损失函数（Loss Function）和代价函数（Cost Function）的区别：

损失函数是针对单个样本定义的，而代价函数是对所有训练样本而言综合的度量。代价函数需要定义为一个凸函数以避免出现局部最优。

GPU与CPU的区别：

深度学习之所以耗费时间就是由于大量的循环延缓了程序的执行速度。Numpy等包的向量化操作实际是将循环体通过SIMD计算并行执行加快运算速率。而GPU比CPU更加擅长并行计算。

正则化（注意区别归一化/标准化）：

本质干的事就是通过调整代价函数来实现权重衰减，权重衰减的目的是减小网络中各个神经元的权重值，这么做使得整个网络的线性化程度加深，一定程度上削弱了网络拟合复杂映射的能力同时也抑制了过拟合的发生。常见的正则化方法有L1正则化、L2正则化、Drop正则化。所谓的L1和L2正则化，指的是在代价函数中人为的加入了一项有关参数w的一范数或二范数的项，以此来实现权重衰减。

Drop正则化：

对于每一层的输出，以一定概率将输出置为0，并将其他有效的未置0的输出除以（1-prob）。首先将输出置0简化了网络结构，这与正则化的基本思想是契合的。将有效输出除以（1-prob）是为了保证每一层输出的期望值保持不变。每一层的keep-prob值可以不同。

梯度消失和梯度爆炸：

层数的增加使得参数的微小变化极大的反映到输出中去。