CSCI 145 Problem 10: Where the Sigmoid Comes From

July 20, 2026

Submission Instructions

Please upload your work to Gradescope by **11:59pm on DATE**.

- Turn in a single PDF of *handwritten* solutions. Typed (including LaTeX) solutions will not be accepted.
- Problems are grouped into three-week **cycles**. Once per cycle, you will meet with the grader and present the solution to *one* problem from that cycle, selected at random when you arrive — so bring your work and be ready to explain any of them.

Problem 10: Where the Sigmoid Comes From

You test positive for a disease that 1% of people have, on a test with a 5% false-positive rate and a 10% false-negative rate, and the Probability lecture put your chance of being sick at about 15%. Now a second, independent test comes back positive too. Rerunning Bayes' rule from scratch for each new piece of evidence gets old fast, so this problem sets up bookkeeping in which independent evidence simply *adds*, and then converts the running total back into a probability. What comes out is the sigmoid of a linear score, which is the model lecture asked you to take on faith. Throughout, $Y \in \{0, 1\}$ is the label and $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ the observed features, the same $\mathbf{x}$ the reading feeds to the model.

Part A: Posterior Odds

The **odds** of an event with probability p are $p/(1-p)$. For one piece of evidence $X = x$, write Bayes' rule for $\Pr(Y = 1 \mid X = x)$ and for $\Pr(Y = 0 \mid X = x)$, divide one by the other, and show that $\Pr(X = x)$ cancels:

$$\frac{\Pr(Y = 1 \mid X = x)}{\Pr(Y = 0 \mid X = x)} = \underbrace{\frac{\Pr(Y = 1)}{\Pr(Y = 0)}}_{\text{prior odds}} \cdot \underbrace{\frac{\Pr(X = x \mid Y = 1)}{\Pr(X = x \mid Y = 0)}}_{\text{likelihood ratio}}.$$

Take the log of both sides, so odds become *log-odds* and the product becomes a sum. Then check it on the medical test, with $Y = 1$ meaning "sick" and $X = +$ a positive result: compute the prior odds and the likelihood ratio separately, multiply, and confirm you recover the $\approx 15\%$ posterior.

Part B: Independent Evidence Adds in Log-Odds

Now take d pieces of evidence $X_1, \dots, X_d$ (d different words in an email, say, for a spam classifier), assumed **conditionally independent given** Y: $\Pr(X_1, \dots, X_d \mid Y) = \prod_{j=1}^{d} \Pr(X_j \mid Y)$, the "naive Bayes" assumption. Repeat Part A with the full evidence vector $\mathbf{x}$ in place of x, apply conditional independence to the likelihood ratio, and show the log posterior odds splits into $d + 1$ separate contributions:

$$\log \frac{\Pr(Y = 1 \mid \mathbf{x})}{\Pr(Y = 0 \mid \mathbf{x})} = \log \frac{\Pr(Y = 1)}{\Pr(Y = 0)} + \sum_{j=1}^{d} \log \frac{\Pr(X_j = x_j \mid Y = 1)}{\Pr(X_j = x_j \mid Y = 0)}.$$

Each piece of evidence contributes its own log-likelihood-ratio, and those contributions add.

Part C: Inverting Back to a Probability

Suppose further that each feature's log-likelihood-ratio is linear in that feature, $\log \frac{\Pr(X_j = x_j \mid Y = 1)}{\Pr(X_j = x_j \mid Y = 0)} = w_j x_j + c_j$ for constants w_j, c_j (true, for instance, when $X_j \mid Y$ is Gaussian with a Y-independent variance; take it as given). Substitute into Part B and collect every constant, the prior log-odds and all the c_j, into a single bias b, so the log posterior odds becomes $\langle \mathbf{w}, \mathbf{x} \rangle + b$ with $\mathbf{w} = (w_1, \dots, w_d)$. Finally invert: if $z = \log \frac{p}{1-p}$, solve for p in terms of z by exponentiating and rearranging, and conclude

$$\Pr(Y = 1 \mid \mathbf{x}) = \sigma\big(\langle \mathbf{w}, \mathbf{x} \rangle + b\big), \qquad \sigma(z) = \frac{1}{1 + e^{-z}},$$

which is logistic regression's model, with the bias written out rather than folded into $\mathbf{x}$. Conclude in one sentence: where does the sigmoid's shape come from?

Where this goes next. This is one stop on the course's Bayes thread: the Mean Squared Error lecture derived a *loss* from a likelihood, this problem derives a *model* from a posterior, and the Generalization lecture will derive *regularization* from a prior. Part C's move also produces the softmax (run it with k classes instead of 2) and the distribution we sample actions from in the final unit.