ECE/CS/ME 539 – Fall 2024 — Activity 30

1. Backpropagation Through Time

Consider a recurrent linear system with state $h_t \in \mathbb{R}$ and inputs $x_t \in \mathbb{R}$ defined by the equation

$$h_t = ah_{t-1} + bx_t$$

where a is a state transition coefficient and b is an input transformation coefficient. Suppose that we unroll this system for T time steps $x_1, \ldots, x_T$ using a known initial state h_0, and that we want to find the coefficients a and b so that the final state h_T matches a specific target value y by minimizing the mean squared error.

$$L = \frac{1}{2}(h_T - y)^2$$

(a) Derive $\frac{\partial L}{\partial h_T}$.

(b) Show that $\frac{\partial L}{\partial h_t} = \frac{\partial L}{\partial h_T} a^{T-t}$.

(c) Show that

$$\frac{\partial L}{\partial a} = \sum_{t=1}^{T} \frac{\partial L}{\partial h_t} h_{t-1} \quad \text{and} \quad \frac{\partial L}{\partial b} = \sum_{t=1}^{T} \frac{\partial L}{\partial h_t} x_t.$$

(d) Suppose that $a < 1$. Discuss what happens to $\frac{\partial L}{\partial h_t}$ if $T \gg t$. What if $a > 1$?

(e) This problem assumes that both the inputs and states are scalars. In RNNs, we usually have inputs and hidden state vectors, in which case the transition weights A and input transformation weights B are matrices (not scalars). What are the conditions on A or B that would lead to similar issues as those identified in part (d)?

(f) Propose two ways to prevent gradient explosion when training RNNs.