Math for AI

The eight ideas AI runs on, in plain words and in real notation.

OBJECTIVE

Read only the first line of each box for a three-minute refresher, or read all three lines for the real math. Every example here is a worked one: the numbers come from the widgets on the website, where you can move them yourself.

VECTORS

1. A vector is a list of numbers that points somewhere. AI turns words, pictures and sounds into lists like this so it can work with them.
2. v = (x, y), length |v| = sqrt(x^2 + y^2)
3. The arrow (3, 4) has length sqrt(9 + 16) = 5. Double it to (6, 8) and the length doubles to 10.

DOT PRODUCT AND SIMILARITY

1. Multiply two lists number by number and add it all up. Big means they point the same way, zero means they have nothing in common, and below zero means they point opposite ways.
2. a · b = a1b1 + a2b2, cosine = a · b ÷ (|a| |b|)
3. Same direction scores 1. A quarter turn (90°) scores 0. Pointing opposite ways (180°) scores −1.

MATRICES

1. A matrix is a grid of numbers that moves things: it can stretch, turn or flip every point at once. AI models are mostly big grids like this, mixing numbers together.
2. M v = (m11x + m12y, m21x + m22y)
3. Turn = [[0, −1], [1, 0]]. Turn twice: [[0, −1], [1, 0]] x [[0, −1], [1, 0]] = [[−1, 0], [0, −1]], a half turn.

FUNCTIONS AND SLOPE

1. A function is a rule: put a number in, get a number out. The slope says how much the answer changes each time you nudge the input. A model's weights are slopes it gets to pick.
2. y = w·x + b, slope = w
3. With w = 1 and b = 0, one step right always means one step up. Change b and the whole line slides, but the slope stays 1.

GRADIENTS AND LEARNING RATE

1. To learn, a model checks which way is downhill on its mistakes and takes a step that way. The learning rate is how big each step is: too small is slow, too big overshoots.
2. w <- w − rate x slope, slope of (w − 3)^2 is 2(w − 3)
3. Start at w = 0. Rate 0.1 moves to 0.6. Rate 0.5 lands on 3 in one step. Rate 1.1 jumps to 6.6, farther from 3 than where it started.

PROBABILITY

1. Probability says how often something happens if you try many times. AI does not know the answer for sure. It picks what is most likely, and it learns those odds by counting.
2. P(face) = 1 ÷ 6 ~ 16.7%
3. Ten rolls are lumpy: one face may never show up. After 1000 rolls every face lands within 3 points of 16.7%.

SOFTMAX

1. Softmax turns any scores into shares that add up to 100%. The highest score gets the biggest share. Those shares are probabilities, not a promise that the top pick is right.
2. share(i) = e^(si) ÷ (e^(s1) + e^(s2) + e^(s3))
3. Scores 2, 1, 0 become about 67%, 24% and 9%. Add 1 to every score and the shares stay exactly the same.

LOSS

1. Loss is one number for how wrong a model is. Training is just making that number smaller. Squaring each miss means one big miss hurts more than a few small ones.
2. loss = (guess1 − answer1)^2 + (guess2 − answer2)^2 + …
3. One guess off by 4 costs 16. Two guesses off by 2 cost only 4 + 4 = 8.