

循环神经网络(RNN、LSTM)

1、相关背景

1.1 序列数据

定义：数据点之间存在时间或顺序依赖关系。

常见示例：

- 文本：单词的顺序构成句子，句子的顺序构成段落。
- 语音：音素按时间顺序排列，形成单词和句子。
- 时间序列：股票价格、天气数据等，其值随时间变化。
- 视频帧：视频是图像帧的序列，每一帧都依赖于前一帧。
- 基因序列：DNA 碱基的顺序决定了基因的功能。

1.2传统神经网络的局限性：

- **无法处理序列数据:** 传统的前馈神经网络（如多层感知机 MLP）在处理序列数据（如文本、语音、时间序列等）时存在根本的缺陷。它们假设输入之间是相互独立的，无法捕捉序列中固有的时间依赖性和顺序信息。我们以“我爱中国”为例做机器翻译。在MLP中 x_1 、 x_2 、 x_3 、 x_4 之间没有顺序关系,换句话说,“我爱中国”和“中国爱我”都会被 MLP 视为相同的输入。
- **固定输入输出维度:** 传统神经网络的输入和输出维度通常是固定的，这使得它们难以处理变长的序列数据。例如，不同长度的句子或不同持续时间的语音信号，对于固定维度的网络来说是无法直接处理的。

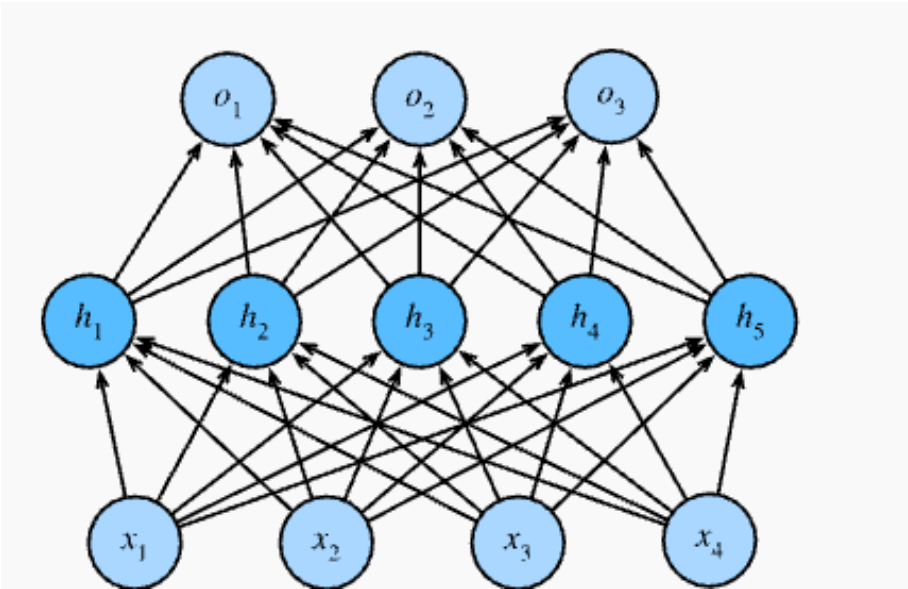


图1MLP结构

2、循环神经网络 (RNN)

2.1 定义

RNN 的核心思想是在处理序列中的每个元素时，将当前输入与先前时间步的隐藏状态相结合。隐藏状态充当网络的“记忆”，它可以捕获有关序列中先前元素的信息。

隐藏状态更新： $h_t = f(W_h x_t + U_h h_{t-1} + b_t)$

输出计算： $y_t = g(W_o h_t + b_o)$

- h_t 是时间步 t 的隐藏状态。
- x_t 是时间步 t 的输入。
- h_{t-1} 是上一个时间步 ($t-1$) 的隐藏状态。
- y_t 是时间步 t 的输出。
- W_h 、 U_h 和 W_o 是权重矩阵。
- b_h 和 b_o 是偏置向量。
- f 和 g 是激活函数。

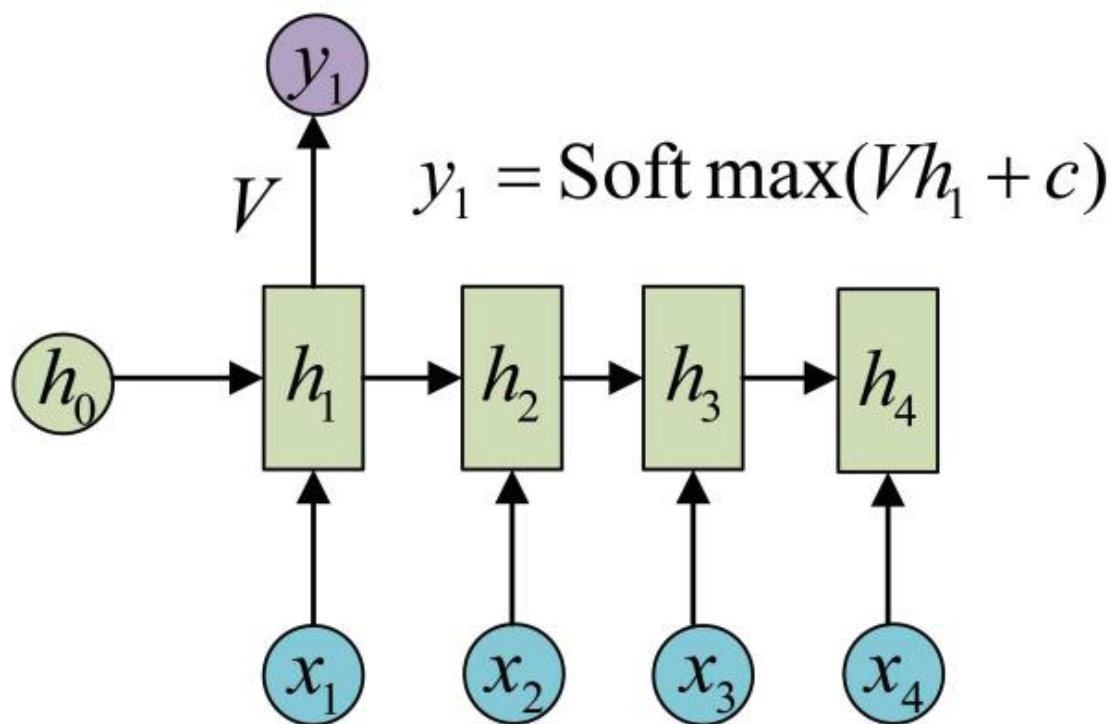


图2RNN展开结构

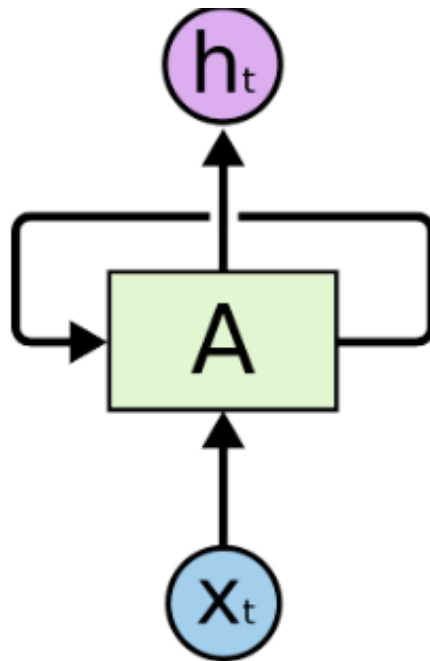


图3RNN循环结构

2.2 RNN的局限性

- 1、训练效率低
- 2、缺乏长期记忆

3、LSTM

3.1定义

LSTM 的核心是细胞状态（cell state），它就像一条传送带，携带信息在整个序列中传递。细胞状态可以通过称为“门”的结构进行修改。门可以选择性地允许信息通过或阻止信息通过。

LSTM 有三种主要的门：

- **遗忘门（Forget Gate）**：决定应从细胞状态中丢弃哪些信息。 $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
- **输入门（Input Gate）**：决定应将哪些新信息添加到细胞状态中。 $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$
- **输出门（Output Gate）**：决定应从细胞状态中输出哪些信息。 $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$

3.2 原理

1. **遗忘门**：遗忘门接收前一个隐藏状态和当前输入，并输出一个介于 0 和 1 之间的向量。该向量中的每个值表示应保留细胞状态中相应信息的程度（0 表示完全丢弃，1 表示完全保留）。
2. **输入门**：输入门接收前一个隐藏状态和当前输入，并输出两个向量：一个用于更新细胞状态的候选值向量，以及一个决定应将多少候选值添加到细胞状态中的向量。 $c_t^1 = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$
3. **细胞状态更新**：首先，细胞状态与遗忘门的输出逐元素相乘，丢弃不相关的信息。然后，将输入门选择的候选值添加到细胞状态中。 $C_t = f_t \odot C_{t-1} + i_t \odot c_t^1$
4. **输出门**：输出门接收前一个隐藏状态和当前输入，并输出一个介于 0 和 1 之间的向量。该向量与经过 \tanh 函数处理的细胞状态逐元素相乘，以确定要输出的最终信息。

5. 隐藏状态更新：最终的隐藏状态是输出门和经过 tanh 函数处理的细胞状态的乘积。 $h_t = o_t \odot \tanh(C_t)$

