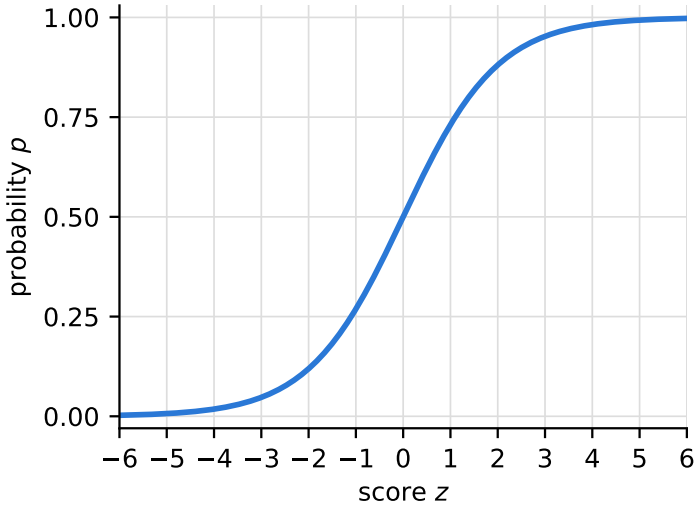
DATA 202 — Logistic Regression · Week 11 Monday

How can a linear model classify? Applications to borrow against a home in Kent County, Michigan, decided in 2025: approved or denied. DTI = debt-to-income, the share of monthly income that would go to debt payments (%). Your professor will tell you when to do each part.

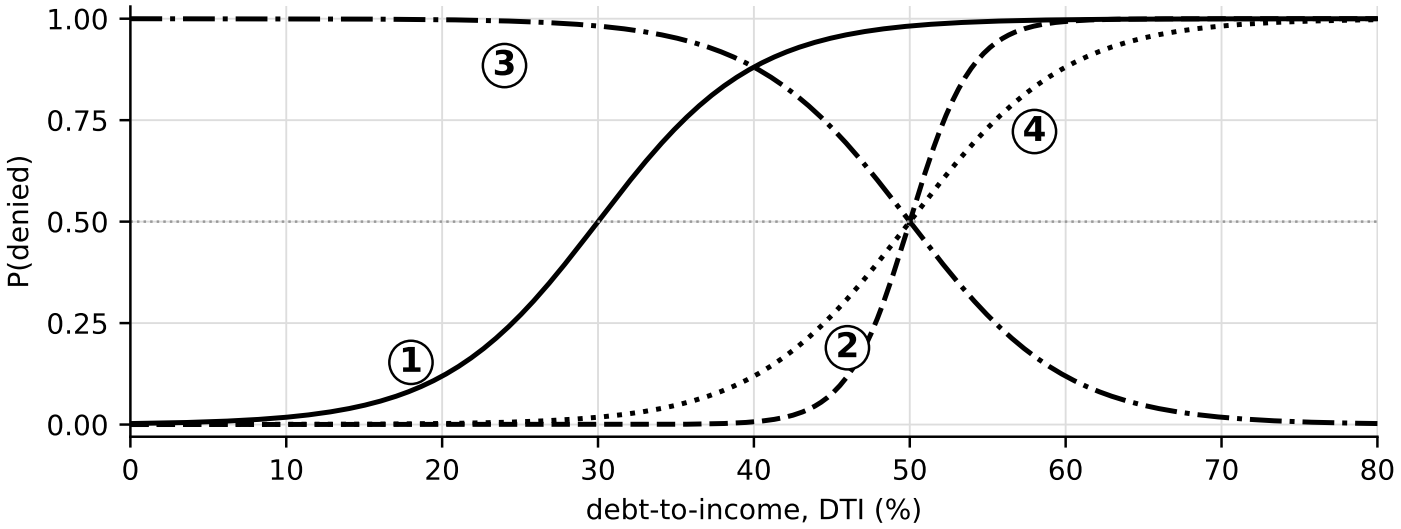
1a · From a score to a probability



Step 1, the score: $z = -10 + 0.2 \times \text{DTI}$. Step 2, the probability: $p = 1/(1 + e^{-z})$, the curve.

1. **Compute** each applicant's score.
   A, DTI 30%: $z =$ _____
   B, DTI 50%: $z =$ **0**
   C, DTI 70%: $z =$ _____
2. **Mark** A, B and C on the curve.
3. Raise each score by 1. **Draw** an arrow along the curve for each applicant. Whose probability moves the most? _____
4. **Pair talk:** whose decision can a small change in the data flip: a clear case or a borderline one? Why might that matter to an applicant?

1b · Four S-curves



Four curves for the chance of denial. **Match** each score to its curve. (*Hint:* a curve crosses 50% where its score is 0.)

score	curve	score	curve
(a) $z = -10 + 0.2 \times \text{DTI}$	_____	(b) $z = -25 + 0.5 \times \text{DTI}$	_____
(c) $z = -6 + 0.2 \times \text{DTI}$	_____	(d) $z = 10 - 0.2 \times \text{DTI}$	_____

5. **Pair talk:** many lenders won't approve a DTI above 50%. Which curve looks most like that rule? What would a curve with an even bigger slope look like?

2 · Many columns, one score

The model now uses DTI, loan-to-value, the loan amount, and the loan's purpose, lien and type. Before the slopes appear on screen, write + if you think the column raises the chance of denial, − if it lowers it.

column	your guess	the model
debt-to-income		
loan amount		
a home purchase (vs. a cash-out refinance)		
a second lien (vs. a first mortgage)		
a conventional loan (vs. an FHA loan)		

A **cash-out refinance** replaces a mortgage with a bigger one and pays the difference in cash. A **second lien** is a second loan on a home that already has a mortgage. An **FHA loan** is insured by the government, for borrowers with smaller down payments or lower credit scores.

Pair talk: the lender saw each applicant's credit score, but this file doesn't have it. Which of these columns might be standing in for it?

3 · Choose the threshold

Ten training applications: the model's chance of denial, and what the lender decided (D = denied, A = approved).

application	1	2	3	4	5	6	7	8	9	10
chance of denial	0.84	0.62	0.55	0.41	0.33	0.27	0.18	0.12	0.08	0.03
lender's decision	D	D	A	D	A	A	D	A	A	A

The model **flags** an application as *denied* when its chance is at or above the threshold. A positive is a denied application. **Precision** = TP ÷ (TP + FP); **recall** = TP ÷ (TP + FN).

1. **Fill in** the table, first for a threshold of 50%, then for 20%.

threshold	TP	FP	FN	TN	precision	recall
50%						
20%						

2. **Pair talk:** a false positive turns away someone the lender would have approved; a false negative approves someone it would have denied. Which threshold would you choose, and who pays for each kind of mistake?