

@AIinMinutes

#9 Cauchy-Schwarz Inequality

A fundamental inequality in statistics used to prove various results (e.g., why Pearson's correlation coefficient is bounded by ± 1).

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2] \cdot \mathbb{E}[Y^2]$$

Terms:

- $\mathbb{E}[XY]$: Expected product of random variables
- $\mathbb{E}[X^2]$: Second non-central moment
- Correlation: Limited to $[-1, 1]$ by this inequality

@AIinMinutes

#10 Adjacency Matrix

Encodes graph connections; core to graph neural networks and multi-hop reasoning using knowledge graphs

For an undirected graph:

$$A_{ij} = \begin{cases} 1 & \text{if vertices } i \text{ and } j \text{ are connected} \\ 0 & \text{otherwise} \end{cases}$$

Terms:

- A_{ij}: Connection between vertices i and j
- Undirected: $A_{ij} = A_{ji}$
- Powers: A^k reveals k-hop connections

@AIinMinutes