课程学习：通过先学习简单的样本，简单的代价函数，后学习困难的样本，困难的代价函数，不断提高模型效果，或是将容易和困难按比例结合在一起，随机提供给学习者，然后逐渐提高困难样本的比例，提高学习效果，在自然语言理解中损失函数定义比较模糊，是否可以采用这种方法。（这里书中提到了训练循环神经网络拜托长期依赖的问题）。

卷积神经网络

使用卷积运算了的神经网络

我感觉卷积在某种特定的情况下是没有丢失信息，起码数值信息是可能都在的，如果是下面这样的卷积核，但由于是特定情况，实际上基本是会丢失信息，与其说是丢失信息，不如说是一种对源域的抽象，抽象出较能代表源域的信息，去掉那部分在训练集中很难代表全体的信息，实际上这在个种机器学习算法中也很常见，只不过神经网络的抽象能力极强，能打到任意函数的抽象。

1 0 0 1 0 0 1 0

0 0 0 0 1 0 0 1

RNN

循环神经网络主要包含三种

（1）每个时间步都有输出，并且隐藏单元之间有循环连接的神经网络，、

（2）每个时间步都产生一个输出，只有当时刻输出到下一时刻隐藏单元之间存在循环连接的循环网络

（3）隐藏单元存在循环连接，但取整个序列后产生单个输出。

任何图灵可计算函数都可以通过一个有限维的循环网络计算。

图灵的基本思想是用机器来模拟人们用纸笔进行数学运算的过程，他把这样的过程看作下列两种简单的动作：

在纸上写上或擦除某个符号；，

把注意力从纸的一个位置移动到另一个位置；

而在每个阶段，人要决定下一步的动作，依赖于 (a) 此人当前所关注的纸上某个位置的符号和（b) 此人当前思维的状态。

这里的关键是每一时刻的书橱依赖于输入和人脑的状态，而人的思维状态具有记忆性，可以获得大量的信息。

RNN的思路就是使用h表示一个当前的状态，这一状态，由x前一时刻的状态h（t-1），这里对状态的可以通过矩阵W和U表示，如果W和U是一个近似可逆的矩阵，则基本可根据当前状态推断出前面每一时刻的输出，如果用输出o控制矩阵，信息量过少，输出信息过于简单，基本完全不可逆，难以达到捕捉历史信息的作用。与卷积网络相比，卷积网络是只使用梯度下降改变参数，求梯度获得是关于输入x的向量，也就是状态的变化是由输入和标签决定的。

这里将W可以看作状态转移矩阵，由当前状态推出下一状态，有点马尔可夫链的意思。