**Lec12**

**一、理解**

Memory network分为四个模块：

I:将特征提取表现为I(x)。

G:更新记忆m并给出新的输入。

O:计算新的特征并得出新的输入与记忆。

R:解码输出特征并给出最后回答。

**二、memory network解决的问**

实现长期记忆（大量的记忆），并且实现如何从长期记忆中读取和写入，此外还加入了推理功能

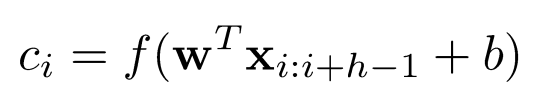
为什么长期记忆很重要：因为传统的RNN连复制任务都不行，LSTM估计也够玄乎。

在QA问题中，长期记忆是很重要的，充当知识库的作用，从其中获取长期记忆来回答问题

上面这个问题就是，当遇到有若干个句子并且句子之间有联系的时候,RNN和LSTM就不能很好地解决，以为是长期依赖，需要从记忆中提取信息

**Lec13**

**一、理解**

简单梳理一下，Lec13讲的就是用RNN对一句话进行提取特征进行分类或者翻译等等。还有一些细节和技巧的分析，还有一些句子模型的比较。

首先用word2vec or Glove对词进行初始化成k维词向量，课中设窗口数为3。将相邻的3个词向量进行串联，变成k\*3的矩阵，将权重W（3\*k维矩阵）的转制乘上这个矩阵再加上一个3\*3维的偏置矩阵得出的一个一维向量 作为一个卷积核（filter）提取的该句子的特征，利用多个卷积核得出一个一组，再用max-pooling进行降维，再经过一层softmax层得出最后用于处理的向量。

**二、拓展**

关于pooling，课中使用的max-pooling，缺点是保证特征的位置与旋转不变性，这在图片处理中是好的，但是在自然语言处理中就不好了，因为单词的位置会影响句子的意思。好处是MaxPooling能减少模型参数数量，有利于减少模型过拟合问题。再者，对于NLP任务来说，Max Pooling有个额外的好处；在此处，可以把变长的输入X整理成固定长度的输入。

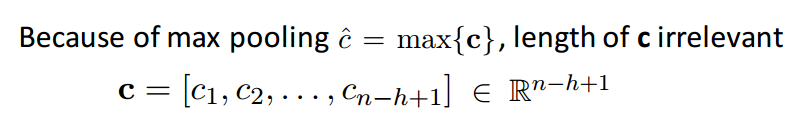
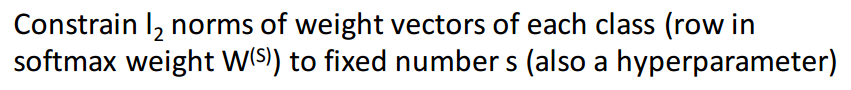
还有k-max-pooling和Chunk-Max Pooling两种pooling方法可以解决特征位置信息丢失的问题。

K-MaxPooling的意思是：原先的Max Pooling Over Time从Convolution层一系列特征值中只取最强的那个值，那么我们思路可以扩展一下，K-Max Pooling可以取所有特征值中得分在Top –K的值，并保留这些特征值原始的先后顺序。

Chunk-MaxPooling的思想是：把某个Filter对应的Convolution层的所有特征向量进行分段，切割成若干段后，在每个分段里面各自取得一个最大特征值，比如将某个Filter的特征向量切成3个Chunk，那么就在每个Chunk里面取一个最大值，于是获得3个特征值。

**三、问题：**

**Lec13**

1. max是用什么公式算出来的，只是单纯地求欧式距离吗？
2. 为什么与向量c的长度无关？
3. 没太明白