

A number or a yes/no. Pick the model.

D8A.ACADEMY/RESOURCES/LOGISTIC-VS-LINEAR

LINEAR

predicts	a number: price, demand, minutes
output	any value, minus to plus
loss	squared error

REMEMBER

both	are linear in the inputs
imbalance	move the threshold, check recall

LOGISTIC

predicts	a probability: churn, fraud, click
output	0 to 1, through the sigmoid
threshold	← 0.5 is a choice, not a law

WORTH MEMORIZING

01 Same inputs, two outputs

Both models draw a line through the features. One reads the line as a number, the other pushes it through a sigmoid and reads a probability.

LINEAR

- $y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2$
- the line is the prediction
- fit by least squares
- read: "price is about 240"

LOGISTIC

- $p = \text{sigmoid}(b_0 + b_1 \cdot x_1 + b_2 \cdot x_2)$
- the line is the log-odds
- fit by maximum likelihood
- read: "72% chance of churn"

02 Trap: the threshold is a decision

A logistic model gives a probability. The yes/no comes from a threshold you choose.

THRESHOLD	FLAGS MORE	USE WHEN
0.5	balanced	classes are balanced, errors cost the same
0.2	yes	missing a case is expensive: fraud, churn
0.8	no	a false alarm is expensive: a manual review

→ Interview phrasing: logistic regression gives me a probability; I pick the threshold with the business, from the cost of each kind of error.