

# Similarity Based Constraint Score

## Mid-term Research Project Report

Gadegbeku Fabio

January 16, 2024



# Contents

1	Introduction	2
2	Laplacian Score	3
3	Constraint Score 1	3
4	Constraint Score 4	3
5	Similarity Based Constraint Score	3
6	Next Steps	3

# 1 Introduction

In Machine Learning having too many features is counter productive, this is called the curse of dimensionality. To avoid this phenomenon there exists feature selection methods that evaluate the relevance of the features. More precisely in classification problems we can use *must link*<sup>1</sup> and *cannot link*<sup>2</sup> constraints to define constraint scores to evaluate how well each feature respects the constraints. These constraint scores typically compute distances between the samples in the original feature space to evaluate them, so still suffer from the curse of dimensionality. In this report we will present and implement the Similarity Based Constraint Score (SBCS) described by [1]. That has the unique capabilities of evaluating a whole subset of features at once and calculating distances in a lower dimensional space.

## Goals

The goals of this project are to first of all implement the SBCS and compare it to other constraint scores on different datasets on multiple criteria. Secondly we will try to improve the SBCS by using constraints directly instead of available labels to generate the constraints.

---

<sup>1</sup>When two samples have the same class

<sup>2</sup>When two samples have different classes

## **2 Laplacian Score**

## **3 Constraint Score 1**

## **4 Constraint Score 4**

## **5 Similarity Based Constraint Score**

## **6 Next Steps**

## **References**

- [1] Abderezak Salmi, Kamal Hammouche, and Ludovic Macaire. “Similarity-Based Constraint Score for Feature Selection”. In: *Knowledge-Based Systems* 209 (Dec. 2020), p. 106429. ISSN: 09507051. DOI: 10.1016/j.knosys.2020.106429. (Visited on 09/23/2023).