Recap: Logistic Regression and Regularization

DS701 Session 13 — Mon Oct 19, 2026

Today’s plan

Knowledge-Check Review

KC 1: The model, in one line

Write down the logistic regression model: what linear function does it fit, what quantity is that linear function equal to, and how do you get from it to a probability $p(x)$*? Then say how the coefficients are chosen (which principle) and why, unlike linear regression, they cannot be found in closed form.*

KC 2: Why the coefficients blow up, and two ways to stop them

When two columns of the design matrix are nearly linearly dependent (multicollinearity), the least-squares coefficients tend to become very large in magnitude. Explain in a sentence or two ***why*** *that happens, how ridge regression's penalty* $c\|\beta\|_2^2$ *addresses it, and what LASSO's* $c\|\beta\|_1$ *penalty does that ridge cannot.*

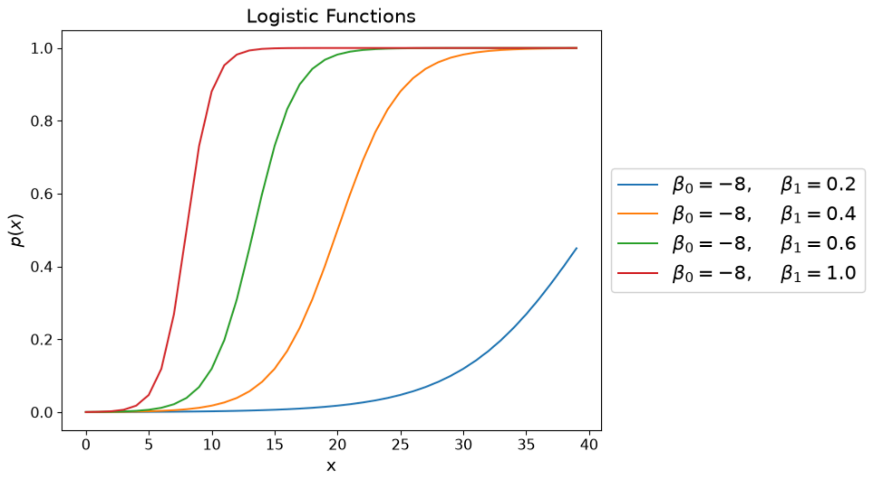
KC 3: 0.5 is not a law of nature

To turn logistic regression into a classifier the lecture used the rule "say **yes** *if $p(x) > t$" and found on the admissions data that F1 peaked near $t \approx 0.3$, not 0.5. Explain what lowering the threshold does to precision and to recall, why 0.5 is not automatically the right choice, and give one real setting where you would deliberately pick a low threshold and one where you would pick a high one.*

Highlights

Logistic regression = a linear model + a sigmoid

Last week: a continuous y as a linear function of x. This week: a **0/1** y, and we want a **probability**.

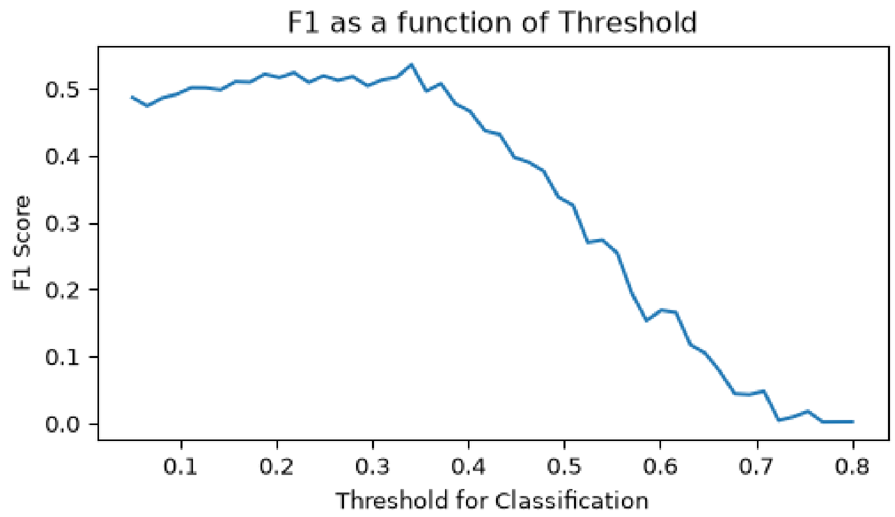


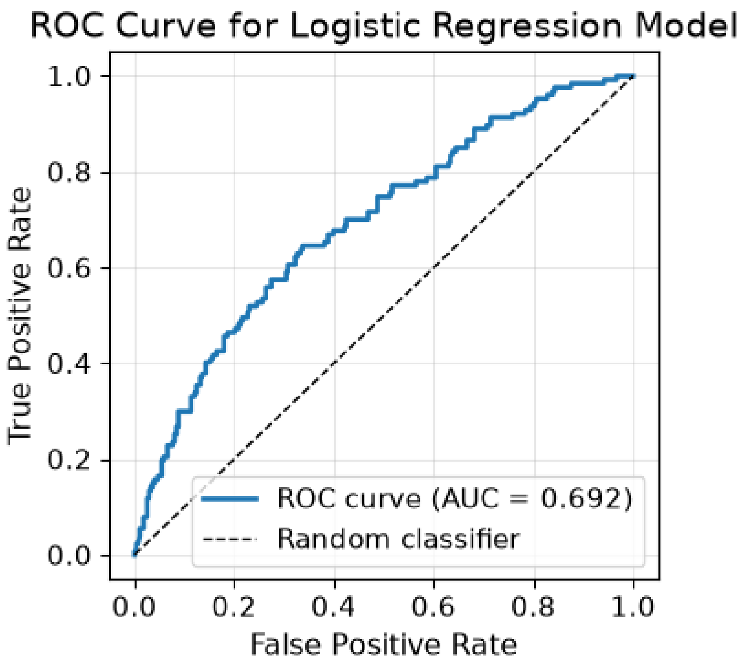
Fit by maximum likelihood — same principle as Lecture 05

One observation: $\mathrm{P}(y_i \mid x_i) = p_i^{y_i}\,(1-p_i)^{1-y_i}, \qquad p_i = \sigma(\boldsymbol{\beta}^\top \mathbf{x}_i).$

The boundary is still a line

Evaluation beyond accuracy: the threshold is a decision





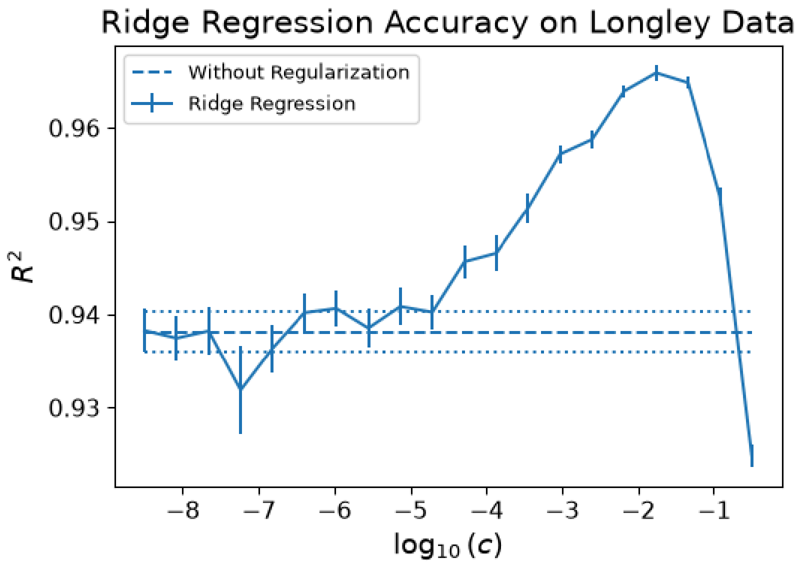
Admissions: F1 peaks near $t \approx 0.3$ — the model rarely says "admit" at 0.5 because admits are the minority.

ROC: TPR vs FPR, one point per threshold; diagonal = random; **AUC** summarizes ranking quality *without* choosing a

Regularization: a penalty that trades bias for variance

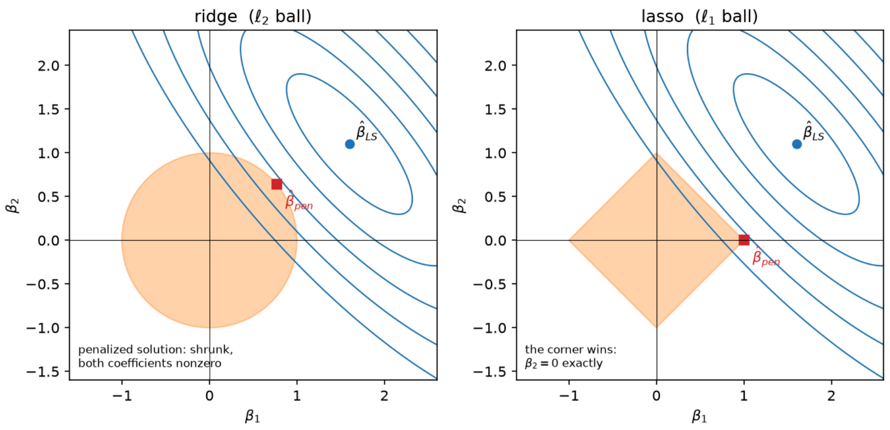
Lecture 07: flexible fits chase noise → high variance. Regularization is the knob:

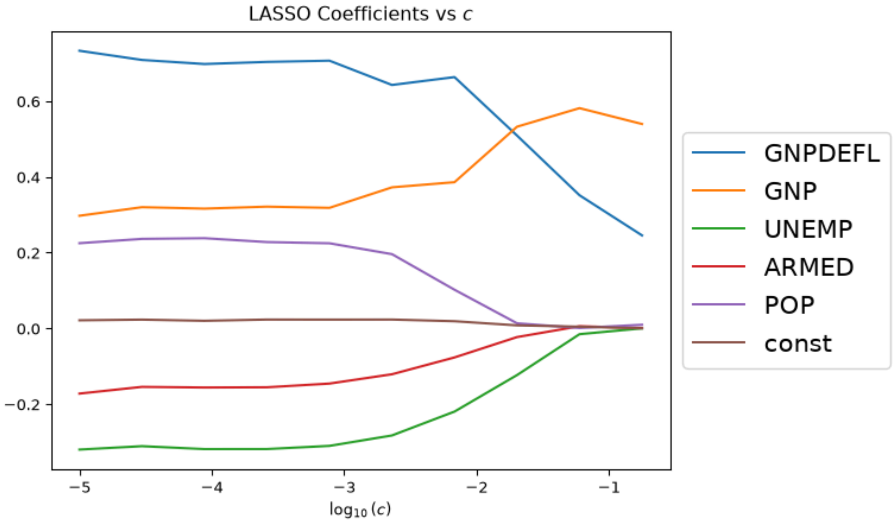
$$\min_{\beta} \|X\beta - \mathbf{y}\|_2^2 + c\,R(\beta)$$



Longley, ridge: held-out R^2 rises with c, then collapses — the sweet spot is found by **cross-validation**.

ℓ_2 shrinks, ℓ_1 selects





Same contours, different penalty balls. The ℓ_1 diamond has **corners on the axes** — the contours usually hit a corner first, and a corner means a coefficient is *exactly* zero.

Longley, LASSO: as c grows, ARMED, UNEMP, POP go to zero; GNP and GNPDEFL remain — the near-dependent variables are the ones dropped.

Two rules you must not skip

Where this goes next

In-Class Activity

Activity: from the likelihood to a threshold you would ship

Goal: build logistic regression from its likelihood and check it against sklearn; trace ridge and lasso paths on data with correlated and irrelevant features, pick α by CV, and see what forgetting to standardize costs; then take a **90/10 (A1) or 95/5 (B1) imbalanced** problem, watch accuracy lie, sweep the threshold, plot ROC and precision–recall, choose a threshold for a stated cost, and try `class_weight="balanced"`.

- Work in groups of 2–3. Open **your section's** notebook.
- Parts marked **(autograded)** are submitted to Gradescope; the rest is participation.
- Staff will circulate — be ready to explain any part of your work.
- Dependencies: `numpy`, `pandas`, `matplotlib`, `scikit-learn`.

Section A1 Open in Colab Open on GitHub

Section B1 Open in Colab Open on GitHub

Going deeper

Full lecture notes: Logistic Regression and Regularization

- Odds and Log-Odds — the logit, the sigmoid, what β_0 and β_1 do to the curve
- Logistic vs Linear Regression — the likelihood, and why there is no closed form
- How Gradient Descent Works
- Logistic Regression in Perspective — precision, recall, F1 vs threshold on the admissions data
- ROC-AUC Curve
- What is regularization? and Condition Number
- Ridge Regression — the geometry of large coefficients, the closed form, Scaling
- The LASSO — coefficient paths on the Longley data; Model Selection
- Background: bias–variance in Generalization; MLE in Probabilistic Modeling