

Branch: master

Algorithm_Interview_Notes-Chinese / A-机器学习 / A-机器学习基础.md

Find file Copy path

imhuay 结构调整

0c9535f on Sep 24 2018

1 contributor

152 lines (119 sloc) 9.79 KB

ML-机器学习基础

Index

- [偏差与方差](#)
 - [导致偏差和方差的原因](#)
 - [深度学习中的偏差与方差](#)
 - [偏差/方差 与 Boosting/Bagging](#)
 - [偏差与方差的计算公式](#)
 - [偏差与方差的权衡（过拟合与模型复杂度的权衡）](#)
- [生成模型与判别模型](#)
- [先验概率与后验概率](#)

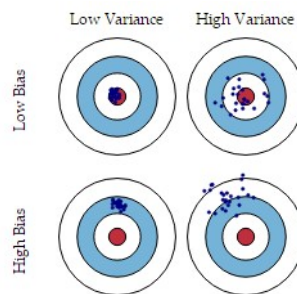
偏差与方差

《机器学习》2.5 偏差与方差 - 周志华

- **偏差与方差**分别是用于衡量一个模型**泛化误差**的两个方面；
 - 模型的**偏差**，指的是模型预测的**期望值**与**真实值**之间的差；
 - 模型的**方差**，指的是模型预测的**期望值**与**预测值**之间的差平方和；
- 在**监督学习**中，模型的**泛化误差**可分解为偏差、方差与噪声之和。

$$Err(x) = Bias^2 + Variance + Irreducible Error$$

- **偏差**用于描述模型的**拟合能力**；
方差用于描述模型的**稳定性**。



导致偏差和方差的原因

- **偏差**通常是由于我们对学习算法做了**错误的假设**，或者模型的复杂度不够；
 - 比如真实模型是一个二次函数，而我们假设模型为一次函数，这就会导致偏差的增大（欠拟合）；
 - **由偏差引起的误差**通常在**训练误差**上就能体现，或者说训练误差主要是由偏差造成的
- **方差**通常是由于模型的**复杂度相对于训练集过高**导致的；
 - 比如真实模型是一个简单的二次函数，而我们假设模型是一个高次函数，这就会导致方差的增大（过拟合）；
 - **由方差引起的误差**通常体现在测试误差相对训练误差的**增量**上。

深度学习中的偏差与方差

- 神经网络的拟合能力非常强，因此它的**训练误差**（偏差）通常较小；
- 但是过强的拟合能力会导致较大的方差，使模型的测试误差（**泛化误差**）增大；
- 因此深度学习的核心工作之一就是研究如何降低模型的泛化误差，这类方法统称为**正则化方法**。

../深度学习/正则化

偏差/方差 与 Boosting/Bagging

../集成学习专题/Boosting/Bagging 与 偏差/方差 的关系

偏差与方差的计算公式

- 记在**训练集 D** 上学得的模型为

$$f(\boldsymbol{x}; D)$$

模型的**期望预测**为

$$\hat{f}(\boldsymbol{x}) = \mathbb{E}_D[f(\boldsymbol{x}; D)]$$

- **偏差** (Bias)

$$bias^2(\boldsymbol{x}) = (\hat{f}(\boldsymbol{x}) - y)^2$$

偏差度量了学习算法的期望预测与真实结果的偏离程度，即刻画了学习算法本身的拟合能力；

- **方差** (Variance)

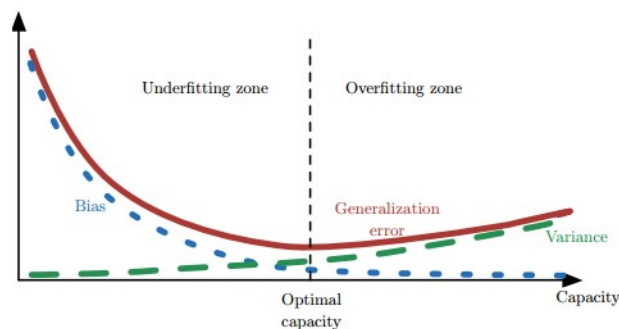
$$var(\boldsymbol{x}) = \mathbb{E}_D \left[\left(f(\boldsymbol{x}; D) - \hat{f}(\boldsymbol{x}) \right)^2 \right]$$

方差度量了同样大小的**训练集的变动**所导致的学习性能的变化，即刻画了数据扰动所造成的影响（模型的稳定性）；

- **噪声**则表达了在当前任务上任何学习算法所能达到的期望泛化误差的下界，即刻画了学习问题本身的难度。
- “**偏差-方差分解**”表明模型的泛化能力是由算法的能力、数据的充分性、任务本身的难度共同决定的。

偏差与方差的权衡（过拟合与模型复杂度的权衡）

- 给定学习任务，
 - 当训练不足时，模型的**拟合能力不够**（数据的扰动不足以使模型产生显著的变化），此时**偏差**主导模型的泛化误差；
 - 随着训练的进行，模型的**拟合能力增强**（模型能够学习数据发生的扰动），此时**方差**逐渐主导模型的泛化误差；
 - 当训练充足后，模型的**拟合能力过强**（数据的轻微扰动都会导致模型产生显著的变化），此时即发生**过拟合**（训练数据自身的、非全局的特征也被模型学习了）
- 偏差和方差的关系和**模型容量**（模型复杂度）、**欠拟合**和**过拟合**的概念紧密相联



- 当模型的容量增大（x 轴）时，偏差（用点表示）随之减小，而方差（虚线）随之增大
- 沿着 x 轴存在**最佳容量**，**小于最佳容量会呈现欠拟合**，**大于最佳容量会导致过拟合**。

《深度学习》5.4.4 权衡偏差和方差以最小化均方误差

Reference

- [Understanding the Bias-Variance Tradeoff](#)
- [机器学习中的Bias\(偏差\), Error\(误差\), 和Variance\(方差\)有什么区别和联系?](#) - 知乎

生成模型与判别模型

《统计学习方法》 1.7 生成模型与判别模型

- 监督学习的任务是学习一个模型，对给定的输入预测相应的输出
- 这个模型的一般形式为一个**决策函数**或一个**条件概率分布**（后验概率）：

$$Y = f(X) \quad \text{or} \quad P(Y|X)$$

- **决策函数**：输入 X 返回 Y ；其中 Y 与一个**阈值**比较，然后根据比较结果判定 X 的类别
- **条件概率分布**：输入 X 返回 X **属于每个类别的概率**；将其中概率最大的作为 X 所属的类别
- 监督学习模型可分为**生成模型与判别模型**
 - **判别模型**直接学习决策函数或者条件概率分布
 - 直观来说，**判别模型**学习的是类别之间的最优分隔面，反映的是不同类数据之间的差异
 - **生成模型**学习的是联合概率分布 $P(X, Y)$ ，然后根据条件概率公式计算 $P(Y|X)$

$$P(Y|X) = \frac{P(X, Y)}{P(X)}$$

两者之间的联系

- 由生成模型可以得到判别模型，但由判别模型得不到生成模型。
- 当存在“**隐变量**”时，只能使用**生成模型**
 - 隐变量：当我们找不到引起某一现象的原因时，就把这个在起作用，但无法确定的因素，叫“隐变量”

优缺点

- **判别模型**
 - 优点
 - 直接面对预测，往往学习的准确率更高
 - 由于直接学习 $P(Y|X)$ 或 $f(X)$ ，可以对数据进行各种程度的抽象，定义特征并使用特征，以简化学习过程
 - 缺点
 - 不能反映训练数据本身的特性
 - ...
- **生成模型**
 - 优点
 - 可以还原出联合概率分布 $P(X, Y)$ ，判别方法不能
 - 学习收敛速度更快——即当样本容量增加时，学到的模型可以更快地收敛到真实模型
 - 当存在“隐变量”时，只能使用生成模型
 - 缺点
 - 学习和计算过程比较复杂

常见模型

- 判别模型
 - K 近邻、感知机（神经网络）、决策树、逻辑斯蒂回归、**最大熵模型**、SVM、提升方法、**条件随机场**
- 生成模型
 - 朴素贝叶斯、隐马尔可夫模型、混合高斯模型、贝叶斯网络、马尔可夫随机场

Reference

- [机器学习---生成模型与判别模型](#) - CSDN博客
-

先验概率与后验概率

[先验概率](#)，[后验概率](#)，[似然概率](#)，[条件概率](#)，[贝叶斯](#)，[最大似然](#) - CSDN博客

条件概率（似然概率）

- 一个事件发生后另一个事件发生的概率。
- 一般的形式为 $P(X|Y)$ ，表示 y 发生的条件下 x 发生的概率。
- 有时为了区分一般意义上的**条件概率**，也称**似然概率**

先验概率

- 事件发生前的预判概率
- 可以是基于历史数据的统计，可以由背景常识得出，也可以是人的主观观点给出。
- 一般都是**单独事件**发生的概率，如 $P(A)$ 、 $P(B)$ 。

后验概率

- 基于先验概率求得的**反向条件概率**，形式上与条件概率相同（若 $P(X|Y)$ 为正向，则 $P(Y|X)$ 为反向）

贝叶斯公式

$$P(Y|X) = \frac{P(X|Y) * P(Y)}{P(X)}$$