支持向量机

1. 前言
2. 采用基本形式的SVM分类器就能得到低错误率的结果，SVM能够对训练集之外的数据点作出很好的分类决策。
3. 这里，我们仅仅关注序列最小优化算法SMO（Sequential Minimal Optimization）,之后介绍如何使用一种称为核函数的方式将SVM扩展到更多数据集上。
4. 基于最大间隔分隔数据
5. 缺点：对参数调节和核函数的选择敏感，原始分类器不加修改仅仅适用于处理二类问题
6. 用于将数据集分隔开来的直线或面称为分隔超平面，我们希望找到离分隔超平面最近的点，确保它们离分隔面的距离尽可能远。这里点到分隔面的距离被称为间隔。
7. 支持向量就是离分隔超平面最近的哪些点。
8. 寻找最大间隔
9. 点到直线(Ax+by+c)的距离公式：
10. 因此可以得到点A到分隔超平面的距离是：
11. 使用类似单位阶跃函数对作用得到，当
12. 当确定分隔面的放置位置时，就可以计算点到分隔面的距离，间隔通过来计算。现在的目标就是找出分类器定义中的w和b。所以，首先找到具有最小间隔的数据点，也就是支持向量。一旦找到具有最小间隔的数据点，我们就需要对该间隔最大化。写成公式为：
13. 直接求解这个问题将相当困难，可以把它拆开成多个步骤，分开求解。先考虑最里层的内容：，可以固定其中一个因子然后最大化其他因子。约束条件是：，可以使用拉格朗日乘子法来解决这个问题。
14. SVM的大部分时间都源自训练，该过程主要实现两个参数的调优。而且SVM本身是一个二分类器，多分类问题应用SVM需要对代码做一些修改。
15. SMO高效优化算法