Comprehensive Tutorial on RNN, LSTM, and GRU

Contents

1 Introduction

Recurrent neural networks (RNNs), Long Short-Term Memory networks (LSTMs), and Gated Recurrent Units (GRUs) are fundamental architectures for processing sequential data. This booklet provides an in-depth mathematical and conceptual understanding of these models.

2 Recurrent Neural Networks (RNN)

2.1 Overview

RNNs process sequences by maintaining a hidden state that captures past information.

2.2 Equations

At time step t:

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t + b)$$
$$y_t = W_{hy}h_t + c$$

2.3 Limitations

- Vanishing and exploding gradients.
- Difficulty in learning long-term dependencies.

3 Long Short-Term Memory (LSTM)

3.1 Architecture

LSTMs introduce gates to control the flow of information:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad \text{(forget gate)}$$
$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad \text{(input gate)}$$
$$g_t = \tanh(W_g[h_{t-1}, x_t] + b_g) \quad \text{(candidate memory)}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad \text{(cell state)}$$
$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad \text{(output gate)}$$
$$h_t = o_t \odot \tanh(c_t) \quad \text{(hidden state)}$$

3.2 Advantages

- Capable of learning long-term dependencies.
- More stable training compared to vanilla RNNs.

4 Gated Recurrent Units (GRU)

4.1 Architecture

GRUs simplify LSTMs by combining forget and input gates:

$$z_t = \sigma(W_z[h_{t-1}, x_t]) \quad \text{(update gate)}$$
$$r_t = \sigma(W_r[h_{t-1}, x_t]) \quad \text{(reset gate)}$$
$$\tilde{h}_t = \tanh(W[r_t \odot h_{t-1}, x_t]) \quad \text{(candidate state)}$$
$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad \text{(final state)}$$

4.2 Comparison with LSTM

- Fewer parameters.
- Comparable performance in many tasks.

5 Comparison Table

Feature	RNN	LSTM	GRU
Handles long-term dependencies	Poor	Excellent	Good
Gating mechanisms	None	Forget, Input, Output	Update, Reset
Training stability	Low	High	High
Parameter count	Low	High	Medium

6 Applications

- Natural Language Processing
- Speech Recognition
- Time Series Forecasting
- Machine Translation