

# 一种有效求解非凸正则化线性 支持向量机的并行与分布式方 法

汇报人：

2024-01-21





# CONTENTS

- 引言
- 非凸正则化线性支持向量机模型
- 并行计算技术在求解中的应用
- 分布式计算技术在求解中的应用
- 实验结果与分析
- 总结与展望



01

引言



# 研究背景和意义

非凸正则化线性支持向量机 ( Non-convex Regularized Linear Support Vector Machine , 简称NR-LSVM ) 在分类、回归等机器学习任务中具有广泛应用。

随着数据规模的扩大和计算资源的增加，传统的串行求解方法已无法满足大规模数据集的处理需求，因此研究并行与分布式求解方法具有重要意义。

并行与分布式求解方法能够充分利用计算资源，提高求解效率，为处理大规模数据集提供有力支持。

$$\frac{\bar{X} - \bar{Y}}{S_p} \sim F_{n_x-1, n_y-1}$$
$$\sigma^2 = \sigma^2 \text{ (unknown),}$$
$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{S_p^2 \left( \frac{1}{n_x} + \frac{1}{n_y} \right)}} \sim$$



# 国内外研究现状及发展趋势

目前，国内外学者在并行与分布式求解NR-LSVM方面已取得一定成果，如基于MapReduce、Spark等框架的求解方法。

然而，现有方法在处理大规模数据集时仍存在问题，如通信开销大、收敛速度慢等。

未来发展趋势将更加注重算法的优化和加速，以及针对不同应用场景的定制化求解方法。





## 本文主要工作和贡献

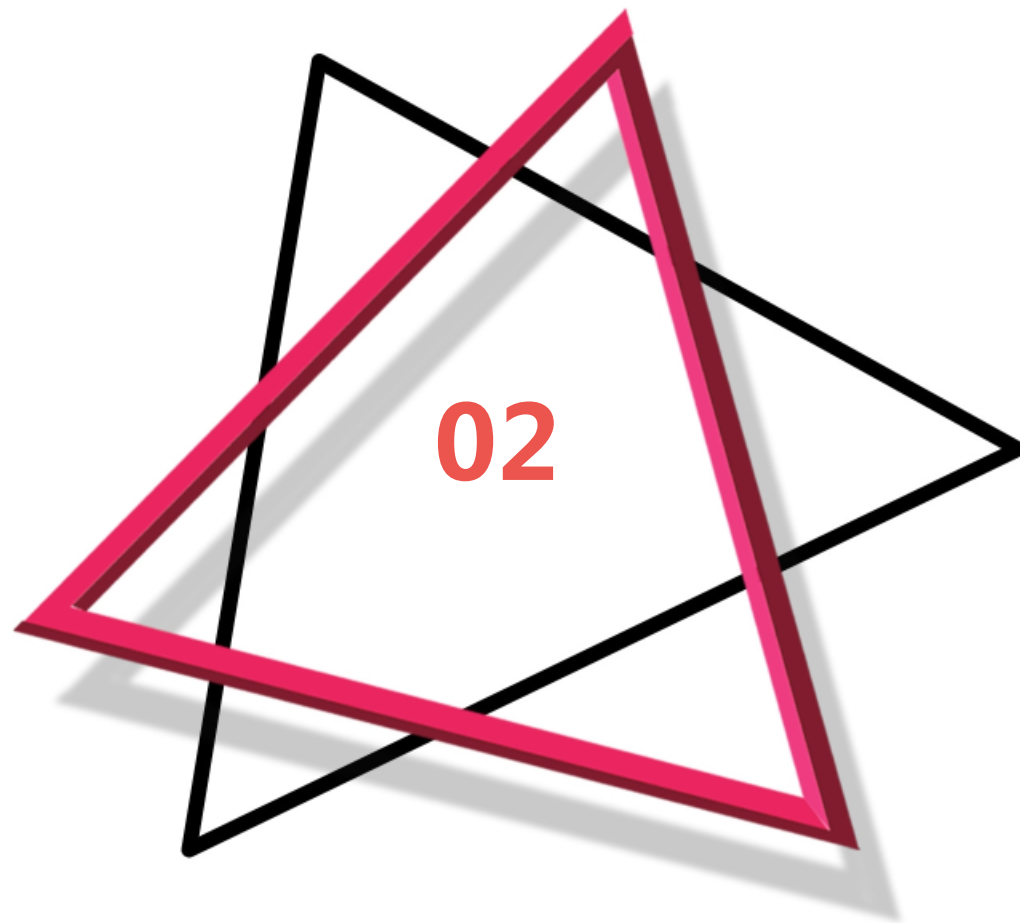


本文提出了一种基于分布式计算的NR-LSVM求解方法，通过设计合理的并行策略和优化算法，实现了高效求解。

在理论方面，本文对所提方法进行了详细的分析和证明，包括收敛性、复杂度等方面的讨论。



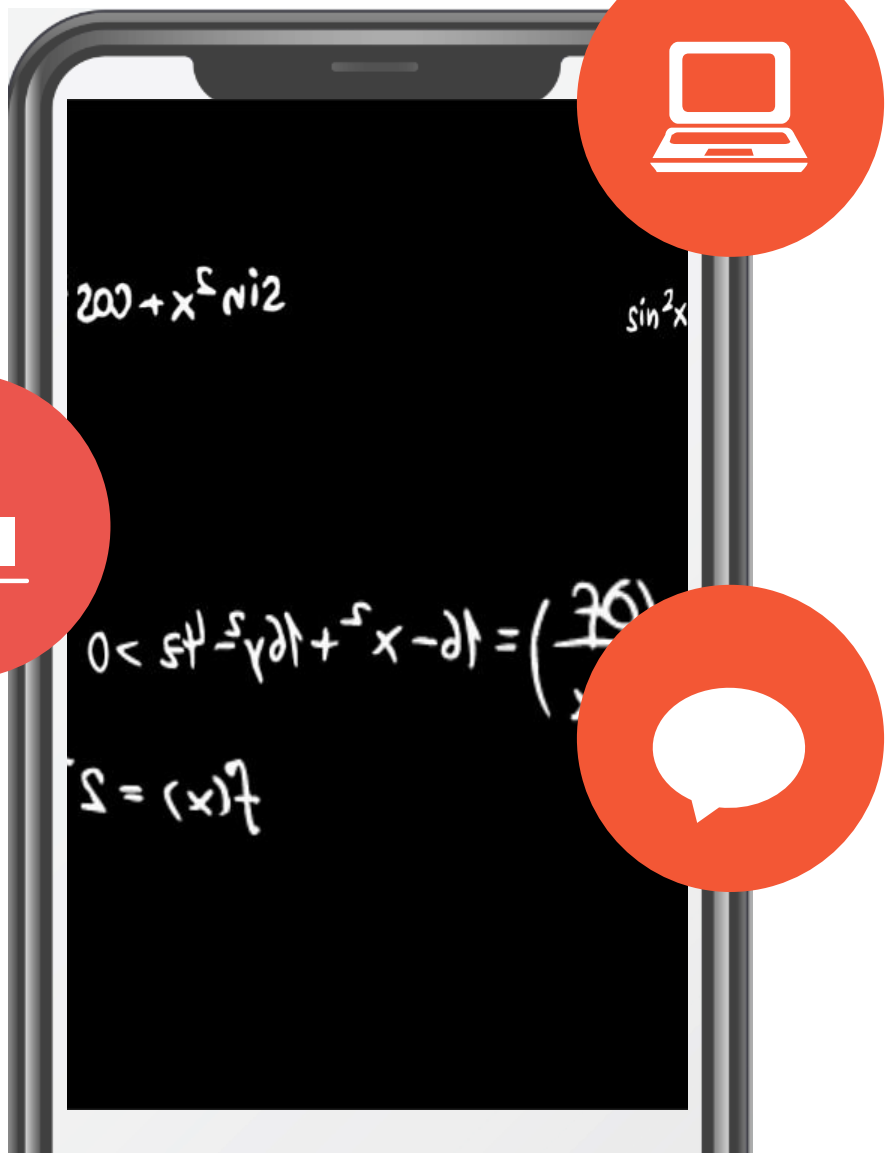
在实验方面，本文对所提方法进行了充分的验证和评估，通过与现有方法的对比实验，证明了所提方法的有效性和优越性。



## 非凸正则化线性支持向量机模型

# 线性支持向量机模型介绍

线性支持向量机 ( Linear Support Vector Machine , 简称LSVM ) 是一种广泛应用于分类和回归问题的机器学习算法。



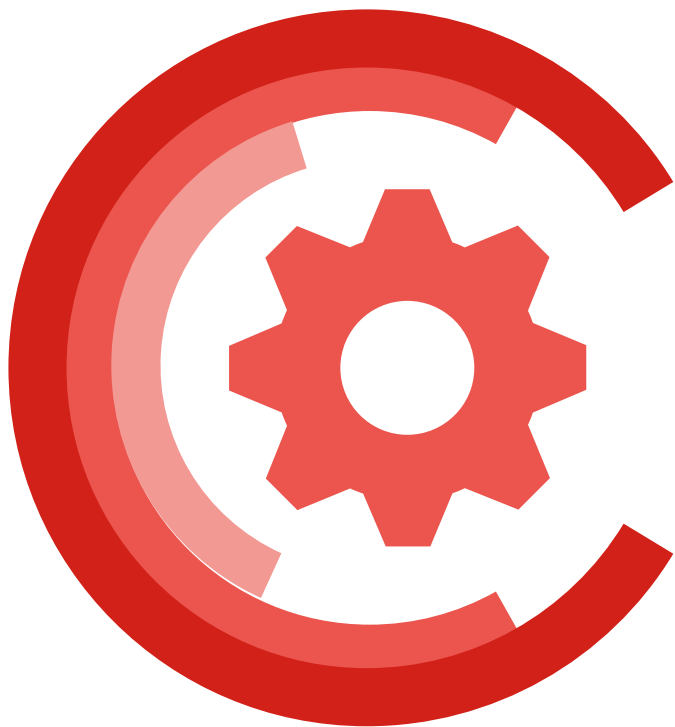
LSVM通过在高维空间中寻找一个超平面,使得不同类别的样本在该超平面上的投影最大化间隔,从而实现分类或回归。

LSVM具有简单、高效、易于理解和实现等优点,在文本分类、图像识别、生物信息学等领域得到了广泛应用。





# 非凸正则化项引入及作用



## 01

非凸正则化项是指那些不满足凸函数性质的正则化项，如L0范数、L1/2范数等。

## 02

引入非凸正则化项的目的是为了增强模型的稀疏性和特征选择能力，使得模型更加关注于那些对分类或回归结果有重要影响的特征。

## 03

非凸正则化项可以使得模型的解更加稀疏，即更多的特征权重被压缩为零，从而实现特征选择。



# 模型优化目标和约束条件

01

非凸正则化线性支持向量机的优化目标是 minimized 经验风险（即训练误差）和正则化项的和。



02

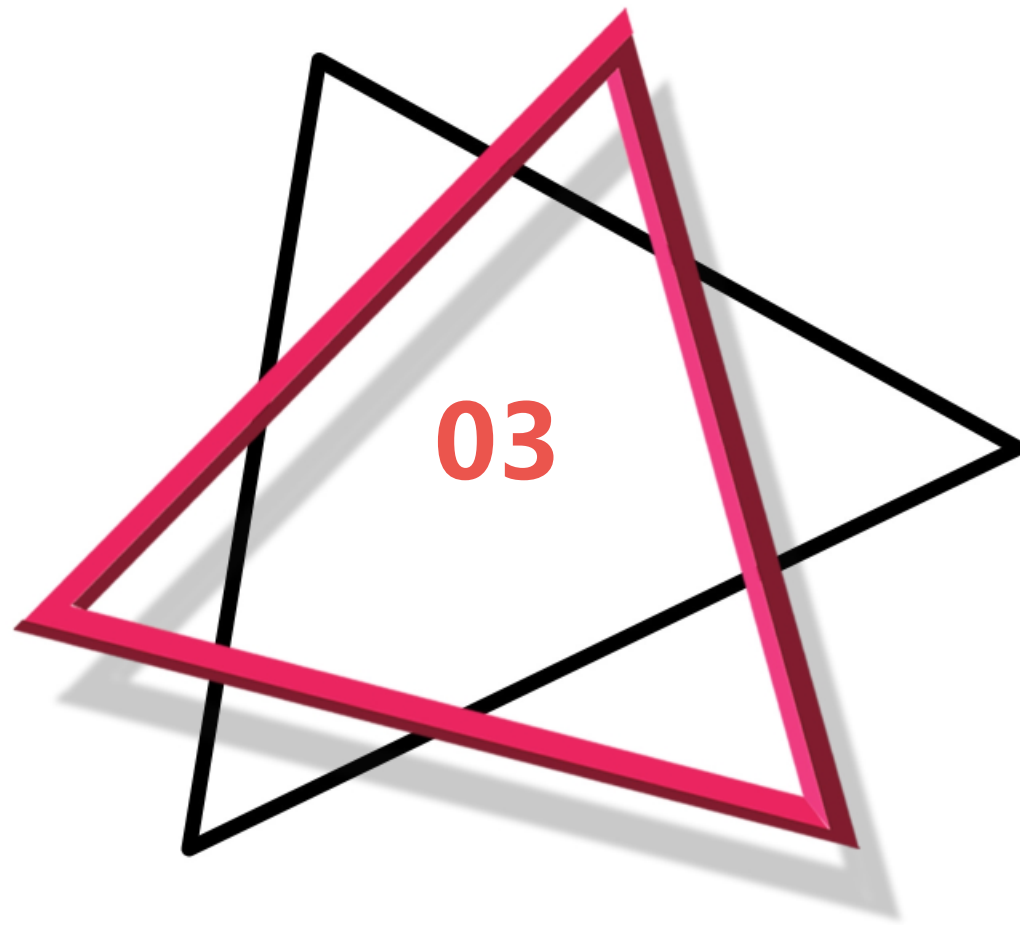
其中，经验风险通常采用平方损失函数或交叉熵损失函数等来衡量，而正则化项则采用非凸正则化项。



03

模型的约束条件通常包括特征权重的非负性约束、稀疏性约束等，以确保模型的解具有实际意义并满足特定需求。





## 并行计算技术在求解中的应用



# 并行计算基本概念和原理

## 并行计算定义

并行计算是指同时使用多种计算资源解决计算问题的过程，其主要目的是快速解决大型且复杂的计算问题。

## 并行计算原理

并行计算通过将问题分解成多个小部分，并同时在多个处理器上进行处理，从而显著提高了计算速度。

## 并行计算模型

常见的并行计算模型包括共享内存模型、消息传递模型和数据并行模型等。



# 并行算法设计思路及实现方法



## 问题分解

将大问题分解为多个小问题，以便并行处理。

## 数据划分

将数据划分为多个部分，每部分数据在单独的处理单元上进行处理。

# 并行算法设计思路及实现方法





# 并行算法设计思路及实现方法



## OpenMP

OpenMP是一种支持多平台共享内存并行编程的API，在C/C和Fortran中广泛使用。

## MPI

MPI ( Message Passing Interface ) 是一种用于分布式内存系统的消息传递标准，支持跨平台并行编程。



## CUDA

CUDA ( Compute Unified Device Architecture ) 是NVIDIA推出的并行计算平台和API，支持GPU加速计算。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：  
<https://d.book118.com/528037101107006101>