CODE BASICS

DEEP LEARNING TAKEAWAYS

codebasics.io

Chapter:

Sequence Models

What are Sequence Models?

1. Sequence models process sequential data, making them ideal for tasks like time series, speech, and text processing.
2. Examples of sequence models include RNNs, LSTMs, GRUs, and Transformers, each designed for specific challenges.
3. Sequence models capture dependencies between elements, enabling predictions based on context.
4. Applications range from language translation and text generation to stock price prediction and speech recognition.

RNN in Depth

1. Recurrent Neural Networks (RNNs) are specialized for sequential data, processing inputs step-by-step while maintaining a memory of past information.
2. RNNs use hidden states to capture temporal dependencies, enabling predictions based on sequence history.
3. Ideal for tasks like text generation, speech recognition, and time-series forecasting.

Types of RNN

1. One-to-Many RNNs generate sequences from a single input, like caption generation from an image.
2. Many-to-One RNNs summarize sequences into a single output, such as sentiment analysis of a sentence.
3. Many-to-Many RNNs handle sequence input and output, such as machine translation or video frame labeling.

Vanishing Gradients

1. Vanishing gradients occur when gradients become too small during backpropagation, hindering effective weight updates.
2. It primarily affects deep networks with activation functions like sigmoid or tanh, leading to slow or stalled learning.
3. Layers closer to the input experience smaller gradients, causing them to learn much slower than deeper layers.
4. Solutions include using activation functions like ReLU, batch normalization, or architectures like LSTMs with gating mechanisms.
5. Addressing vanishing gradients is critical for training deep neural networks effectively and efficiently.

LSTM - Long Short Term Memory Networks

1. Long Short Term Memory network addresses short term memory problem in RNN by introducing long term memory cell (a.k.a cell state)
2. It has both short term and long term memory
3. It has 3 gates, Forget, Input and Output
4. LSTM, GRU etc. have become less popular in the shiny world of transformers

GRU - Gated Recurrent Unit Network

1. GRU combines long and short term memory in one cell
2. It is light weight, more efficient. LSTM however performs well on longer sequences
3. GRU has only two gates, Update and Reset