

3. 梯度下降与CNN

笔记本： 【课】原理-李宏毅 deep learning

创建时间： 2023/4/20 19:30

更新时间： 2023/4/21 15:34

作者： 1256876216@qq.com

URL: <https://baike.baidu.com/item/%E5%8F%8D%E5%90%91%E4%BC%A0%E6%...>

Gradient Descent

一、梯度下降Gradient Descent： 调整学习率

最优化原则：随着update次数变小，不同参数给不同的学习率

最优化方法：

- Adagrad方法，利用二次微分看平缓控制步调大小
- Stochastic Gradient Descent: LOSS函数只考虑一个观测值（非all）去更新参数，速度更快
- Feature Scaling: 特征归一化（根据一维特征的均值及方差，将各维控制统一分布到（均值0,标准差1）的分布上），例如二维是圆，三维是球。达到在学习率（步长）相同的情况下，各维保持相同更新速度，更快的迭代找到最优解

梯度下降缺点：卡在局部最优解、卡在斜率为0的平台

$$w^{t+1} \leftarrow w^t - \frac{\eta^t}{\sigma^t} g^t$$
$$\eta^t = \frac{\eta}{\sqrt{t+1}} \quad 1/t \text{ decay}$$
$$\sigma^t = \sqrt{\frac{1}{t+1} \sum_{i=0}^t (g^i)^2}$$
$$w^{t+1} \leftarrow w^t - \frac{\eta}{\sqrt{\sum_{i=0}^t (g^i)^2}} g^t$$

二、BP反向传播 BackPropagation

误差反向传播算法：通过将输出误差反传，将误差分摊给各层所有单元，从而获得各层单元的误差信号，进而修正各单元的权值（其过程，是一个权值调整的过程）

卷积神经网络应用

图片分类：

图片 => [Convolution => pooling] (repeat) => Flatten

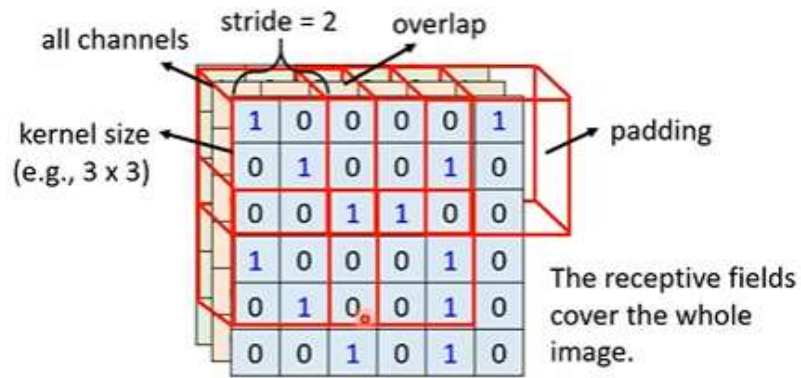
一、Convolutional Layer

1、常应用于图片相关领域

2、数据：图片RGB三通道channel，每个像素点以及周围局部数据做运算，每个局部称为 receptive field,

3、经典的receptive field 设置方式：

- 考虑全部通道，kernel size 一般3*3
- 每个receptive field通常有一组神经元维护
- 下一个receptive field通常与当前有重叠，偏移单位（可自行设置）为stride，重叠部分为overlap，边界补足部分为padding



4、共享参数，常见设置：

- receptive field共用权重weight,

二、polling 将图片缩小,减少运算量

- Max pooling：分块后选择区域内最大值
- mean pooling：选择均值

下围棋

棋盘：19*19维向量 => CNN（全连接也可以，效果差点） => 预测下一步最佳位置
用48个数字描述每个位置的状态（包括周围子的情况）