

机器学习 2024 级期末考试回忆版

线性回归(3)

$$L = \frac{1}{m} \sum_{i=1}^m (y_i - (w^T x_i + b))^2$$

- (1) 解释 L 的含义,然后求 $\nabla_w L(w, b)$
- (2) 给定学习率 ν ,写出更新权重 w 的公式并说明学习率太小或者太大的缺点
- (3) 一般会对 L 进行 L2 正则化,问正则化的作用,对模型有什么影响

SVM(3)

样本序号	x_1	x_2	标签
1	2	3	-1
2	3	5	1
3	5	1	-1
4	6	4	1

- (1) 假设已经使用 SVM 计算出 $\omega = [1, 1]$, $b = -7$, 写出决策边界的表达式,计算所有点到边界的距离,指出支持向量。
- (2) 如果将点 3 的(5,1)改成(4,2),决策边界发生什么改变 (不需要计算 ω 和 b)
- (3) 线性核和高斯核对比下,高斯核函数的优点

集成学习(5)

Boosting 用了决策桩($x \leq c$ 的为 一类, $x > c$ 的为 一类)

坐标点 $x=1,3,5$;其中 1,5 为 $y=-1$,3 为 $y=+1$;

- (1) 求数据点的初始权重
- (2) 画出建第一个弱分类器的一个可能的决策边界,标出正侧,负侧
- (3) 根据(2),圈出在构建第二个弱分类器的时候权重变大的点.
- (4) 能不能训练出在训练集上完美分类的 boosting 分类器?能的话说明迭代次数,不能的话说明原因.
- (5) boosting 分类器的损失函数是否单调递减,说明理由.

PCA(3)

假设对 $\{x_i\}_{i=1}^K$ 做 PCA 得到 u_1, u_2, \dots, u_m 个主成分向量

- (1) 主成分向量 u_1, u_2, \dots, u_m 等于什么
- (2) 给出 PCA 的两个解释视角
- (3) 如何用这些主成分 u_1, u_2, \dots, u_m 表示数据 x 能够使误差最小? 并证明。

EM(4)

隐变量模型形如 $p(x) = p(x|z)p(z)$

- (1) 简述 EM 算法中的 E 步和 M 步，并给出公式
- (2) 对 GMM 用 EM 算法，求 Q 函数的表达式（无需展开高斯分布）
- (3) 假设已经有一个训练好的 GMM，参数为 $\theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^K$ 。考虑新样本点 x_n ，如何判断应该划分到哪个簇？
- (4) 有时候Q函数没有闭式表达，怎么解决，至少给出两种解决方法

推荐系统(3)

题目描述是用户的协同过滤

- (1) 给出用户协同过滤的过程描述,以及相似度表达公式
- (2) 用户协同过滤冷启动问题怎么解决
- (3) 说出两个评价指标解释含义