

机器学习重点总结

笔记本： 机器学习重点总结
创建时间： 2018/8/9 8:51
作者： 王

更新时间： 2018/8/29 15:08

1.简要介绍一下TensorFlow中的计算图

TensorFlow可以分为两个部分，一个是计算图的构造，另一个是计算图的执行，TensorFlow中的每一个结点都是一个计算

2.请问GBDT和XGBoost的区别是什么？

两者都是属于模型融合中的boosting的方法（voting, bagging, averaging, boosting, stacking），boosting每一次分类都更加关心

被分类错误的样本，给他们更多的权重，将这些弱分类器相加

区别：

- 1.损失函数是用泰勒展式二项逼近，而不是像GBDT里的就是一阶导数
- 2.对树的结构进行了正则化约束，防止模型过度复杂，降低了过拟合的可能性
- 3.节点分裂的方式不同，GBDT是用的基尼系数，XGBoost是经过优化推导后的

3.在k-means或kNN，我们是用欧氏距离来计算最近的邻居之间的距离。为什么不用曼哈顿距离？

曼哈顿距离只能计算水平与垂直方向的距离，有维度的限制，而欧式距离能够延伸到多维空间中去

4.简单说一下特征工程

俗话说，数据和特征决定了机器学习的上限，而模型和算法只是逼近这个上限而已。所以说特征工程是十分重要的，大多数情况下，

模型的精进工作是少数站在金字塔顶端的人在做的，大多数人（我们）在公司中做的都是一些简单的数据清洗工作，利用LR分析提取特征

（水里没有鱼的情况下，再好的渔具也没用，特征工程就是找有鱼的那片水域。）

5. LR与SVM的区别

联系：两者都可以用以处理分类问题，且一般都用于二分类问题

两者都可以添加L1, L2正则化

区别：（1）LR是参数模型，SVM是非参数模型（参数模型是指假设总体服从一个分部（伯努利分部），该分部由一些参数确定；非参数模型对于数据总体的

分部不做任何假设，只能在给定一定样本的情况下，依据非参数统计的方法进行推断）

（2）LR优化目的是使分隔面到所有点的距离的总和最小，但是SVM的优化目的是使到分隔超平面最近的两个点距离分隔超平面的距离进行大

6.LR与线性回归的区别与联系

逻辑回归与线性回归都是广义的线性回归，线性模型优化的目标函数是最小二乘，逻辑回归是似然函数

逻辑回归是一种将预测值限定在0-1范围内的一种线性回归模型

7.L1正则化和L2正则化的区别

L1正则化：各个元素的绝对值的和

L2正则化：各个元素平方和的1/2次方

L1正则化有利于形成一个稀疏的权值矩阵，有利于特征的选择，在一些不重要的特征上面的权值为0

L2正则化主要的作用是防止过拟合，提升模型的泛化能力

在一个就是L1正则化不可导，L2正则化可导

8.朴素贝叶斯的朴素怎么理解

因为它假设所有的特征是同样重要且相互独立、互不影响的，但是在自然界中是很不真实的，所以说它朴素

P(有女朋友|性格好，有钱，家教好)，朴素贝叶斯假设的是后面的几个条件是相互之间没有联系的

但是往往家教好性格就会好

9.为什么要进行归一化

- （1）加快梯度下降的收敛速度

(2) 可能会提高精度，减少异常值带来的影响

10.简述一个机器学习的项目过程

(1) 抽象问题为数学模型

了解是一个分类还是回归问题，对比以前相似的项目

(2) 获取与分析数据

数据是本质，对于数据的敏感是最重要的，数据决定了模型的上限

通过可视化的方法对数据进行统计分析

(3) 特征选择

同样是可以结合可视化的方法衡量一个特征的重要性，这一个阶段经验就比较重要

(4) 模型的选择

做不同的尝试，调整参数

(5) 模型的融合

(6) 上线测试

11.如何解决梯度消失和梯度爆炸问题

梯度消失一般出现的原因是网络太深，或者是选择了不合适的损失函数

梯度爆炸一般出现在初始化参数过大的时候

解决梯度消失与梯度爆炸的问题可以采用：

(1) 采用更好地激活函数 (Relu)

(2) batch normalization

(3) LSTM不会容易发生梯度消失，因为其中含有门的结构，具有一定的短期记忆功能

12.协方差与相关性

协方差：两个变量在变化的过程中是同向变化还是反向变化，你变大我变大，你变大我变小

相关系数：X与Y的协方差除以X的标准差和Y的标准差，相关系数相对于协方差，消除了两个变量变化幅度的影响，而只是反映了两个变量单位变化幅度的相似程度

13.线性分类器和非线性分类器的区别以及优劣

线性分类器速度快，编程简单，但是拟合的效果可能不会太好

非线性分类器编程复杂但是效果拟合能力强

SVM是线性的还是非线性的，取决于它的核函数是线性核还是高斯核

14.简述什么是生成对抗网络

GAN之所以是对抗的，是因为GAN的内部存在竞争关系，由一个generator和一个discriminator构成生成模型的任务就是生成一个实例来欺骗判别模型，判别模型判断一个实例是由生成模型生成的还是真实的实例，在对抗的过程中不断成长，最后达到预想的效果

15.梯度下降法找到的一定是下降最快的方向么？

不是的，它只是在目标函数当前点的切平面上下降最快的方向

16.梯度下降法的问题和挑战

(1) 梯度的计算困难，运算量太大

(2) 学习率的选择，太小会导致收敛太慢，太大导致不收敛

17.梯度下降法的分类

(1) 批梯度下降

(2) 随机梯度下降

(3) 小批量随机梯度下降

18.什么是最小二乘法

数学技术，通过最小化误差的平方和来寻找数据的最佳函数匹配，自然

最小二乘是在欧氏距离为误差度量的情况下，由系数矩阵所张成的向量空间内对于观测向量的最佳逼近点。

19.什么是表示学习？为什么深度学习是有效的？

当我们在解决一个问题的时候，如果能够得到对于原始数据更好的表达，往往可以得到更好的模型

提升原始数据表达的两种方法：

- (1) 特征学习：又称为表示学习，自动学习有用的数据特征
- (2) 特征工程：人为地处理数据，也称为数据清洗

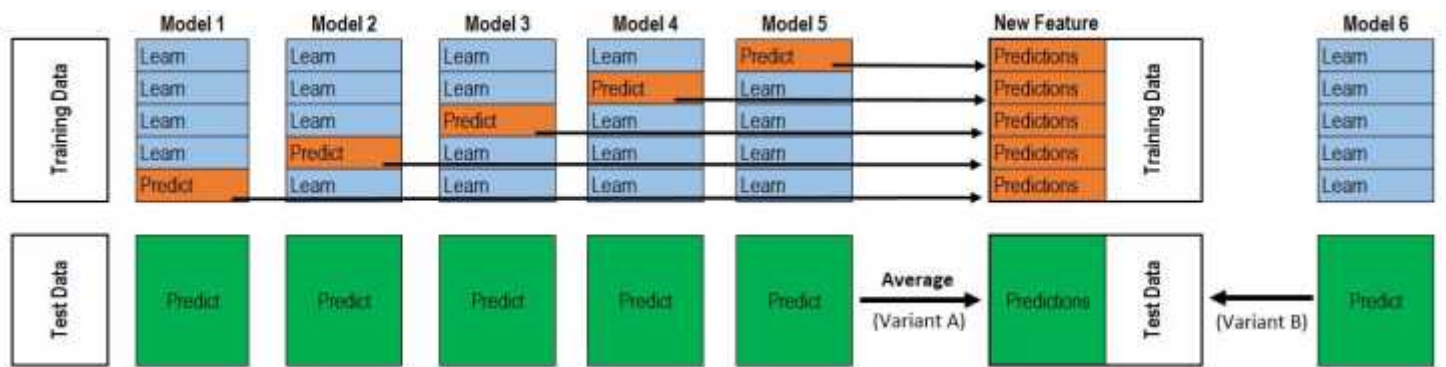
当我们谈及什么时候使用特征学习，什么时候使用特征工程的时候，我们往往基于假设的角度。当我们在选择一个模型的时候，我们会假设它的概率分布，假设输入与输出之间的关系。从这种角度上来看，这是一种先验，是一种贝叶斯模型。在小的数据集上面人的假设是一种强假设，当数据量变大的时候，数据越来越复杂，人们对于数据的假设也越来越肤浅，这时就需要采用特征学习的方法，自动学习特征。这一点也间接解释了为什么深度学习需要大量的数据。

20.模型融合

- (1) voting 采取投票的方式，票数多的种类获胜
- (2) averaging 一个简单的思路是求平均，再一个是加权平均
- (3) bagging 采用有放回的方法进行抽样，用抽样的样本对建立子模型，对子模型进行训练，重复多次，最后进行融合，分类问题采用voting，回归采用average（随机森林）
- (4) boosting 每一次迭代的时候都更加关心分类错误的样例，给这些分类错误的样例更大的权重，下一次迭代的目标就是能够更容易辨别出上一轮分类错误的样例，最终将这些弱分类器相加（AdaBoost, GBDT, xgboost）



- (5) stacking



21.为什么relu函数优于sigmoid函数

- (1) 计算导数的时候，涉及到除数的运算，sigmoid函数的运算量太大
- (2) 利用sigmoid函数进行训练，很容易出现梯度饱和的情况，导数趋向于0，这种情况会造成信息的丢失（防止梯度弥散）
- (3) relu会使得一些输出为0，这就造成了网络的稀疏性，减少了参数的相互依赖，缓解了过拟合问题的发生