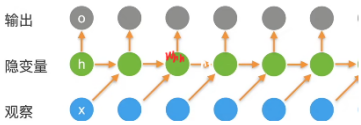
# RNN

- 潜变量自回归模型：使用潜变量ht总结过去信息 $P(h_t|h_{t-1}, x_{t-1})$；$P(x_t|h_t, x_{t-1})$
- 循环神经网络：




去掉就退化为MLP

$$h_t = \varphi(W_{hh} h_{t-1} + W_{hx} x_{t-1} + b_h)$$

$$o_t = W_{ho} h_t + b_o$$

- 困惑度（Perplexity）
  - 衡量一个语言模型好坏可以用平均交叉熵 $\pi = \frac{1}{n}\sum_{i=1}^{n} -\log p(x_t|x_{t-1}\cdots)$，n为 seq 长度
  - 历史原因NLP使用困惑度exp(π)来衡量，1表示完美，∞是最差情况
- 梯度剪裁
  - 迭代中计算T个时间步上的梯度，反向传播产生O(T)的矩阵乘法，导致数值不稳定
  - 梯度剪裁能有效预防梯度爆炸，如果提督长度超过θ，那么拖回长度θ
  - 这样做后梯度范数不会超过θ，且更新后的梯度仍与g原始方向对齐

$g \leftarrow \min\left(1, \frac{\theta}{\|g\|}\right) g$，g为所有层梯度

$\|g\|$ 代表梯度长度（$L_2$ 范数），超过θ就压缩

相当于将每个分量压缩了

也为模型赋与了稳定性