

李沐 · AWS

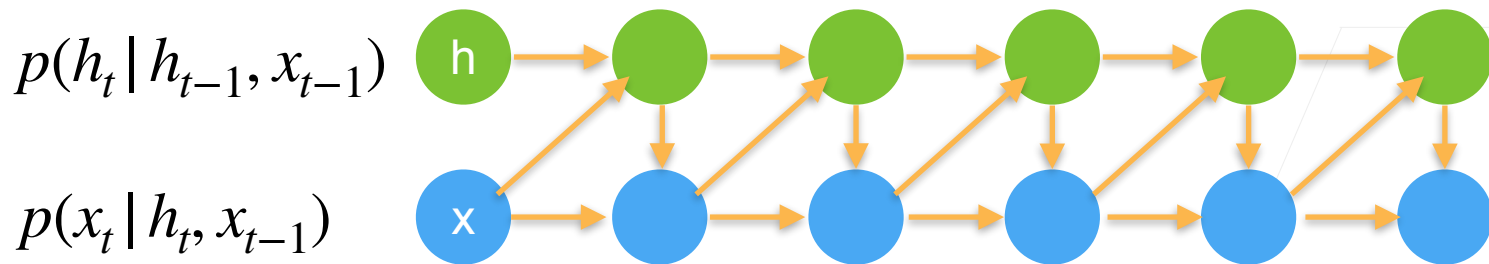
# 循环神经网络



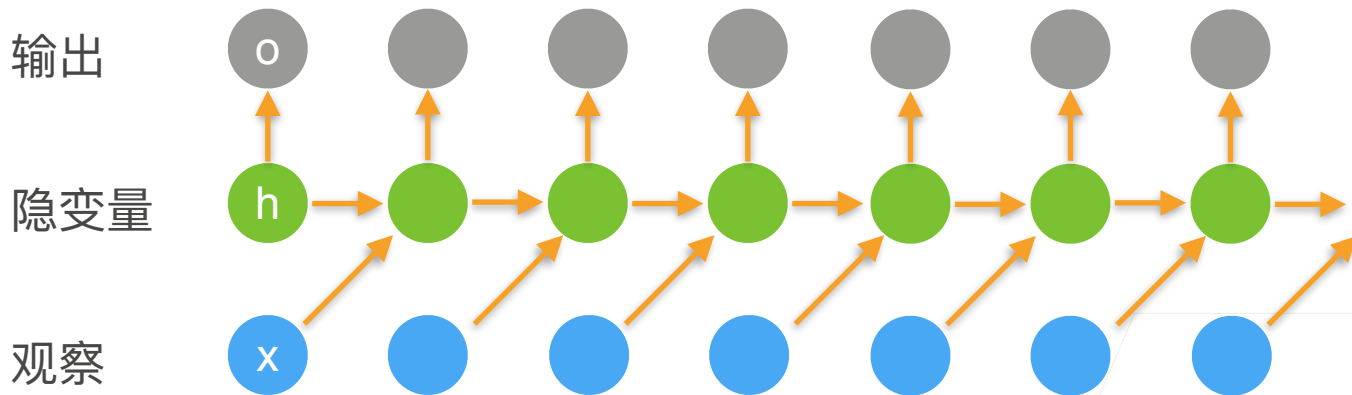
# 潜变量自回归模型



- 使用潜变量  $h_t$  总结过去信息



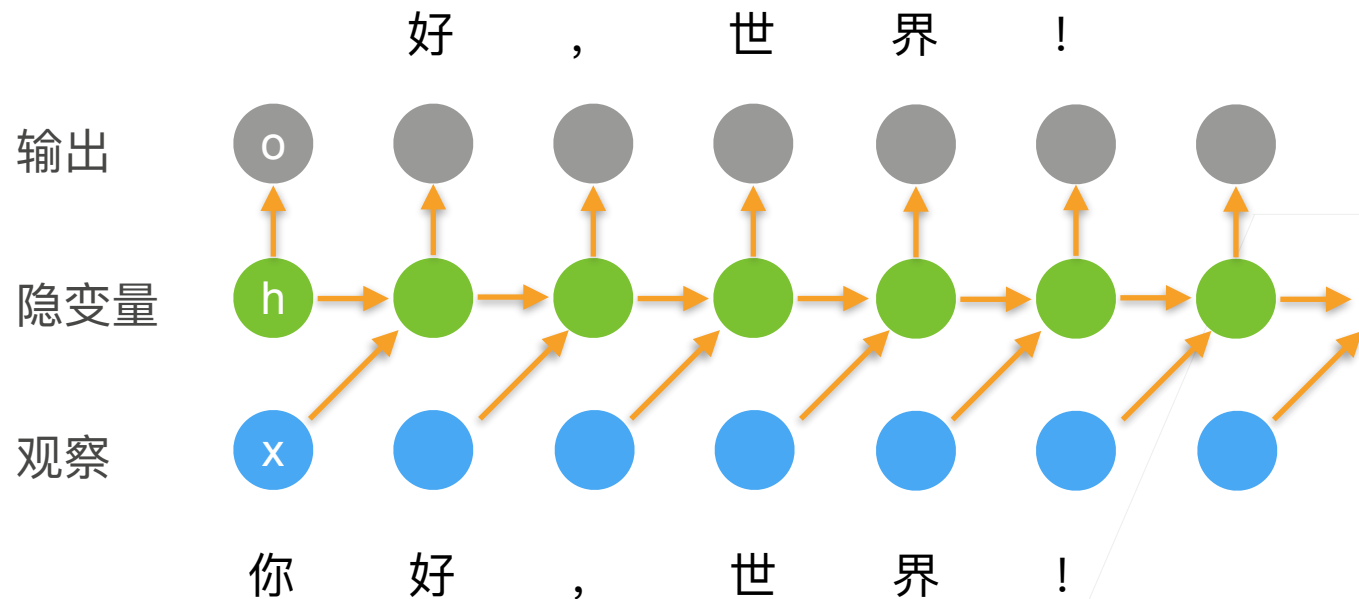
# 循环神经网络



- 更新隐藏状态:  $\mathbf{h}_t = \phi(\mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{W}_{hx}\mathbf{x}_{t-1} + \mathbf{b}_h)$
- 输出:  $\mathbf{o}_t = \phi(\mathbf{W}_{ho}\mathbf{h}_t + \mathbf{b}_o)$

去掉这一项就  
退化成MLP

# 使用循环神经网络的语言模型





# 困惑度 (perplexity)

- 衡量一个语言模型的好坏可以用平均交叉熵

$$\pi = \frac{1}{n} \sum_{i=1}^n -\log p(x_i | x_{1:i-1})$$

$p$  是语言模型的预测概率， $x_i$  是真实词

- 历史原因NLP使用困惑度  $\exp(\pi)$  来衡量，  
是平均每次可能选项
  - 1表示完美，无穷大是最差情况



# 梯度裁剪

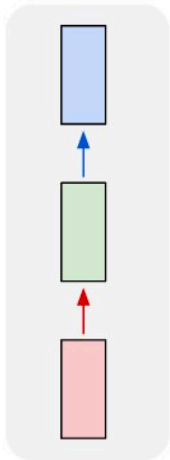
- 迭代中计算这 $T$ 个时间步上的梯度，在反向传播过程中产生长度为  $O(T)$  的矩阵乘法链，导致数值不稳定
- 梯度裁剪能有效预防梯度爆炸
  - 如果梯度长度超过  $\theta$ ，那么拖影回长度 $\theta$

$$\mathbf{g} \leftarrow \min \left( 1, \frac{\theta}{\|\mathbf{g}\|} \right) \mathbf{g}$$

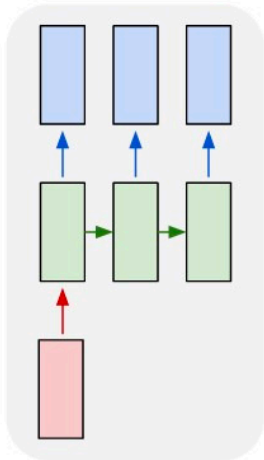
# 更多的应用 RNNs



one to one

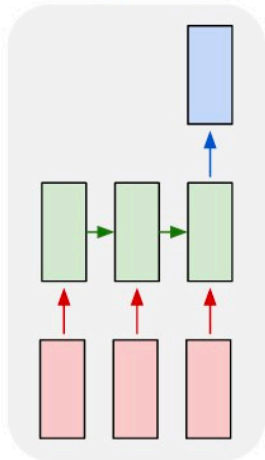


one to many



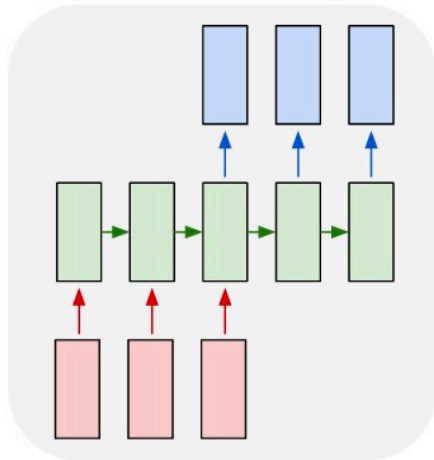
文本生成

many to one



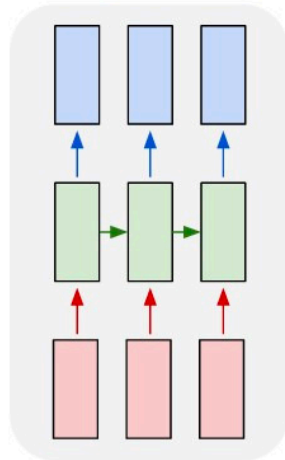
文本分类

many to many



问答、机器翻译

many to many



Tag生成

# 总结



- 循环神经网络的输出取决于当下输入和前一段时间的隐变量
- 应用到语言模型中时，循环神经网络根据当前词预测下一次时刻词
- 通常使用困惑度来衡量语言模型的好坏