ECE/CS/ME 539 – Fall 2024 — Activity 33

Problem Statement

Consider a Gated Recurrent Unit (GRU) with scalar inputs $x_t \in \mathbb{R}$, hidden states $h_t \in \mathbb{R}$, and output $o_t \in \mathbb{R}$. The GRU equations for a single timestep are given by:

$$\begin{aligned} z_t &= \sigma(W_{\mathrm{xz}} x_t + W_{\mathrm{hz}} h_{t-1} + b_{\mathrm{z}}) \\ r_t &= \sigma(W_{\mathrm{xr}} x_t + W_{\mathrm{hr}} h_{t-1} + b_{\mathrm{r}}), \\ \tilde{h}_t &= \tanh(W_{\mathrm{xh}} x_t + W_{\mathrm{hh}}(r_t \odot h_{t-1}) + b_h) \\ h_t &= z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t \end{aligned}$$

where σ is the sigmoid function, and $\odot$ denotes element-wise multiplication.

The system is unrolled for T timesteps, with known input values $x_1, \ldots, x_T$ and initial state $h_0 = 0$. The goal is to minimize the mean squared error between the output at the final timestep and a target value y:

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

Questions

1. **Forward Pass:**
   - Compute the forward pass for $T = 2$, given:

     $$x_1 = 0, \quad x_2 = 1.0, \quad y = 2, \quad h_0 = 0$$

     and the parameter values:

     $$\begin{aligned} W_{\mathrm{xz}} &= 0.5, \quad W_{\mathrm{hz}} = 0.1, \quad b_z = 0.0, \\ W_{\mathrm{xr}} &= 0.4, \quad W_{\mathrm{hr}} = 0.2, \quad b_r = 0.0, \\ W_{\mathrm{xh}} &= 0.3, \quad W_{\mathrm{hh}} = 0.3, \quad b_h = 0.0, \end{aligned}$$

     Compute h_t for $t = 1, 2$.
2. **Backward Pass:**
   - Compute $\frac{\partial L}{\partial h_T}$.
   - Compute $\frac{\partial L}{\partial W_{\mathrm{xz}}}$.
   - Compute $\frac{\partial L}{\partial W_{\mathrm{xh}}}$.
3. **Gradient Stability:**
   - Why do GRUs help mitigate gradient vanishing or explosion compared to simple RNNs?