

使用支持向量机（SVM）和不同损失函数的线性分类器——大作业

1

问题描述：

手动实现线性分类模型，分别使用 Hinge Loss 和 Cross-Entropy Loss 进行比较。

除此之外，使用多项式实现一个分类器，画出拟合曲线，并且比较不同参数量下的结果。

此外，训练一个采用不同核函数的 SVM 二分类器。

请根据以下要求完成实验并提交报告和代码。

数据集：

UCI 的乳腺癌数据集 (Breast Cancer Wisconsin)

数据集地址：[UCI Breast Cancer Dataset](#)

任务要求：

1. 实现一个线性分类器，使用两种不同的损失函数：

Hinge Loss （对应 SVM 的损失函数）

Cross-Entropy Loss （对应于逻辑回归的损失函数）

比较两种损失函数在分类任务中的表现。

2. 使用多项式回归实现一个分类器，并比较不同参数量下的结果，从中体会什么是过拟合。

3. 支持向量机 SVM：选择两种核函数来训练 SVM 模型：

线性核函数

高斯核函数（RBF）

2 和 3 可以使用现成的 SVM 软件包，例如 Python 的 scikit-learn 库。

实验报告（内容包括但不限于以下部分）：

i. 比较两种损失函数（Hinge Loss 和 Cross-Entropy Loss）在**线性分类模型**中的表现。

ii. 比较多项式分类器不同参数情况下的情况，**可视化拟合曲线**。

iii. SVM 模型的一般理论：简要介绍 SVM 的基本原理，包括支持向量、间隔最大化、核函数等内容。

iv. 不同核函数的 SVM 模型与性能比较：比较线性核函数和高斯核函数在给定数据集上的分类性能，并分析结果。

v. SVM 的训练过程：详细描述训练过程，包括：

模型初始化方法

超参数选择（例如惩罚参数 C、核函数参数等）

使用的训练技巧（例如正则化、梯度下降等）

vi. 实验结果分析与讨论：包括模型的准确率、精确率、召回率、F1 值等评价指标，并进行分析和讨论。

提交要求：

将实验报告（pdf 格式）和代码（不要包含数据集）打包成一个压缩包。

压缩包文件名格式：学号+姓名+大作业 1

提交至邮箱 jzk1310153668@163.com

截止时间:

提交截止日期: 2024.10.20 24: 00

工具和技术:

编程语言: Python

必要库: scikit-learn, numpy, pandas, matplotlib