

## 一、梯度优化方法

### 1、批量(batch)梯度降优化：

(1)、原理：**每一次迭代都需要所有样本的梯度参与计算**，求和平均后迭代更新权重和偏置

(2)、优缺点：

- 优点：全局最优解；易于并行实现；
- 缺点：当样本数目很多时，训练过程会很慢。

### 2、随机梯度降(SGD)优化：

(1)、原理：**每一次迭代只使用一个样本的梯度参与计算**，再迭代更新权重和偏置

(2)、优缺点：

- 优点：训练速度快；
- 缺点：准确度下降，并不是全局最优；不易于并行实现

SGD 算法对每一个样本都调整权重，所以在训练过程中，神经网络的性能是变化的。“随机”一词暗含了训练过程的随机性

### 3、小批量(Mini-batch)梯度降优化(最常用)

- 每一次迭代选出  $m$  个样本的梯度参与计算，求和平均后迭代更新权重和偏置

## 神经网络训练中的 Batch Size、和 Iteration、Epoch

• **Batch size**(批量大小)：即1次迭代使用的样本量

• **Iteration**(迭代)：1次iteration，即使用1个batch size样本训练网络参数1次

$$\text{总iteration} = \frac{\text{训练集大小}}{\text{batch size}}$$

• **Epoch** (期)：1个epoch表示遍历了1遍训练集中的所有样本

## 二、非线性激活函数作用（为什么使用非线性激活函数）

激活函数的主要作用是完成数据的非线性变换，解决线性模型的表达、分类能力不足的问题

线性激活函数无论叠加多少层，都是线性的，只是斜率和截距不同，叠加网络对实际问题没有多大帮助；如果没有非线性，深度神经网络的工作原理与线性变换相同；因为需要神经网络解决的实际问题基本都是非线性的。

**线性函数的问题在于,不管如何加深层数，总是存在与之等效的“无隐藏层的神经网络”。**这里我们考虑把线性函数  $h(x) = cx$  作为激活函数,把  $y(x) = h(h(h(x)))$  的运算对应 3 层神经网络 A。这个运算会进行  $y(x) = c \times c \times c \times x$  的乘法运算,但是同样的处理可以由  $y(x) = ax$  (注意, $a = c^3$ ) 这一次乘法运算(即没有隐藏层的神经网络)来表示。如本例所示,**使用线性函数时,无法发挥多层网络带来的优势**。因此，为了发挥叠加层所带来的优势，激活函数必须使用非线性函数。

## 三、CNN 各层的作用

### 1、卷积层：

卷积层通过**局部连接和权值共享**的方法提取一些**初级视觉特征**。局部连接指卷积层上的每一个神经元与前一层特征图中固定区域的神经元建立连接；权值共享指同一特征图中的神经元用一组相同的权重与前一层局部连接。

### 2、池化层：

池化层是将初级的视觉特征筛选并结合成更高级、抽象的视觉特征的过程。通过采样实现，经过池化层后，特征图数量不变，尺寸变小。因此池化主要功能是：

- 保留主要特征的同时减少参数和计算量。经过 MaxPool 可以减小卷积核的尺寸，同时又可以保留相应特征，所以主要用来降维。
- 在一定程度上能防止过拟合，由于这一层没有参数，不需要学习。
- 特征不变性，这种不变性包括 translation(平移)，rotation(旋转)，scale(尺度)。这就使网络的鲁棒性增强了，有一定抗扰动的作用。

#### (3)、全连接层 (fully connected layers, FC)

全连接层在整个卷积神经网络中起到“分类器”的作用，将学到的“分布式特征表示”映射到样本标记空间。

全连接相当于一个特征空间变换，可以把有用的信息提取整合，再加上激活函数的非线性映射，多层全连接层理论上可以模拟任何非线性变换，增强网络的非线性能力。

FC 层鉴别与高层特征最相关的类别，最终其权重用于得到不同类别的正确概率。

#### (4)、激活函数：增加非线性因素。

### 四、特征图计算公式

#### 卷积后输出特征图大小计算

输入图像大小： $H_{in} \times W_{in} \times n_c$

每个滤波器(卷积核)大小： $k \times k \times n_c$

滤波器(卷积核)个数： $K$

加边填充 padding:  $P$

卷积核滑动步幅(stride):  $S$

输出特征图大小：

$$\underbrace{\left( \frac{H_{in}-k+2P}{S} + 1 \right)}_{\text{高}} \times \underbrace{\left( \frac{W_{in}-k+2P}{S} + 1 \right)}_{\text{宽}} \times \underbrace{K}_{\text{通道数}}$$

计算卷积层输出特征图大小，当除不尽时，一般向下取整。

## 池化后输出特征图的大小计算

- 输入图像大小:  $H_{in} \times W_{in} \times n_c$
- 每个滤波器(卷积核)大小:  $k \times k \times n_c$
- 滤波器(卷积核)个数:  $K$
- 卷积滑动步幅(stride):  $S$
- 加边填充 padding:  $P$

□ 输出图像大小:

$$\left( \frac{H_{in}-k+2P}{S} + 1 \right) \times \left( \frac{W_{in}-k+2P}{S} + 1 \right) \times K$$

计算池化层输出特征图大小, 当除不尽时, 通常向上取整。

### 五、1\*1 卷积作用

1\*1 卷积可以控制输入特征图的深度, 可以减小或增大它, 或者在不改变深度时仅添加非线性。

融合不同特征图通道间同一位置的信息并且减小通道数。输出特征图的通道数与卷积核的个数有关, 当卷积核的个数小于输入特征图的通道数时, 可以起到减小通道数的作用, 同时将输入特征图的多通道特征通过单个卷积核进行卷积操作融合。

融合不同特征图通道间同一位置的信息并且增加非线性。当输入特征图和输出的特征图的通道数完全相同时, 1\*1 卷积核可以为输入数据的特征图增加非线性。

### 六、时间复杂度

### 七、空间复杂度