

CSCI 145 Problem 20: Build an Associative Memory

August 27, 2026

Submission Instructions

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

- Turn in a single PDF of *handwritten* solutions. Typed (including L^AT_EX) 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; bring your work and be ready to explain any of them.

Problem 20: Build an Associative Memory

2 points

A memory contains at most $n \geq 2$ records $(t_j, \mathbf{u}_j)$, where token t_j comes from a vocabulary of size V and $\mathbf{u}_j \in \mathbb{R}^p$ is its stored value. A query token q should return the value paired with it. Treat token identities as one-hot vectors and the $\mathbf{u}_j$ as the value vectors already produced by $\mathbf{W}_V$. Assume $0 < \delta < 1/2$.

Consider the single-head construction $\mathbf{W}_Q = \mathbf{W}_K = s\mathbf{I}_V$, with key dimension $d_k = V$. Audit how well it acts as an associative memory. Whenever q occurs exactly once, determine how s must depend on n and δ to give its record attention weight at least $1 - \delta$, uniformly over memories of length at most n .

Now allow the same token to occur twice with different stored values, and require the most recent occurrence. Prove that a content-only head cannot meet this requirement for every memory, regardless of its learned query and key matrices. Describe one concrete kind of positional information that would break the tie in favor of the later duplicate, and explain how it removes the obstruction. You do not need to optimize a positional scale.

Results you may use. You may use the scaled dot-product attention and row-wise softmax definitions from the reading. A content-only attention layer is permutation equivariant.