
Implementation of Multi-Label Classification in Sparse Matrices

Napoleon Maraidonis¹ under the supervision of Michalis Titsias¹

1 Department of Informatics, Athens University of Economics and Business

Abstract

Introduction

Introduction [1] [2] [4]

Notation

- Lowercase bold letter denotes a $n * k$ matrix where n is rows and k is columns :
 $\mathbf{w} \in \mathbb{R}^{n*k}$
- Lowercase bold letter with subscript denotes the n -nth row of the matrix :
 $\mathbf{x}_n \in \mathbb{R}^k$
- Lowercase letter will denote a real number unless stated otherwise : $x \in \mathbb{R}$
- Uppercase Italics letter denotes a set : T
- The letter \hat{y} will denote predicted labels.
- The letter X will denote train data and the letter Y train data labels.
- The letter h will denote a classifier function $h(\mathbf{x}) = \mathbf{y}$

Classifier Chains

Classifier Chains is a transformation of the problem which takes into account label dependence. Classifier Chains works in the following way [4]:

$$\begin{aligned}\hat{y}_1 &= h_1(X), \quad X = [X|\hat{y}_1], \\ \hat{y}_2 &= h_2(X), \quad X = [X|\hat{y}_2], \\ &\dots \\ \hat{y}_n &= h_n(X)\end{aligned}$$

Each classifier is trained using results from previous classifiers in a chain, the order which labels are predicted is arbitrary. There are more complicated schemes that aim to optimise the order that labels are predicted such as Bayes Optimal CC [4].

Prototype Structure and Documentation

General Structure

The prototype consists of three main parts : (a) Binary Classification Algorithm, (b) Multy Label Classifiactation Interface (c) Score function. More specifically :

a Firstly, the implementation of any Binary Classification Algorithm will be required: multiclass logistic regression was chosen for simplicity. The sparsity of the training data will be put into account while implementing the algorithm and stochastic gradient descent will be used for the optimization of the loss function.

b Classifier chains as explained above will be used.

c Scoring will be made according to the formula:

$$accuracy \triangleq \frac{|T \cap P|}{|T \cup P|}$$

Notes

- The implementation of any Multiclass Classification Algorithm will follow the following contract : The methods `train(Xtrain,Ytrain)`, and `predict(Xtest)` will be implemented.
- The implementation of any Multy Label Classifiactation Interface will follow the following contract using the implemented Multiclass Classification Algorithm : The methods `train(Xtrain,Ytrain)`, and `predict(Xtest)` will be implemented.

Results

Discussion

Acknowledgments

References

1. *Multiclass-Multilabel Classification with More Classes than Examples*. AISTATS, 2010.
2. Y. P. M. V. Kush Bhatia, Himanshu Jain. The extreme classification repository: Multi-label datasets and code.
3. J. Read. Multi-label classification. Technical report, Universidad Carlos III de Madrid. Department of Signal Theory and Communications, 2013.
4. J. Read. Multi-label classification. Technical report, Department of Information and Computer Science Helsinki, Finland, 2015.