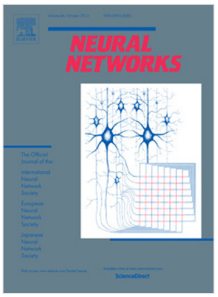
Accepted Manuscript



Incremental sparse Bayesian ordinal regression

Chang Li, Maarten de Rijke

PII: S0893-6080(18)30214-4
DOI: https://doi.org/10.1016/j.neunet.2018.07.015
Reference: NN 3998

To appear in: *Neural Networks*

Received date : 8 February 2018
Revised date : 9 June 2018
Accepted date : 25 July 2018

Please cite this article as: Li, C., de Rijke, M., Incremental sparse Bayesian ordinal regression. *Neural Networks* (2018), https://doi.org/10.1016/j.neunet.2018.07.015

Incremental Sparse Bayesian Ordinal Regression

Chang Li[a,*], Maarten de Rijke[a]

[a]*University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands*

Abstract

Ordinal Regression (OR) aims to model the ordering information between different data categories, which is a crucial topic in multi-label learning. An important class of approaches to OR models the problem as a linear combination of basis functions that map features to a high-dimensional non-linear space. However, most of the basis function-based algorithms are time consuming. We propose an incremental sparse Bayesian approach to OR tasks and introduce an algorithm to sequentially learn the relevant basis functions in the ordinal scenario. Our method, called Incremental Sparse Bayesian Ordinal Regression (ISBOR), automatically optimizes the hyper-parameters via the *type-II maximum likelihood* method. By exploiting fast marginal likelihood optimization, ISBOR can avoid big matrix inverses, which is the main bottleneck in applying basis function-based algorithms to OR tasks on large-scale datasets. We show that ISBOR can make accurate predictions with parsimonious basis functions while offering automatic estimates of the prediction uncertainty. Extensive experiments on synthetic and real word datasets demonstrate the efficiency and effectiveness of ISBOR compared to other basis function-based OR approaches.

Keywords: Ordinal regression, sparse Bayesian learning, basis function-based method

1. Introduction

The task of modeling ordinal data has attracted attention in various areas, including computer vision [1, 2], information retrieval [3], recommender systems [4] and machine learning [5, 6, 7, 8, 9]. Because of the explicit or implicit relationship between labels, simple regression or multi-classification algorithms may fail to find optimal decision boundaries, which motivates the development of dedicated methods.

Generally, OR algorithms can be classified into three categories: naive approaches, ordinal binary decompositions, and threshold models [5]. For naive approaches, OR tasks are simplified into traditional multi-classification or regression tasks, omitting ordering information, and solved by simple machine learning algorithms, e.g., Support Vector Machine (SVM) Regression [10]. For ordinal binary decomposition, the ordinal labels are decomposed into several binary pairs, which are then modeled by a single or multiple classifiers. For the threshold models, the OR problem is addressed by training a threshold model, which models the hidden score function and an implicit set of thresholds that derive the ordinal paradigm. Among these three categories, the

*Corresponding author.

Email addresses: c.li@uva.nl (Chang Li), derijke@uva.nl (Maarten de Rijke)