embedding

降维，把单词表中的每个单词对应到一个embedding\_dim维的连续向量。

处理数据时，先把每个句子处理成一个个单词（分词器），然后再将单词对应到相应的向量。

batch

每次选取一部分数据进行训练，从而减小内存压力。

神经元

，括号内称为净输入，非线性函数f称为激活函数，a称为神经元的活性值。

前馈网络

前后定序的神经层。

迭代公式：。

通用近似定理：任意连续函数，可以用一个前馈网络去近似。

卷积神经网络

，K为卷积核的大小。

在应用中，通常用。

可以实现一个节点信息（高度）的增大或减小。

卷积层，下一层的节点仅和上一层的部分节点关联，且该层间卷积核相同。

注意维度无法被卷积核大小整除时，需要在边界补零。

汇聚层，维度缩小为原来的。

在输出层之前，需要把数据展平（节点数目×单个节点的信息维度）。

记忆网络

神经元可以存储自己过去的信息。

循环神经网络

隐藏层活性值，其中为时刻的输入。

通用近似定理：任意非线性动力系统，可以用一个循环神经网络去近似。

循环神经网络是图灵完备的。