

算法面试

方鑫

2019 年 4 月 18 日

目录

1	数学基础	1
1.1	SGD,Momentum,Adagard,Adam	1
1.2	PCA原理	1
1.3	L1不可导如何处理	1
1.4	sigmoid函数特性	1
2	图像处理算法	2
2.1	图像中的锐化平滑操作	2
2.2	sift特征提取和匹配的具体步骤	2
3	机器学习算法	2
3.1	训练集中类别不平衡怎么处理?	2
4	深度学习	2

1 数学基础

1.1 SGD,Momentum,Adagard,Adam

SGD: 随机梯度下降, 每一次迭代计算数据集的一个数据就更新梯度, 然后对参数进行更新。

Momentum: 参考物理中动量的概念, 前几次的梯度也会参与到当前的计算中, 但前几轮的梯度叠加在当前计算中会有一定的衰减。

Adagrad: 在训练过程中可以自动变更学习的速率, 设置一个全局的学习率, 而实际的学习率与以往的参数模和的平方成反比。

Adam利用梯度的一阶矩估计和二阶矩估计动态调整每个参数的学习率, 在经过偏置的校正后, 每一次迭代后的学习率都有个确定的范围, 使得参数较为平稳。

1.2 PCA原理

PCA是一种线性降维方法, 通过将高维数据映射到低维数据中, 使得在新的投影上自身方差尽量大, 方差越大特征越有特效, 尽量使产生的新特征间相关性越小

PCA算法具体操作为对所有样本进行中心化操作, 计算样本的协方差矩阵, 然后对协方差矩阵做特征值分解, 取最大的 n 个特征值对应的特征向量构成投影矩阵

1.3 L1不可导如何处理

1.4 sigmoid函数特性

函数形式 $f(x) = \frac{1}{1+e^{-x}}$, 定义域 $(-\infty, +\infty)$, 值域 $(0, 1)$

2 图像处理算法

2.1 图像中的锐化平滑操作

2.2 sift特征提取和匹配的具体步骤

检测流程：对图像用不同的卷积核做卷积，得到不同的卷积结果（即不同的尺度空间），然后对得到的结果做差分。缩小图像比例，重复这一过程得到高斯差分金字塔。对金字塔每个点的邻域，搜索26邻域（三维），找到候选极值。

生成高斯差分金字塔

尺度空间构建

空间极值点检测

关键点精确定位

关键点主方向计算

描述子构造

3 机器学习算法

3.1 交叉熵

3.2 训练集中类别不平衡怎么处理？

类别不均衡指分类任务中不同类别的训练样例数目差别很大的情况。

类别不均衡的一个基本策略：再缩放。

正常情况下，如下情况下预测为正

$$\frac{y}{1-y} > 1 \quad (1)$$

若正反比例不相同，为了平衡正负样本，预测几率需要跟随观测几率做出改变

$$\frac{y}{1-y} > \frac{m^+}{m^-} \quad (2)$$

最终在预测时，根据下式得到结果

$$\frac{y'}{1-y'} = \frac{y}{1-y} * \frac{m^+}{m^-} \quad (3)$$

实际操作中，再缩放很困难，因为训练集是真实样本总体的无偏采样这个假设往往不成立。若反例比正例多，实际操作总体有三类做法：1）直接对训练集反类样例欠采样；2）对正类样例过采样；3）直接基于原始训练集进行学习，但在预测时乘上修正比例

4 深度学习