

伯禹学习笔记篇叁

2020年2月18日 19:12

梯度消失和梯度爆炸

当神经网络的层数较多时，模型的数值稳定性容易变差

随机初始化模型参数

1. PyTorch的默认随机初始化
2. Xavier随机初始化

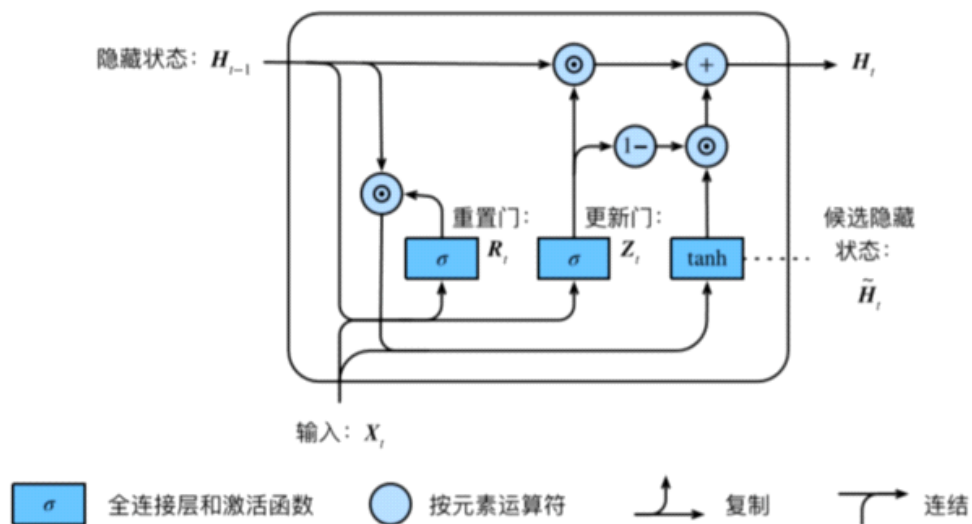
还有一种比较常用的随机初始化方法叫作Xavier随机初始化。假设某全连接层的输入个数为 a ，输出个数为 b ，Xavier随机初始化将使该层中权重参数的每个元素都随机采样于均匀分布

$$U\left(-\sqrt{\frac{6}{a+b}}, \sqrt{\frac{6}{a+b}}\right).$$

环境因素

1. 协变量偏移
2. 标签偏移
3. 概念偏移

门控循环神经网络



$$\begin{aligned} R_t &= \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r) \\ Z_t &= \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z) \\ \tilde{H}_t &= \tanh(X_t W_{xh} + (R_t \odot H_{t-1}) W_{hh} + b_h) \\ H_t &= Z_t \odot H_{t-1} + (1 - Z_t) \odot \tilde{H}_t \end{aligned}$$

- 重置门有助于捕捉时间序列里短期的依赖关系；
- 更新门有助于捕捉时间序列里长期的依赖关系。

长短期记忆网络

遗忘门

控制上一时间步的记忆细胞

输入门

控制当前时间步的输入

输出门

控制从记忆细胞到隐藏状态

记忆细胞:

一种特殊的隐藏状态的信息的流动

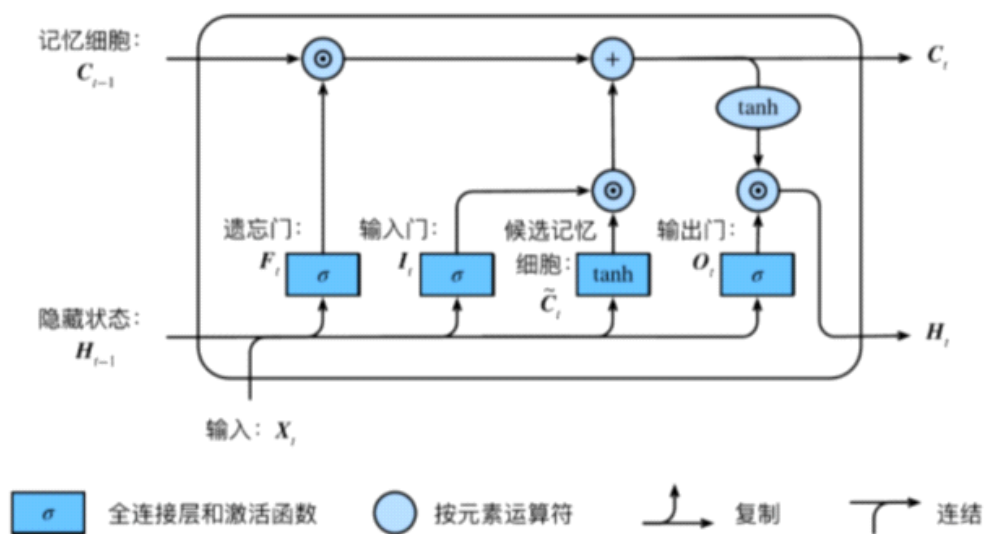
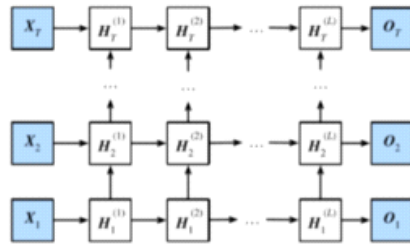


图 6.10: 长短期记忆中隐藏状态的计算。这里的 \odot 是按元素乘法

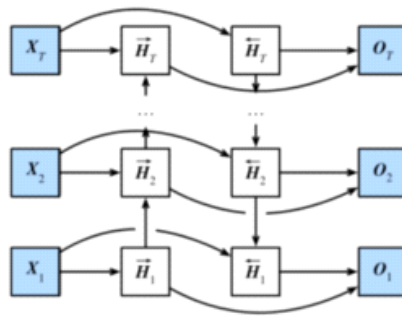
$$\begin{aligned} I_t &= \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i) \\ F_t &= \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f) \\ O_t &= \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o) \\ \tilde{C}_t &= \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c) \\ C_t &= F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \\ H_t &= O_t \odot \tanh(C_t) \end{aligned}$$

深度循环神经网络



$$\begin{aligned}
 H_t^{(1)} &= \phi(X_t W_{xh}^{(1)} + H_{t-1}^{(1)} W_{hh}^{(1)} + b_h^{(1)}) \\
 H_t^{(\ell)} &= \phi(H_t^{(\ell-1)} W_{xh}^{(\ell)} + H_{t-1}^{(\ell)} W_{hh}^{(\ell)} + b_h^{(\ell)}) \\
 O_t &= H_t^{(L)} W_{hq} + b_q
 \end{aligned}$$

双向循环神经网络



$$\begin{aligned}
 \vec{H}_t &= \phi(X_t W_{xh}^{(f)} + \vec{H}_{t-1} W_{hh}^{(f)} + b_h^{(f)}) \\
 \overleftarrow{H}_t &= \phi(X_t W_{xh}^{(b)} + \overleftarrow{H}_{t+1} W_{hh}^{(b)} + b_h^{(b)}) \\
 H_t &= (\vec{H}_t, \overleftarrow{H}_t) \\
 O_t &= H_t W_{hq} + b_q
 \end{aligned}$$