

Recurrent Neural Networks

Understanding Sequential Models

July 19, 2025

What are RNNs?

RNN Mechanics

Training RNNs

Variants of RNNs

Applications

Summary

What are RNNs?

Why Recurrent Neural Networks?

RNNs are designed to handle sequential data.

- Traditional neural networks assume independent inputs
- RNNs capture dependencies across time steps
- Suitable for time series, language, speech, etc.

Given a sequence $x = [x_1, x_2, \dots, x_T]$, RNN learns a function:

$$h_t = f(h_{t-1}, x_t)$$

- h_t : hidden state at time t
- Encodes past information from sequence

RNN Mechanics

Basic RNN Cell

At each time step:

$$h_t = \tanh(W_h h_{t-1} + W_x x_t + b)$$

$$y_t = W_y h_t + b_y$$

- $x_t \in \mathbb{R}^{n_x}$: input vector
- $h_t \in \mathbb{R}^{n_h}$: hidden state
- $y_t \in \mathbb{R}^{n_y}$: output

Shape Calculations

Let:

- Input shape: (T, n_x)
- Hidden state size: n_h
- Output size: n_y

Then:

- $W_x \in \mathbb{R}^{n_h \times n_x}$
- $W_h \in \mathbb{R}^{n_h \times n_h}$
- $W_y \in \mathbb{R}^{n_y \times n_h}$

Training RNNs

Backpropagation Through Time (BPTT)

Gradients are computed over all time steps:

$$\frac{\partial L}{\partial W} = \sum_{t=1}^T \frac{\partial L}{\partial h_t} \cdot \frac{\partial h_t}{\partial W}$$

- Computationally expensive for long sequences
- Suffers from vanishing/exploding gradients

- **Vanishing gradients:** long-term dependencies hard to learn
- **Exploding gradients:** unstable training
- **Short memory:** can't retain info over long sequences

Variants of RNNs

Variants of RNNs

- **LSTM (Long Short-Term Memory):**
 - Uses gates to control information flow
 - Remembers information over longer periods
- **GRU (Gated Recurrent Unit):**
 - Simplified LSTM
 - Fewer parameters, faster training

When to Use LSTM or GRU

- Use LSTM for long sequences with complex dependencies
- Use GRU for faster training with decent performance
- Start simple, benchmark both

Applications

Applications of RNNs

- Language modeling
- Machine translation
- Speech recognition
- Music generation
- Time series forecasting

Summary

Key Concepts Recap

- RNNs process sequences step-by-step
- Hidden states capture memory
- Training uses backpropagation through time
- LSTM and GRU address gradient issues

Thank you!

Questions?