权值向量

梯度向量

代价函数=损失函数(loss)

训练阶段,向机器送入一张图像,产生一个得分向量,每个元素对应一个分类. 我们希望真实类别在所有类别中得分最高(实际上在训练前是不可能发生的). 我们通过计算目标函数来度量输出得分与期望形式得分的差异(或距离),之后机器改变内部可调参数来降低这个目标函数。可调参数为实数，通常称为**权值**。被看成是定义及其输入-输出函数的“旋钮”。

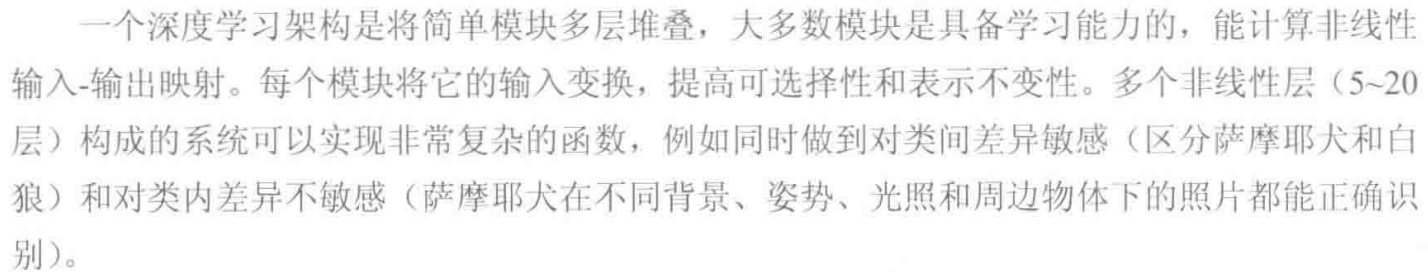
为了恰当调节权值向量，学习算法 计算梯度向量，表示每个权值增加一个微小值时目标函数的变化量，之后权值向量根据梯度向量的相反方向来调节。

目标函数在所有训练样本上取平均，可看作权值向量在高维空间的“丘陵平面”。负梯度向量指示当前平面最陡的下降方向，将目标函数带入距离最小值更近的地方，输出误差在平均意义上更低。

在实际中，大多数工程师陡使用随机梯度下降(Stochastic Gradient Descent,SGD)的方法。包括输入少量样本、计算输出和误差、计算这些样本的平均梯度、根据梯度调节权值。对训练集中的大量小样本集重复该过程，直到目标函数平均之停止下降。它之所以被称为SGD，是因为每个小样本子集提供了所有样本平均梯度的带噪声估计。相比更精确的优化技术，这个简单方法通常能更快的找到一组好的权值。

训练完成后，在一组不同的样本(称为测试集)上测试系统性能，目的是测试及其泛化能---它在训练阶段没有见过的新输入数据，产生合理输出的能力。

深度学习的有点，在于不需要设计特征提取器(以前是需要专家或者非常有经验的工程师进行特征提取的，提取关心的特征，去除不相关的因素)。

利用反向传播算法计算目标函数相对多层网络权值的梯度过程，其实就是《高等数学》中求导链式法则的工程应用。