Similarity Based Constraint Score For Feature Selection

Research Project Presentation

Gadegbeku Fabio

Supervised By : Dr. Ludovic Macaire

April 15, 2024

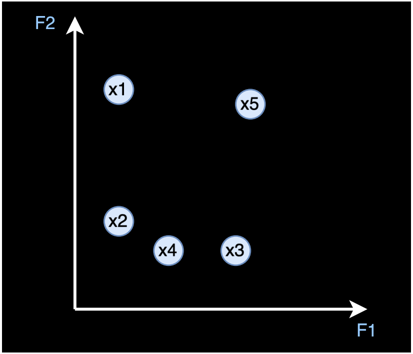
Table of Contents

Introduction

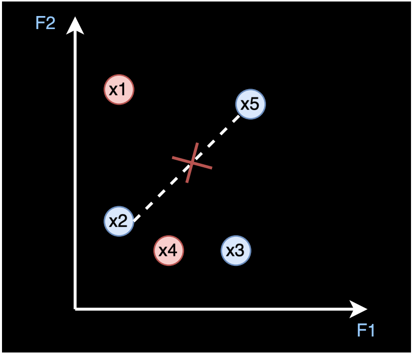
Semi Supervised Learning

- We can define **must link** and **cannot link** constraints
  - must link : When two samples belong to the same class
  - cannot link : When two samples belong to different classes
- **Constraint Scores** to evaluate how well each feature matches the constraints

Introduction



((a)) Unsupervised



((b)) Semi Supervised

Introduction

Schema

- Read and understand the papers on the subject
- Implement 3 supervised scores and 3 semi-supervised scores described in [1]
- Compare results of these scores on multiple datasets

1

[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).

Laplacian Score

Our n samples of d features

$$X = \begin{bmatrix} x_{11} & \dots & x_{1r} & \dots & x_{1d} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & \dots & x_{ir} & \dots & x_{id} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nr} & \dots & x_{nd} \end{bmatrix}$$

A sample of our Data

$$x_i = (x_{i1}, \dots, x_{ir}, \dots, x_{id})^T \in \mathbb{R}^d$$

A feature vector

$$f_r = (x_{1r}, \dots, x_{ir}, \dots, x_{nr})^T \in \mathbb{R}^n$$

Laplacian Score

Similarity Matrix

$$W = \begin{bmatrix} 1 & w_{12} & \dots & w_{1n} \\ w_{21} & 1 & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & 1 \end{bmatrix}$$

Similarity between two samples

$$w_{ij} = S(x_i, x_j)$$

For example :

$$S(x_i, x_j) = \exp(-\frac{||x_i - x_j||^2}{2\sigma^2})$$

Laplacian Score

Degree Matrix

$$D = \begin{bmatrix} d_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & d_{nn} \end{bmatrix}$$

Where :

$$d_{ii} = \sum_{j=1}^{n} w_{ij}$$

Laplacian Score

Laplacian Matrix

$$L = D - W$$

Laplacian Score of feature r

$$L_r = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ir} - x_{jr})^2 s_{ij}}{\sum_{i=1}^{n}(x_{ir} - \bar{f}_r)p_i}$$

Where :

$$pi = \frac{d_i}{\sum_{k=1}^{n} d_k}$$

And we have :

$$L_r = \frac{{f_r}^T L f_r}{{f_r}^T D f_r}$$

Constraint Score 1

Constraints

$$\mathcal{M} = \{(x_i, x_j) \in X \times X \,|\, \text{such that } x_i \text{ and } x_j \text{ belong to the same class}\}$$

$$\mathcal{C} = \{(x_i, x_j) \in X \times X \,|\, \text{such that } x_i \text{ and } x_j \text{ belong to different classes}\}$$

Binary Constraint Matrices

$$w_{ij}^{\mathcal{M}} = \begin{cases} 1 & \text{if } (x_i, x_j) \in \mathcal{M} \text{ or } (x_j, x_i) \in \mathcal{M} \\ 0 & \text{else} \end{cases}$$

$$w_{ij}^{\mathcal{C}} = \begin{cases} 1 & \text{if } (x_i, x_j) \in \mathcal{C} \text{ or } (x_j, x_i) \in \mathcal{C} \\ 0 & \text{else} \end{cases}$$

Constraint Score 1

Minimize

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ir}-x_{jr})^2 w_{ij}^{\mathcal{M}}$$

Maximize

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ir}-x_{jr})^2 w_{ij}^{\mathcal{C}}$$

Constraint Score 1

Laplacian Matrix with constraints

$$L^{\mathcal{M}} = D^{\mathcal{M}} - W^{\mathcal{M}} \quad \text{and} \quad L^{\mathcal{C}} = D^{\mathcal{C}} - C^{\mathcal{C}}$$

Where :

$$D_{ii}^{\mathcal{M}} = \sum_{j=1}^{n} w_{ij}^{\mathcal{M}} \quad D_{ii}^{\mathcal{C}} = \sum_{j=1}^{n} w_{ij}^{\mathcal{C}}$$

SC_r^1 of feature r

$$SC_r^1 = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (x_{ir} - x_{jr})^2 w_{ij}^{\mathcal{M}}}{\sum_{i=1}^{n} \sum_{j=1}^{n} (x_{ir} - x_{jr})^2 w_{ij}^{\mathcal{C}}} = \frac{f_r^T L^{\mathcal{M}} f_r^T}{f_r^T L^{\mathcal{C}} f_r^T}$$

Constraint Score 2

SC_r^2 of feature r

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ir}-x_{jr})^2 w_{ij}^{\mathcal{M}} - \lambda \sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ir}-x_{jr})^2 w_{ij}^{\mathcal{C}} = f_r^T L^{\mathcal{M}} f_r^T - \lambda f_r^T L^{\mathcal{C}} f_r^T$$

λ is a scaling parameter (set to 1 in the paper)

Semi-Supervised Constraint Scores

Supervised VS Semi-Supervised

- Use the same constraints as the supervised scores
- Also use the unlabeled data to gain more information

Constraint Score 3

A redefined similarity Matrix

$$w_{ij}^{\mathrm{kn1}} = \begin{cases} \gamma & \text{if } (x_i, x_j) \in M \\ 1 & \text{if } x_i \in X^U \text{ or } x_j \in X^U \text{ but } x_i \in \mathrm{KNN}(x_j) \text{ or } x_j \in \mathrm{KNN}(x_i) \\ 0 & \text{otherwise} \end{cases}$$

γ set to 100

Redefined Laplacian

$L^{n1} = D^{kn1} - W^{kn1}$ is the Laplacian matrix of W^{kn1}, D^{kn1} being the degree matrix computed from W^{kn1}

Constraint Score 3

SC_r^3 of feature r

$$SC_r^3 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ir} - x_{jr})^2 w_{ij}^{kn1}}{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ir} - x_{jr})^2 w_{ij}^{C}} = \frac{f_r{}^T L^{kn1} f_r}{f_r{}^T L^{C} f_r}$$

Constraint Score 4

SC_r^4 of feature r

$$SC_r^4 = \frac{f_r^T L f_r}{f_r^T D f_r} \cdot \frac{{f_r}^T L^{\mathcal{M}} f_r}{{f_r}^T L^{\mathcal{C}} f_r} = SL_r \cdot SC_r^1$$

Similarity Based Constraint Score

Similarity Based Constraint Score ε^* * = S or SS VS $SC_r^{\#}, \# \in \{1, 2, 3, 4\}$

- ε^* Computes the score of a whole subset of m selected features F_m
- ε^* Only computes distances between samples in F_m

Similarity Based Constraint Score

Target Similarity Matrix of F_m

$$w_{ij}(F_m) = \exp\left(-\frac{\delta^2\left(x_i^{(m)}, x_j^{(m)}\right)}{2\sigma^2}\right)$$

Supervised Similarity matrix of F_m

$$\hat{W}_{ij}^{S} = \begin{cases} 1 & \text{if } (x_i, x_j) \in M \\ 0 & \text{if } (x_i, x_j) \in C \end{cases}$$

Semi-Supervised Similarity matrix of F_m

$$\hat{W}_{ij}^{SS} = \begin{cases} 1 & \text{if } (x_i, x_j) \in M^{SS} \\ 0 & \text{if } (x_i, x_j) \in C \end{cases}$$

Similarity Based Constraint Score

Semi Supervised Must Link

$$M^{SS} = \left\{ (x_i, x_j) \in X^2 | \exists l = 1, \ldots, k \quad NP(x_i) \in X^l \quad \text{and} \quad NP(x_j) \in X^l \right\}$$

Similarity Based Constraint Score of F_m

$$\varepsilon^*(F_m) = ||W(F_m) - \hat{W}^*||_2^2$$

Algorithm 0: Feature Selection Procedure

Input: Set of d features $F_d = \{f_1, \ldots, f_r, \ldots, f_d\}$.
Output: Subset of $\hat{m}$ relevant features $F_{\hat{m}}$.
$F_0 \leftarrow \{\emptyset\}$
for $m = 1$ *to* d **do**

Select the most relevant feature f_r^+:

$$f_r^+ = \arg \min_{f_r \in F_d \setminus F_{m-1}} (\varepsilon^*(F_{m-1} \cup \{f_r\}))$$

Update $F_m \leftarrow F_{m-1} \cup \{f_r^+\}$

Select the number $\hat{m}$ of features such that:

$$\hat{m} = \arg \min_{m=1,2,\ldots,d} (\varepsilon^*(F_m))$$

Output: $F_{\hat{m}}$

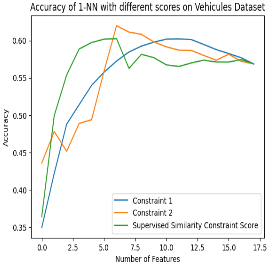
Experimental Results

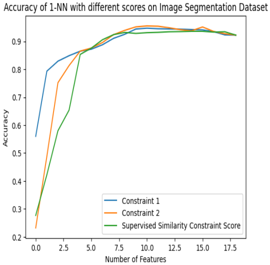
Datasets

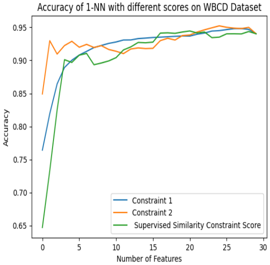
Dataset	Features d	# Training	# Test	Classes k	Prototypes $k * p$
Vehicles	18	473	473	4	12
Image Segmentation	19	578	577	7	21
WBCD	30	285	284	2	6
Ionosphere	34	176	175	2	6
Sonar	60	104	104	2	6
Libras Movement	90	180	180	15	45

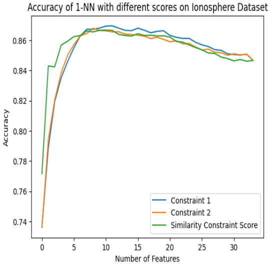
Table: Datasets used in the experiments

Experimental Results [Supervised]

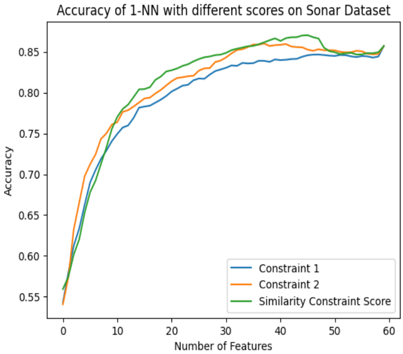


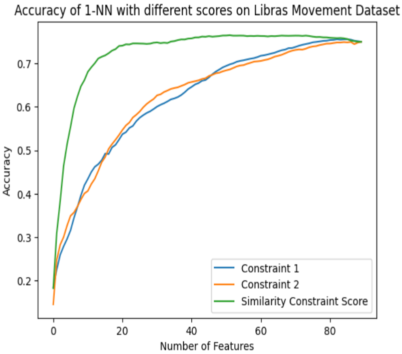




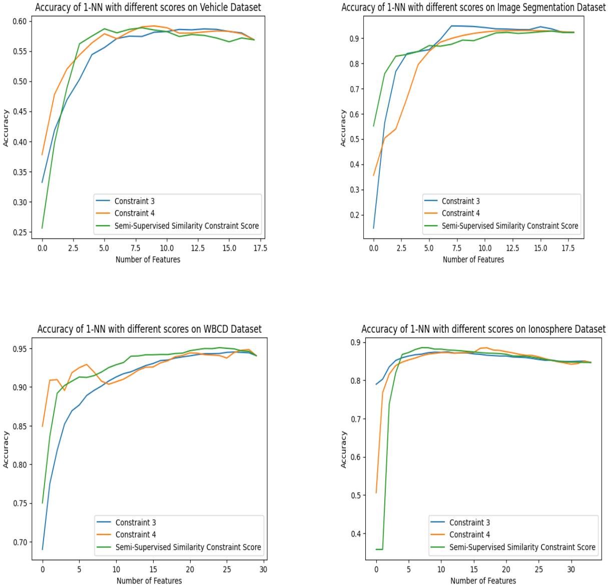


Experimental Results [Supervised]

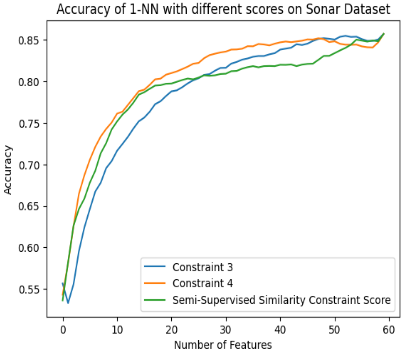


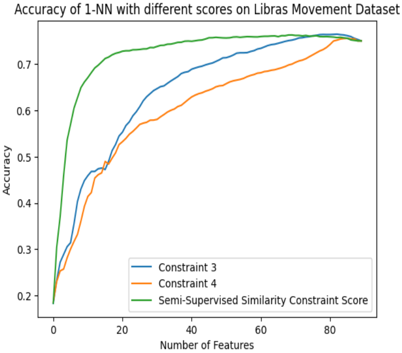


Experimental Results [Semi-Supervised]



Experimental Results [Semi-Supervised]





Experimental Results

To go even further for the ε^{SS} we can see if the set of must link M^{SS} deduced from the prototype set are correct or not.

Database	*CR*	*COV*	*SMC*
Vehicles	34.18%	38.34%	0.08%
Image Segmentation	52.72%	56.03%	0.03%
WBCD	77.66%	80.31%	0.01%
Ionosphere	61.12%	76.17%	0.02%
Sonar	51.81%	54.87%	0.08%
Libras Movement	42.49%	49.33%	1.5%

CR : Correct Ratio, *COV* : Coverage, *SMC* : Standard Must Link Coverage