STAT3007/7007 Deep Learning, Tutorial 7

2022 Semester 2

1. (RNN)

(a) Design an RNN that takes in a sequence of real numbers $x_1, x_2, \ldots$, with x_t being the input at time step t, and outputs at time step t the simple moving average $y_t = \frac{1}{3}(x_t + x_{t-1} + x_{t-2})$. Here $x_t = 0$ for $t \leq 0$.

(b) Repeat (a) to compute the exponential moving average $y_t = \begin{cases} x_1 & t = 1, \\ \alpha x_t + (1-\alpha) y_{t-1}, & t > 1. \end{cases}$

(c) Consider learning the parameter α for the RNN in (b) using gradient descent. Explain how we can compute the derivative of a loss function wrt α.

2. (Exploding/vanishing gradients)

(a) What are the exploding gradient problem and the vanishing gradient problem?

(b) How do exploding/vanishing gradients affect gradient-based learning?

(c) How does GoogLeNet avoid vanishing gradients?

(d) How does LSTM avoid exploding/vanishing gradients?