

Machine Learning God Knows

Fduzjrqlw

目录

| | |
|---|----------|
| 1 机器学习理论 | 5 |
| 1.1 简述机器学习问题. | 5 |
| 1.2 误差分解公式. | 5 |
| 1.3 偏差和方差 | 5 |
| 1.4 风险 (risk), 经验风险 (empirical risk) 的含义. | 5 |
| 1.5 过拟合, 欠拟合和误差的关系. | 5 |
| 1.6 估计的无偏, 有效, 一致分别在说什么? | 5 |
| 1.7 为什么样本方差的分母是 $n - 1$? | 5 |
| 2 数据预处理 | 6 |
| 2.1 归一化 (normalization) 和标准化 (standardization), 以及它们的相同点和不同点. | 6 |
| 2.2 为什么要使用归一化/标准化? | 6 |
| 2.3 什么时候用归一化, 什么时候用标准化? | 6 |
| 2.4 一定要归一化吗? 举出一些不需要归一化的例子. | 6 |
| 2.5 处理缺失值一般有哪些方法? | 6 |
| 3 特征工程 | 6 |
| 3.1 特征抽取的目的. | 6 |
| 3.2 独热编码. | 6 |
| 3.3 TF-IDF 的公式和含义? | 6 |
| 4 类别不均衡 | 6 |
| 4.1 什么叫类别不平衡问题? | 6 |
| 4.2 类别不平衡问题用什么评估? | 7 |

| | | |
|-------|---|----|
| 4.3 | 解决不平衡问题有哪些方法? | 7 |
| 4.4 | 欠采样和过采样的缺点? | 7 |
| 5 | 性能度量指标 (metrics) | 8 |
| 6 | 正则化 | 8 |
| 6.1 | 做正则化的原因. | 8 |
| 6.2 | 模型复杂程度的评价指标. | 8 |
| 6.3 | 线性回归需要对偏差项 (bias term) 做正则吗? | 8 |
| 6.4 | 限制经验风险最小化 (Constrained ERM) v.s. 惩罚项经验风险最小化 (Penalized ERM) 的等价性. | 8 |
| 6.5 | L_1 正则的作用和原因. | 8 |
| 6.6 | L_2 正则的作用和原因. | 9 |
| 6.7 | 为什么要引入 Elastic Net 中的 L_1 和 L_2 的组合正则? | 9 |
| 6.8 | dropout 的原理和过程? | 9 |
| 6.9 | dropout 起到正则化作用和原因? | 9 |
| 7 | 优化算法 | 10 |
| 7.1 | GD,SGD 与 mini_batch GD 之间的区别和联系. | 10 |
| 7.2 | 什么时候用 SGD?mini_batch GD 中 batch_size 的选择. | 10 |
| 7.3 | 梯度下降算法的收敛率 (convergence rate). | 10 |
| 7.4 | SGD 的使用技巧. | 10 |
| 7.5 | SGD 的理论依据. | 10 |
| 7.6 | 如果测试集数据增加 100 倍,batch 大小 m 需要调整吗? | 10 |
| 7.7 | 在线学习的动机和方法. | 11 |
| 7.8 | 次梯度和次梯度下降算法. | 11 |
| 7.9 | SMO 算法 | 11 |
| 8 | 模型 | 11 |
| 8.1 | 支持向量机 (Supported Vector Machine) | 11 |
| 8.1.1 | SVM 的损失函数是什么? | 11 |
| 8.1.2 | SVM 的推导过程? | 11 |
| 8.1.3 | 为什么要引入对偶问题? 为什么要引入松弛变量 ξ ? | 11 |
| 8.1.4 | Slater 条件是什么, 如何验证 SVM 问题满足 Slater 条件? | 11 |

| | | |
|--------|--|----|
| 8.1.5 | 叙述 SVM 问题的 KKT 条件? | 11 |
| 8.1.6 | 模型偏置 bias term b 的作用? | 12 |
| 8.1.7 | 什么叫核方法? | 12 |
| 8.1.8 | 核化 (kernelized) 的好处有哪些? | 12 |
| 8.1.9 | 表示定理以及其作用? | 12 |
| 8.1.10 | 如何验证核函数? | 12 |
| 8.1.11 | 为什么叫径向基函数? | 12 |
| 8.1.12 | RBF 核对应特征空间的维数. | 12 |
| 8.1.13 | 正则化 RBF 核. | 13 |
| 8.1.14 | 参数 γ 以及 C 的作用 | 13 |
| 8.1.15 | 叙述 SVM 问题的 KKT 条件? | 13 |
| 8.1.16 | 什么是 SVM 的退化? | 13 |
| 8.1.17 | 如何构造新的输入数据, 使得一个线形可分的 SVM 问题退化? | 13 |
| 8.1.18 | SVM 的优化算法 | 13 |
| 8.1.19 | 核函数的选择 | 14 |
| 8.1.20 | SVM 的优缺点 | 14 |
| 8.2 | 感知机 (perceptron) | 15 |
| 8.2.1 | 感知机的工作原理? | 15 |
| 8.2.2 | 感知机与 SVM 的异同点有哪些? | 15 |
| 8.2.3 | 感知机如何优化? | 15 |
| 8.2.4 | 感知机解的性质 | 15 |
| 8.3 | 决策树 (Decision Tree) | 15 |
| 8.3.1 | 简述决策树的工作原理. | 15 |
| 8.3.2 | 决策树区域的特性. | 16 |
| 8.3.3 | 叙述决策树的建树过程? | 16 |
| 8.3.4 | 不纯度是什么? | 16 |
| 8.3.5 | Gini 值的含义? | 16 |
| 8.3.6 | Gini 值和熵的联系? | 16 |
| 8.3.7 | 什么是多变量决策树? | 17 |
| 8.3.8 | ID3 决策树的特点有哪些? | 17 |
| 8.3.9 | C4.5 决策树有哪些细节? | 17 |
| 8.3.10 | CART 有哪些细节? | 18 |

| | | |
|--------|---|----|
| 8.3.11 | 决策树的缺失值问题和解决策略? | 20 |
| 8.3.12 | 决策树的剪枝策略有哪些? | 20 |
| 8.3.13 | ID3,C4.5,CART 之间的差异有哪些? | 20 |
| 8.3.14 | sklearn 关于决策树的缺失值策略如何实现的? | 20 |
| 8.3.15 | 关于决策树的调参, 有哪些技巧? | 20 |
| 8.4 | Bagging | 20 |
| 8.4.1 | 统计量 (statistic) 与点估计 (point estimator) | 20 |
| 8.4.2 | 估计量的方差有什么含义? | 20 |
| 8.4.3 | 如何平衡每组的样本个数与组数 | 20 |
| 8.4.4 | 自助采样 (bootstrap) | 20 |
| 8.4.5 | Bagging 为什么用平均法? 原理是什么? | 21 |
| 8.4.6 | 包外误差估计 (Out-of-Bag Error Estimation) 是什么? | 21 |
| 8.4.7 | Bagging 的适用性? | 21 |
| 8.4.8 | bootstrap 的方差估计 | 21 |
| 8.5 | 随机森林 (Random Forest) | 22 |
| 8.6 | Boosting | 22 |
| 8.6.1 | 简单介绍下 boosting? | 22 |
| 8.7 | Adaboost | 22 |
| 8.7.1 | Adaboost 算法的流程及其推导? | 22 |
| 8.7.2 | Adaboost 的损失函数为什么是指数函数? | 22 |
| 8.7.3 | 分类替代函数 | 22 |
| 8.7.4 | Adaboost 中权重的含义? | 23 |
| 8.7.5 | 如何求加权最优化问题? | 23 |
| 8.7.6 | Adaboost 算法终止的条件? | 23 |
| 8.7.7 | Adaboost 训练集误差的上界估计? | 23 |
| 8.7.8 | 为什么 Adaboost 当训练集上 error 变为 0 时, 继续训练可以降低测试集上的 error? | 23 |
| 8.7.9 | Adaboost 什么情况下会过拟合严重? | 23 |
| 8.7.10 | 如何缓解 Adaboost 的过拟合现象? | 23 |
| 8.7.11 | 前向分布算法和坐标梯度下降法的联系和区别? | 24 |