## 一般问题说明

1. 所有向量均为行向量。

## 获取embedding vectors



W为embedding vectors构成的矩阵，每行为一个embedding vector。编程时，可不必使用one-hot vector，而是输入i，i表示one-hot vector的第i个元素为1，然后直接x=W[ i ]，取W的第i行赋值给x。

## 计算LSTM













其中权重矩阵*W*\*被随机初始化；除遗忘门的偏差*bf*被初始化为全一向量外，其余偏差向量被初始化为零向量。对应的梯度可全部初始化为零。

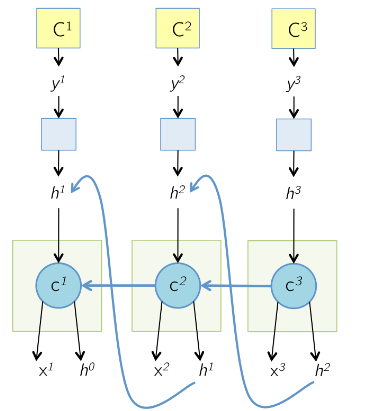
预测当前输出：



## Backward计算梯度

LSTM的梯度：<http://arunmallya.github.io/writeups/nn/lstm/index.html#/7>

The cell state at time T, *c*T receives gradients from *h*T as well as the next cell state *c*T+1. 类似地，*h*T从*i*T+1，*f*T+1，*o*T+1及*c\_hat*T+1接收梯度。所以，在时刻*t*，当计算各个梯度后，需要保存dct-1及dht-1，以便计算之前时刻的梯度。



## Seq2seq模型

Seq2seq模型是一种encoder-decoder结构模型。其中，encoder由embedding layer 和 lstm layer构成，decoder由embedding layer，lstm layer和softmax layer构成。Decoder需要使用softmax决定每一时刻的输出类别的概率分布。



Figure seq2seq

## 类变量-各个layer

**记录时刻。**在seq2seq模型中，显然输入是一个序列，输出也是一个序列。所以，在每一个layer中，我们需要对应的layer类中含有一个self.t整数变量成员记录当前forward或者backward进行到了哪里。

**各个layer还应记录在每一时刻该layer的输入和输出。**这些记录的输入输出变量被定义为字典类型，其key为时刻的记录，其value为对应时刻的输入输出值。以LSTM为例，LSTM在每一时刻需要记录：self.x = {}，self.h = {}，self.c = {}， self.ct = {}，self.input\_gate = {}， self.forget\_gate = {}，self.output\_gate = {}，self.cell\_update = {}。

**各个layer还需要记录对应的layer参数及他们的梯度。**以LSTM为例，它需要记录计算*it*，*ft*，*ot*，**的参数，共12类，包括8类*W*\*及4类*b*\*。

**在seq2seq模型中，各个layer主要涉及两种操作，forward()和backward()。**forward()用于进行predict()及在train()时计算各个时刻函数值。backward()用于计算loss function对各个参数的梯度。

## 编程时Input\_layers与Output\_layers的联系

如Figure 1所示，在seq2seq模型中，Input\_layers与Output\_layers通过两个LSTM串联在一起，其余层无直接联系。

在对整个seq2seq模型进行forward()操作时，Input\_layers中的LSTM给Output\_layers中LSTM提供初始的记忆细胞状态和隐藏细胞输出。

在对整个seq2seq模型进行backward()操作时，Output\_layers中的LSTM给Input\_layers中LSTM提供初始的记忆细胞状态的梯度和隐藏细胞输出的梯度。