Mathematics for New Technologies in Finance

Exercise sheet 9

Background. Associative recall measures a generative model's capability to remember relevant information from past input. This requires storing information in hidden states and is therefore subject to information-theoretic bounds. For transformers, these bounds depend on the size of the context window and also on the choice of attention mechanism. Here, we consider a stylized recall task and ask for a *constructive* solution. The didactical goal is to get familiar with query, key, and value computations.

Definition. Given a sequence of key-value pairs and a key, the associative recall task is to output the value corresponding to this key. Example:

$$\text{A6B8C3B} \to 8$$

Accordingly, a correct input-output pair (x, y) for this example is

$$x = \text{A6B8C3B?} \qquad y = \text{A6B8C3B8}$$

Exercise 9.1 Are transformers able to perfectly solve the associative recall task, for recall sequences with fixed length?

Hint. As the focus is on the attention mechanism, you may for simplicity choose arbitrary functions for token embeddings, position embeddings, and feed-forward networks. Moreover, you may ignore layer normalization. Finally, it might be helpful to note that a transformer block whose attention layer has all-zero weights is just a feed-forward network.

Exercise 9.2 (Memory of Signature transforms) Notice the analogy to the first example:

(a) Given a C^1 trajectory $u : [0, 1] \to \mathbb{R}^2$. How many and which signature components do you need to know to reconstruct the area swept over by the vector from $u(0)$ to $u(t)$ on the interval $[0, 1]$.

(b) Let us assume that the first component is time, i.e. $u^1(t) = t$ for $t \in [0, 1]$. Which signature components do you need to know (possibly countably many) to reconstruct $u^2(t) - u^2(0)$ perfectly for every $t \in [0, 1]$. If you can only store finitely many signature components: which ones to choose to do an approximate reconstruction and what would be a criterion for the approximation quality (Notice that u is C^1 and that one can make a Fourier expansion of $u^2(t) - u(0)$).