CSCI 145 Problem 20: Build an Associative Memory

October 27, 2026

Submission Instructions

Please upload your work to Gradescope by **11:59pm on Wednesday, November 25**.

- Turn in a single PDF of *handwritten* solutions. Typed (including LaTeX) solutions will not be accepted.
- A complete, good-faith attempt earns submission credit. The grader may select any recent assigned problem for a brief conversation, so keep your work and be prepared to explain each solution and answer a related question. Scheduling details will be announced in class.

Problem 20: Build an Associative Memory

A lookup system stores at most $n \geq 2$ records $(\mathbf{t}_j, \mathbf{u}_j)$, where each token $\mathbf{t}_j$ is a one-hot vector from a vocabulary of size V and $\mathbf{u}_j$ is its stored value. A query token $\mathbf{q}$ should retrieve the corresponding record. Consider one content-only attention head with key dimension $d_k = V$ and

$$\mathbf{W}_Q = \mathbf{W}_K = s\mathbf{I}_V, \qquad s \geq 0.$$

Fix a desired failure tolerance $0 < \delta < 1/2$.

1. Assuming $\mathbf{q}$ occurs exactly once, what is the smallest s that guarantees the matching record's attention weight is at least $1 - \delta$ uniformly over memories of length at most n?
2. If the same token occurs twice with different stored values, why can no content-only query and key maps retrieve the more recent occurrence merely by increasing their scale?

Hint: You may use the scaled dot-product attention and row-wise softmax definitions from the reading and the orthogonality of distinct one-hot vectors. For a memory of length ℓ with exactly one matching token, the displayed query and key maps give the matching record weight

$$\frac{\exp(s^2/\sqrt{V})}{\exp(s^2/\sqrt{V}) + \ell - 1}.$$

Takeaway: *Attention can put almost all its weight on a unique matching record by increasing the query and key scale. If two records have the same token, that scale cannot tell it which one came later. Choosing the latest match requires information about position.*