Retos en clasificación ordinal: redes neuronales artificiales y métodos basados en proyecciones
Challenges in ordinal classification: artificial neural networks and projection-based methods

Tesis Doctoral

Universidad de Granada
Doctorado en Tecnologías de la Información y la Comunicación

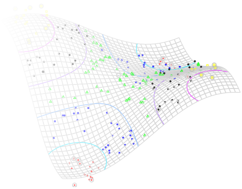
Javier Sánchez Monedero
jsanchezm at uco dot es

Directores:
César Hervás Martínez y Pedro Antonio Gutiérrez Peña

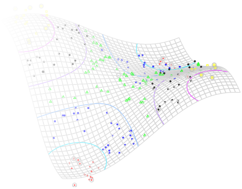
Dept. of Computer Science and Numerical Analysis, University of Córdoba.
Learning and Artificial Neural Networks (AYRNA) research group.
http://www.uco.es/ayrna/

4th September 2013

Outline



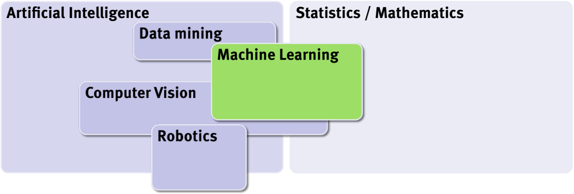
Outline



1. Introduction
2. Objectives
3. Related work
4. Class imbalance
5. Proposals for OR
6. Applications
7. Conclusions and Future Work

Introduction

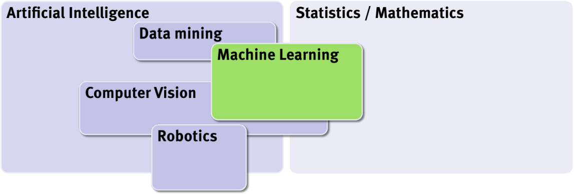
Machine Learning: "Field of study that gives computers the ability to learn without being explicitly programmed"



Machine learning: Where does it fit? What is it not?

Introduction

Machine Learning: "Field of study that gives computers the ability to learn without being explicitly programmed"

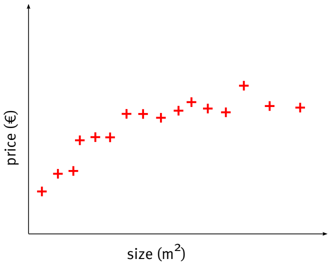


Machine learning: Where does it fit? What is it not?

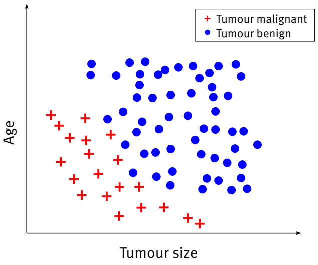
Pattern Recognition/System Modelling:

- Unsupervised learning
- Supervised learning

Supervised learning

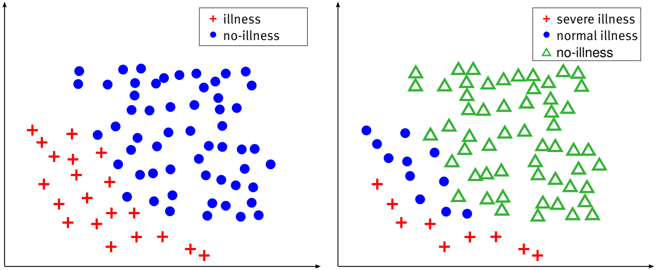


Regression problem "Given these data, a friend has a house of 75 square meters, how much can he expect to get?".



Classification problem "Can you estimate prognosis based on tumour size and known age?"

Binary vs Ordinal Classification



Comparison of binary and ordinal classification

Ordinal regression

Ordinal classification/regression (OR)

Definition: Ordinal classification (so called ranking, sorting or ordinal regression) is a supervised learning problem of predicting categories that have an ordered arrangement.

Ordinal regression

Ordinal classification/regression (OR)

Definition: Ordinal classification (so called ranking, sorting or ordinal regression) is a supervised learning problem of predicting categories that have an ordered arrangement.

Goals/Challenges:

- To exploit the ordinal relationship of the data.
- To minimize errors that consider the order between classes.

Ordinal regression

Ordinal classification/regression (OR)

Definition: Ordinal classification (so called ranking, sorting or ordinal regression) is a supervised learning problem of predicting categories that have an ordered arrangement.

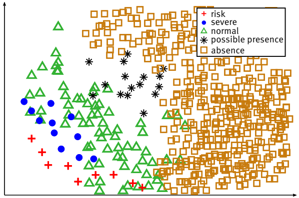
Goals/Challenges:

- To exploit the ordinal relationship of the data.
- To minimize errors that consider the order between classes.

Applications

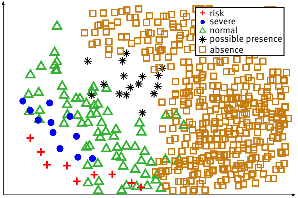
Teaching assistant evaluation, car insurance risk rating, pasture production prediction, breast cancer conservative treatment, credit rating. . .

Ordinal regression example: illness classification



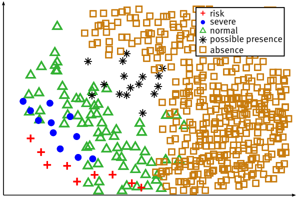
- Illness degrees based on an ordinal scale $\{C_1 = \textit{risk},\ C_2 = \textit{severe},\ C_3 = \textit{normal},\ C_4 = \textit{possible presence},\ C_5 = \textit{absence}\}$

Ordinal regression example: illness classification



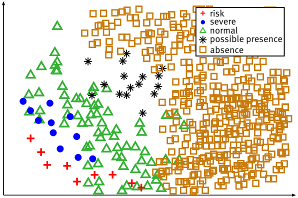
- Illness degrees based on an ordinal scale $\{C_1 = risk,\ C_2 = severe,\ C_3 = normal,\ C_4 = possible\ presence,\ C_5 = absence\}$
- Class labels are imbued with order information

Ordinal regression example: illness classification



- Illness degrees based on an ordinal scale $\{C_1 = risk,\ C_2 = severe,\ C_3 = normal,\ C_4 = possible\ presence,\ C_5 = absence\}$
- Class labels are imbued with order information
- Misclassification costs are not the same for different errors

Ordinal regression example: illness classification



- Illness degrees based on an ordinal scale $\{C_1 = risk,\ C_2 = severe,\ C_3 = normal,\ C_4 = possible\ presence,\ C_5 = absence\}$

- Class labels are imbued with order information
- Misclassification costs are not the same for different errors
- Class imbalance can be very common (e.g. medicine, teaching evaluation. . .)

Problem formulation

- The purpose is to learn a mapping ϕ from the input space $\mathcal{X}$ to a finite set $\mathcal{C} = \{C_1, C_2, \ldots, C_Q\}$ containing Q labels, where the label set has a linear order relation $C_1 \prec C_2 \prec \ldots \prec C_Q$.
- Each pattern is represented by a K-dimensional feature vector $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^K$ and a class label $y \in \mathcal{C}$.
- The training dataset $\mathbf{T}$ is composed of N patterns $\mathbf{T} = \{(\mathbf{X}, \mathrm{Y}) = (\mathbf{x}_i, y_i) : \mathbf{x}_i \in \mathcal{X},\ y_i \in \mathcal{C}\,(i = 1, \ldots, N)\}$, with $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,K})$.

Motivation and challenges

- State-of-the-art in ordinal regression

Motivation and challenges

- State-of-the-art in ordinal regression
- Class imbalance

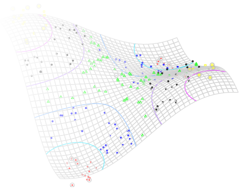
Motivation and challenges

- State-of-the-art in ordinal regression
- Class imbalance
- Data ordering exploitation

Motivation and challenges

- State-of-the-art in ordinal regression
- Class imbalance
- Data ordering exploitation
- Real world problems

Outline



Objectives I

All these challenges result in the following formal objectives considered for the thesis:

1. State of the art in ordinal regression objectives:
   1. *To propose an OR method taxonomy*.
   2. *To review OR evaluation metrics*.
   3. *To select benchmark datasets*.
2. Class imbalance can be divided into the following objectives:
   1. *To perform an analysis of the state of the art for nominal class imbalance*.
   2. *To optimize algorithms that tackle the nominal class imbalance as a multi-objective optimization problem*.
   3. *To explore new solutions considering ordinal class imbalance*.

Objectives II

3. Data ordering exploitation:
    1. *To check if data ordering exploitation improves classification performance in OR problems.*
    2. *To design OR algorithms based on standard regression but avoiding any trivial assumption about the latent variable.*
    3. *To develop latent variable modelling approaches only with restrictions in the labels set.*
    4. *To develop classifiers that exploit the input data ordering.*
    5. *To develop methods that relax the data projection of threshold methods.*

Objectives III

4. Application of OR methods to real problems:
   1. *To develop sovereign credit rating classification methods using ordinal regression.*
   2. *To develop wind speed forecasting systems using ordinal regression.*

Proposals overview

OR State of the Art

- Taxonomy and Review [1] [2]
- Datasets
- Performance Metrics [3] [4]

New Proposals for **Class Imbalance** [5]

- EELM Extensions
  - CCR vs MS [6]
  - RMSE vs MRMSE [7]
  - CEntropy vs MCEntropy

New Proposals for **Ordinal Regression**

- EELM Extensions
  - EELM for Ordinal Regression [8]
- Latent Variable Modelling
  - Indirect Modelling: NVR [9]
  - Direct Modelling: PCDOC [10]

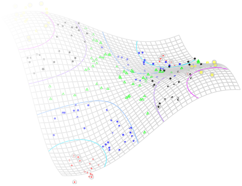
Aplications of Ordinal Regression

- Soverein Debt Rating [12]
- Wind Speed Forecasting [11]

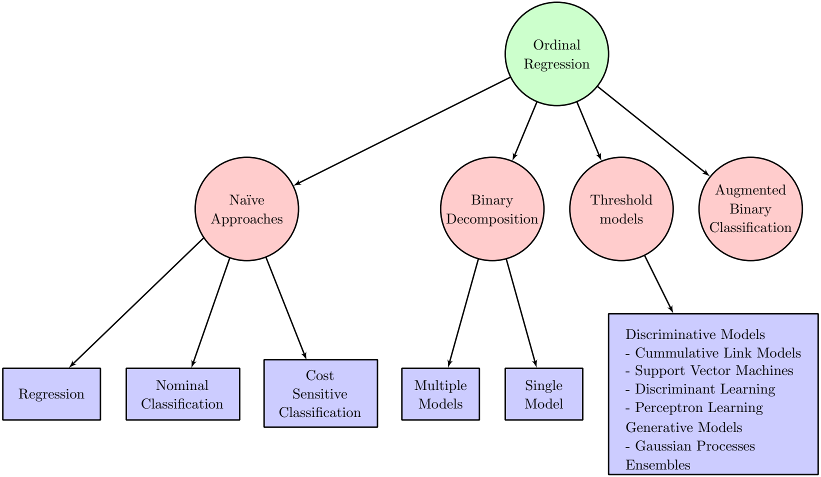
Legend:

- Review
- Evolutionary ANNs
- Latent Variable Modelling
- International Conference
- International Journal

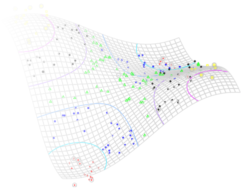
Outline



Proposed Taxonomy



Outline

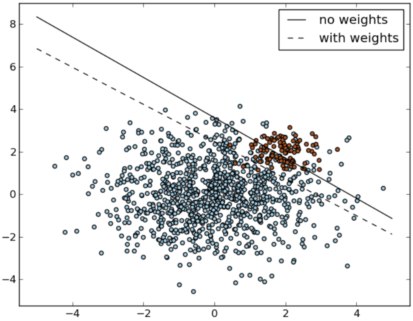


The Class Imbalance Problem

- Data imbalance refers to datasets where the number of patterns belonging to each class varies noticeably
- Classifiers tend to ignore minority classes
  - Typically, those are the most interesting ones (e.g. illness detection)
- Very active research in nominal binary and multi-class fields

The Class Imbalance Problem

- Data imbalance refers to datasets where the number of patterns belonging to each class varies noticeably
- Classifiers tend to ignore minority classes
    - Typically, those are the most interesting ones (e.g. illness detection)
- Very active research in nominal binary and multi-class fields



Imbalance problem depends on the noise and overlap degree

Evaluation of imbalanced problems

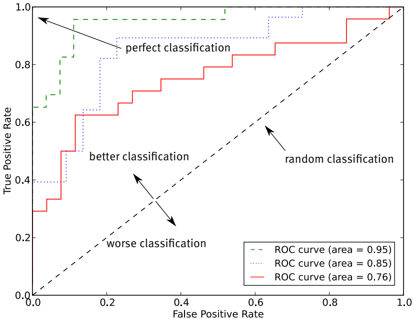
Evaluation metrics:

- Essential to evaluate and guide the learning algorithms

Evaluation of imbalanced problems

Evaluation metrics:

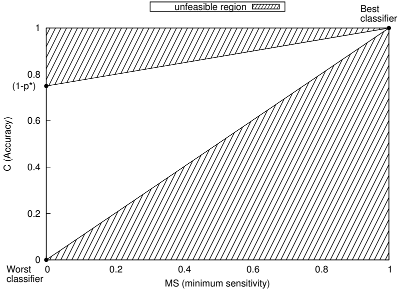
- Essential to evaluate and guide the learning algorithms
- Binary problems: precision/recall, F-measure, ROC and AUC



Evaluation of imbalanced problems

Evaluation metrics:

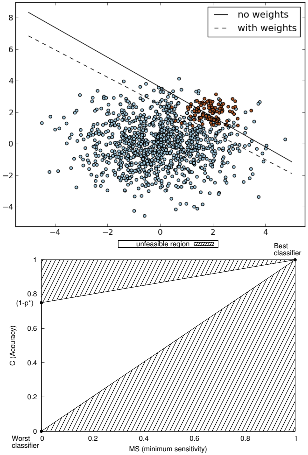
- Essential to evaluate and guide the learning algorithms
- Binary problems: precision/recall, F-measure, ROC and AUC
- Multi-class problems: Geometric Mean of accuracy for each class and the **Accuracy-Minimum Sensitivity**



Solutions to the imbalance problem

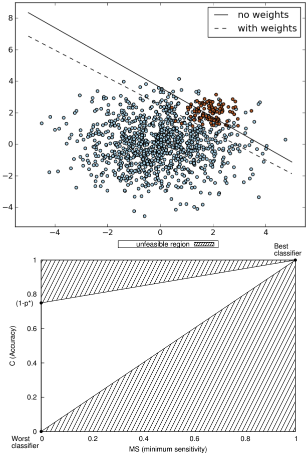
1. Data preprocessing level: the data is preprocessed suppressing or adding patterns $\rightarrow$ *resampling techniques*
2. Model and algorithm level: the models and/or training algorithms consider performance of can be modified for dealing with data imbalance.
3. Hybrid approaches

Accuracy vs. Minimum Sensitivity



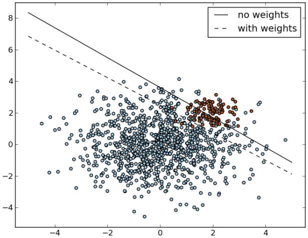
- Accuracy vs. Minimum Sensitivity for ANNs training formulated as a multi-objective optimization problem

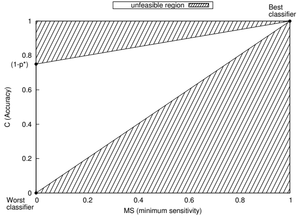
Accuracy vs. Minimum Sensitivity



- Accuracy vs. Minimum Sensitivity for ANNs training formulated as a multi-objective optimization problem
- Multi Objective Evolutionary Algorithm (MOEA)

Accuracy vs. Minimum Sensitivity





- Accuracy vs. Minimum Sensitivity for ANNs training formulated as a multi-objective optimization problem
- Multi Objective Evolutionary Algorithm (MOEA)
- Pareto based algorithms: good classification performance at the cost of high computational cost

Multi-objective reformulation and Training algorithm

The previous Pareto based approach is reformulated as a weighed convex linear optimization problem

Multi-objective reformulation and Training algorithm

The previous Pareto based approach is reformulated as a weighed convex linear optimization problem

. . . the error functions are **not differentiable**. . .

Multi-objective reformulation and Training algorithm

The previous Pareto based approach is reformulated as a weighed convex linear optimization problem

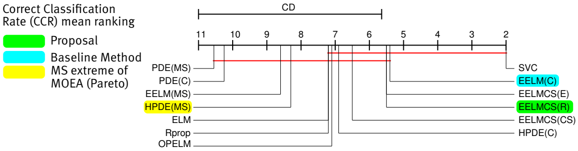
. . . the error functions are **not differentiable**. . .

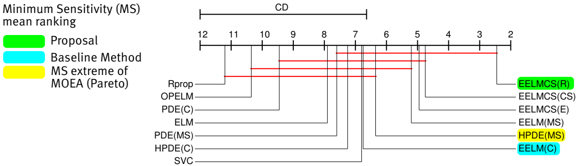
$\Downarrow$

- Meta-heuristic method: Evolutionary Algorithm
- Evolutionary Extreme Learning Machine
  - Based on Differential Evolution
  - Avoids costly gradient descent optimization

Results I

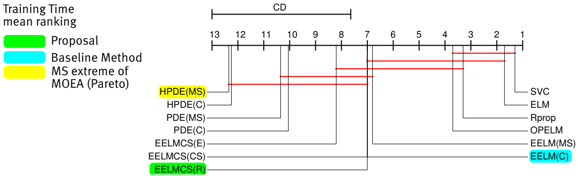
Nemenyi Critical Distance (CD) diagrams comparing CCR and MS mean results:



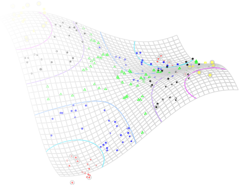


Results II

Nemenyi Critical Distance (CD) diagrams comparing training time mean results:



Outline



Proposals for OR

1. Latent variable modelling with probability distributions
2. Pairwise Class Distances Projection for Ordinal Classification
3. Evolutionary Extreme Learning Machine for Ordinal Regression

Threshold Models I

- Most common models in OR
- Assumption: there exist a latent continuous variable that captures the underlying order of the patterns
- This variable is difficult to measure or cannot be observed

Spaces

- Input space $\mathcal{X}$: observable
- Label space $\mathcal{C}$: observable
- Latent space $\mathcal{Z}$: unobservable or non-directly observable

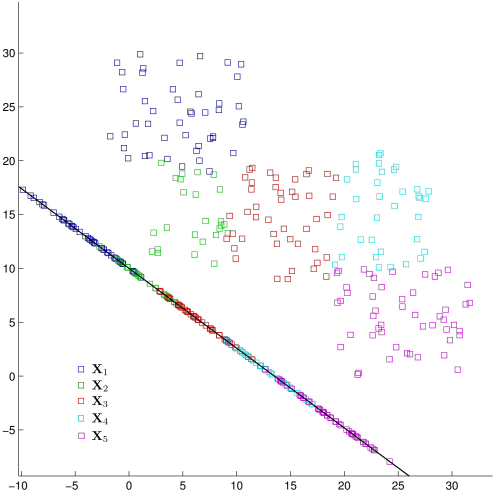
Threshold Models II

The threshold model can be represented with the following general expression:

$$f(\mathbf{x}, \theta) = \begin{cases} C_1, & \text{if } g(\mathbf{x}) \leq \theta_1, \\ C_2, & \text{if } \theta_1 < g(\mathbf{x}) \leq \theta_2, \\ \vdots & \\ C_Q, & \text{if } g(\mathbf{x}) > \theta_{Q-1}, \end{cases} \tag{1}$$

where $g : \mathcal{X} \to \mathbb{R}$ is the function that **projects data space onto the 1-dimensional latent space** $\mathcal{Z} \subseteq \mathbb{R}$ and $\theta_1 \leq \theta_2 \ldots \leq \theta_{Q-1}$ are the thresholds that divide the space into ordered intervals corresponding to the classes.

Threshold Models III



Example projection of Linear Discriminant Analysis for OR

Switching from classification to regression

1. We move from the following **classification problem**...

$$\mathbf{T} = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathcal{X},\ y_i \in \mathcal{C}, i = 1, \ldots, N\}, \mathbf{x}_i = (x_{i1}, \ldots, x_{iK}).$$

2. ...to a **regression problem** where the response variable is generated by the algorithm:

$$\mathbf{T}' = \left\{(\mathbf{x}_i, \phi(\mathbf{x}_i^{(y_i)}) \mid (\mathbf{x}_i, \mathrm{y}_i) \in \mathbf{T}\right\},$$

where ϕ assigns a value in the latent space $\mathcal{Z}$ to each pattern during the training phase

3. At this point we can apply any generic regressor $\hat{g}$ to predict the latent variable
4. Prediction consist on estimating z values for each pattern and then assign it to each class according to a threshold set

NVR I

- **Indirect modelling** of the latent space: NVR algorithm
- The latent variable $\mathcal{Z}$ is considered as a random variable that is sampled, depending on the pattern class, from a set of different probability distributions
- The NVR approach does not assume any ordering in the input space, but only on the labels space, which is the strict definition of Ordinal Regression

NVR II

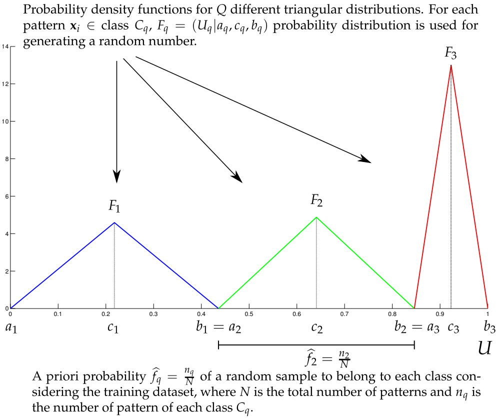


Figure: NVR with triangular probability distributions example

Exploitation of data ordering

- The strict definition of OR limits the order restriction to the labels space $\mathcal{C}$

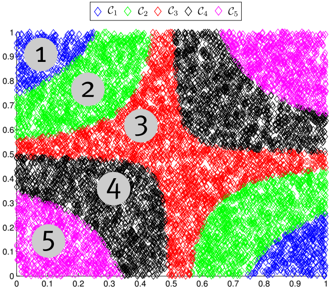
Exploitation of data ordering

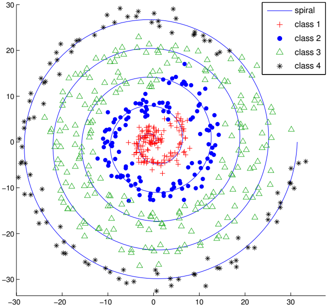
- The strict definition of OR limits the order restriction to the labels space $\mathcal{C}$
- Nevertheless, some authors suggest that the label ordering should be somehow present in the input space $\mathcal{X}$

Exploitation of data ordering

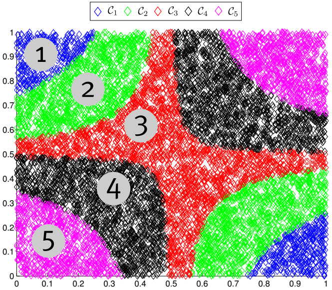
- The strict definition of OR limits the order restriction to the labels space $\mathcal{C}$
- Nevertheless, some authors suggest that the label ordering should be somehow present in the input space $\mathcal{X}$
- ¿Can this order be **exploited** to improve the latent space ($\mathcal{Z}$) modelling?

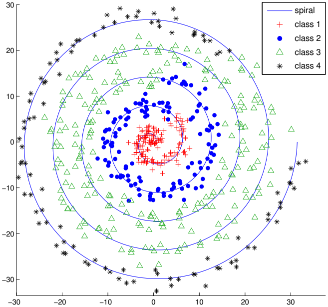
Data ordering restriction





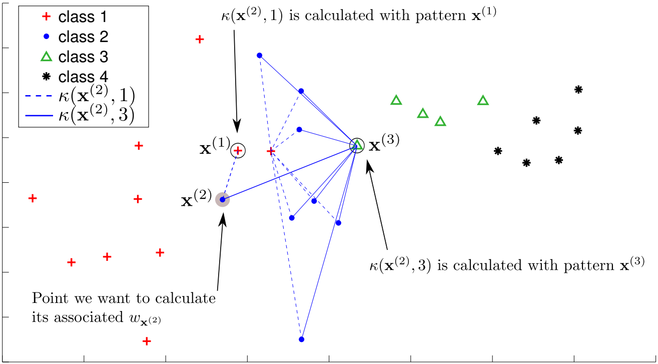
Data ordering restriction





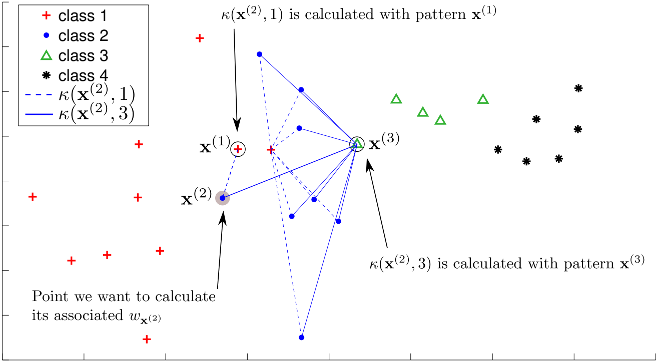
If there is an order in the input space, **this order should be always present between adjacent classes**

How '*well*' is a pattern placed?



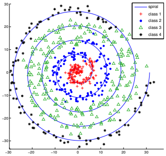
How '*well*' is a pattern placed in the latent space interval corresponding to its class?

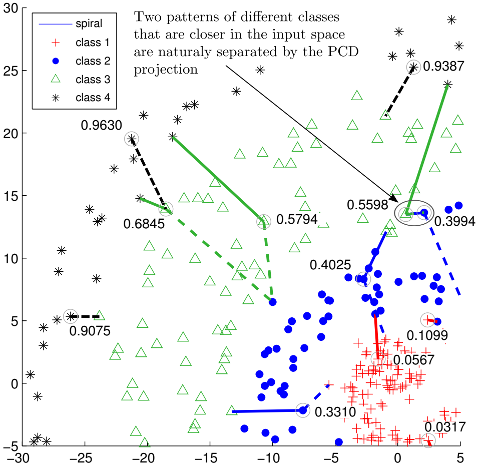
How '*well*' is a pattern placed?

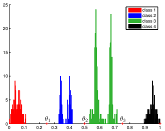


How '*well*' is a pattern placed in the latent space interval corresponding to its class? $\rightarrow$ We estimate this value with the minimum distances to the patterns of neighbour classes.

Pairwise Class Distances projection







Experimental results of SVR-PCDOC

Accuracy Mean$_{SD}$

Method/DataSet	automobile	bondrate	contact-lenses	eucalyptus	newthyroid	pasture	squash-stored	squash-unstored	tae	winequality-red
ASAOR(C4.5)	$0.6962_{0.0585}$	$0.5333_{0.0743}$	$\mathit{0.7500_{0.0848}}$	$0.6391_{0.0360}$	$0.9167_{0.0388}$	$\mathbf{0.7519_{0.1450}}$	$0.6026_{0.1179}$	$\mathit{0.7744_{0.1009}}$	$0.3947_{0.0578}$	$0.6027_{0.0211}$
GPOR	$0.6109_{0.0726}$	$\mathbf{0.5778_{0.0320}}$	$0.6056_{0.0927}$	$\mathbf{0.6855_{0.0341}}$	$0.9660_{0.0244}$	$0.5222_{0.1779}$	$0.4513_{0.1005}$	$0.6436_{0.1622}$	$0.3281_{0.0407}$	$0.6058_{0.0148}$
KLDOR	$\mathbf{0.7218_{0.0575}}$	$0.5422_{0.0871}$	$0.5889_{0.1736}$	$0.6107_{0.0282}$	$0.9716_{0.0187}$	$\mathit{0.6778_{0.1250}}$	$\mathbf{0.7026_{0.1120}}$	$\mathbf{0.8282_{0.1043}}$	$0.5553_{0.0519}$	$0.6029_{0.0165}$
POM	$0.4673_{0.1940}$	$0.3444_{0.1605}$	$0.6222_{0.1379}$	$0.1594_{0.0364}$	$\mathit{0.9722_{0.0222}}$	$0.4963_{0.1537}$	$0.3821_{0.1518}$	$0.3487_{0.1425}$	$0.5123_{0.0890}$	$0.5940_{0.0174}$
SVC	$\mathit{0.6974_{0.0623}}$	$\mathit{0.5556_{0.0686}}$	$\mathbf{0.7944_{0.1290}}$	$\mathit{0.6534_{0.0368}}$	$0.9667_{0.0250}$	$0.6333_{0.1342}$	$0.6564_{0.1273}$	$0.7000_{0.0817}$	$0.5386_{0.0617}$	$\mathbf{0.6358_{0.0210}}$
SVMRank	$0.6840_{0.0548}$	$0.5533_{0.0725}$	$0.7000_{0.1107}$	$0.6511_{0.0244}$	$0.9685_{0.0224}$	$0.6481_{0.1340}$	$0.6641_{0.1040}$	$0.7487_{0.0855}$	$0.5219_{0.0735}$	$0.6178_{0.0215}$
SVOREX	$0.6654_{0.0679}$	$0.5533_{0.0961}$	$0.6500_{0.1265}$	$0.6467_{0.0288}$	$0.9673_{0.0221}$	$0.6296_{0.1249}$	$0.6282_{0.1326}$	$0.7179_{0.1283}$	$0.5807_{0.0602}$	$0.6293_{0.0217}$
SVORIM	$0.6385_{0.0757}$	$0.5467_{0.0916}$	$0.6333_{0.1269}$	$0.6386_{0.0283}$	$0.9691_{0.0214}$	$0.6667_{0.1203}$	$0.6385_{0.1181}$	$0.7641_{0.1029}$	$\mathbf{0.5895_{0.0661}}$	$0.6303_{0.0219}$
SVR-PCDOC	$0.6782_{0.0595}$	$0.5397_{0.1009}$	$0.6889_{0.0952}$	$0.6482_{0.0289}$	$\mathbf{0.9735_{0.0205}}$	$0.6556_{0.1025}$	$\mathit{0.6846_{0.1235}}$	$0.6949_{0.0845}$	$\mathit{0.5816_{0.0642}}$	$\mathit{0.6306_{0.0225}}$

MAE Mean$_{SD}$

Method/DataSet	automobile	bondrate	contact-lenses	eucalyptus	newthyroid	pasture	squash-stored	squash-unstored	tae	winequality-red
ASAOR(C4.5)	$0.4006_{0.0945}$	$0.6244_{0.0792}$	$0.3667_{0.1541}$	$0.3835_{0.0417}$	$0.0833_{0.0388}$	$\mathbf{0.2481_{0.1450}}$	$0.4436_{0.1395}$	$\mathit{0.2385_{0.1094}}$	$0.6860_{0.1464}$	$0.4406_{0.0234}$
GPOR	$0.5942_{0.1307}$	$0.6244_{0.0619}$	$0.5111_{0.1747}$	$\mathbf{0.3310_{0.0378}}$	$0.0340_{0.0244}$	$0.4889_{0.1904}$	$0.6256_{0.1481}$	$0.3564_{0.1622}$	$0.8614_{0.1551}$	$0.4248_{0.0172}$
KLDOR	$\mathbf{0.3340_{0.0760}}$	$0.5867_{0.1071}$	$0.5389_{0.2085}$	$0.4236_{0.0321}$	$0.0284_{0.0187}$	$\mathit{0.3222_{0.1250}}$	$\mathbf{0.3077_{0.1278}}$	$\mathbf{0.1718_{0.1043}}$	$0.4728_{0.0686}$	$0.4434_{0.0188}$
POM	$0.9532_{0.6866}$	$0.9467_{0.3206}$	$0.5333_{0.2413}$	$2.0286_{0.0698}$	$\mathit{0.0278_{0.0222}}$	$0.5852_{0.2041}$	$0.8128_{0.2480}$	$0.8256_{0.2303}$	$0.6263_{0.1262}$	$0.4393_{0.0190}$
SVC	$0.4455_{0.0945}$	$0.6244_{0.0901}$	$\mathbf{0.3111_{0.2220}}$	$0.3944_{0.0424}$	$0.0333_{0.0250}$	$0.3667_{0.1342}$	$0.3769_{0.1595}$	$0.3077_{0.0903}$	$0.5781_{0.0825}$	$0.4076_{0.0202}$
SVMRank	$\mathit{0.3929_{0.0730}}$	$0.5978_{0.0884}$	$0.3778_{0.1691}$	$\mathit{0.3797_{0.0272}}$	$0.0315_{0.0224}$	$0.3593_{0.1420}$	$0.3462_{0.1102}$	$0.2513_{0.0855}$	$0.5149_{0.0865}$	$0.4193_{0.0212}$
SVOREX	$0.4083_{0.0887}$	$\mathit{0.5733_{0.1208}}$	$0.4889_{0.1854}$	$0.3920_{0.0305}$	$0.0327_{0.0221}$	$0.3704_{0.1249}$	$0.3821_{0.1392}$	$0.2821_{0.1283}$	$0.4851_{0.0781}$	$0.4076_{0.0234}$
SVORIM	$0.4244_{0.0895}$	$0.5911_{0.1017}$	$0.5056_{0.1666}$	$0.3953_{0.0348}$	$0.0309_{0.0214}$	$0.3333_{0.1203}$	$0.3718_{0.1263}$	$\mathit{0.2385_{0.1094}}$	$\mathit{0.4605_{0.0805}}$	$\mathit{0.4057_{0.0222}}$
SVR-PCDOC	$0.3974_{0.0932}$	$\mathbf{0.5683_{0.1258}}$	$\mathit{0.3667_{0.1541}}$	$0.3924_{0.0382}$	$\mathbf{0.0265_{0.0205}}$	$0.3481_{0.1041}$	$\mathit{0.3256_{0.1409}}$	$0.3051_{0.0845}$	$\mathbf{0.4570_{0.0713}}$	$\mathbf{0.4001_{0.0233}}$

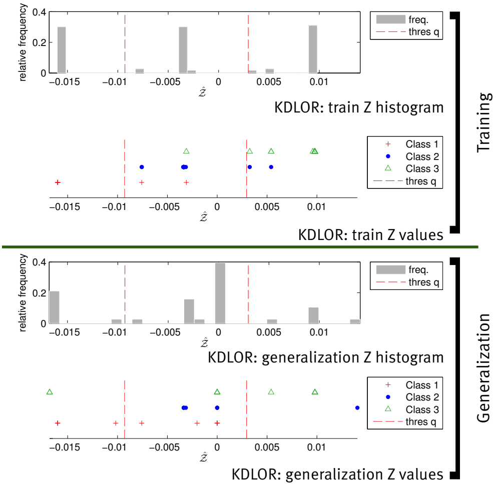
AMAE Mean$_{SD}$

Method/DataSet	automobile	bondrate	contact-lenses	eucalyptus	newthyroid	pasture	squash-stored	squash-unstored	tae	winequality-red
ASAOR(C4.5)	$0.5110_{0.1044}$	$1.2262_{0.1745}$	$\mathit{0.3148_{0.1239}}$	$0.4281_{0.0449}$	$0.1152_{0.0563}$	$\mathbf{0.2481_{0.1450}}$	$0.5019_{0.1924}$	$\mathbf{0.2556_{0.1485}}$	$0.6890_{0.1508}$	$\mathit{1.0453_{0.0801}}$
GPOR	$0.7924_{0.1998}$	$1.3600_{0.1221}$	$0.6509_{0.2861}$	$\mathbf{0.3624_{0.0402}}$	$0.0623_{0.0491}$	$0.4889_{0.1904}$	$0.7974_{0.2342}$	$0.4426_{0.2258}$	$0.8627_{0.1635}$	$1.0647_{0.0650}$
KLDOR	$\mathbf{0.3453_{0.1040}}$	$\mathit{1.0371_{0.2698}}$	$0.5185_{0.2802}$	$0.4259_{0.0379}$	$0.0587_{0.0402}$	$\mathit{0.3222_{0.1250}}$	$\mathbf{0.3493_{0.1556}}$	$\mathit{0.3093_{0.1797}}$	$0.4706_{0.0697}$	$1.2575_{0.0687}$
POM	$1.0263_{0.7996}$	$1.1031_{0.4034}$	$0.5352_{0.2754}$	$1.9898_{0.0483}$	$\mathit{0.0496_{0.0402}}$	$0.5852_{0.2041}$	$0.8152_{0.2510}$	$0.7907_{0.3316}$	$0.6266_{0.1279}$	$1.0852_{0.0374}$
SVC	$0.4862_{0.1247}$	$1.2654_{0.1832}$	$\mathbf{0.3065_{0.2772}}$	$0.4329_{0.0475}$	$0.0599_{0.0512}$	$0.3667_{0.1342}$	$0.4463_{0.1888}$	$0.4444_{0.1631}$	$0.5762_{0.0833}$	$1.1189_{0.0693}$
SVMRank	$0.4677_{0.0963}$	$1.1840_{0.2246}$	$0.3852_{0.1976}$	$0.4143_{0.0301}$	$0.0570_{0.0485}$	$0.3593_{0.1420}$	$0.3907_{0.1485}$	$0.3481_{0.1591}$	$0.5125_{0.0862}$	$1.0679_{0.0689}$
SVOREX	$0.5184_{0.0955}$	$1.0717_{0.2166}$	$0.5167_{0.3029}$	$0.4110_{0.0335}$	$0.0542_{0.0415}$	$0.3704_{0.1249}$	$0.4326_{0.1720}$	$0.4259_{0.1567}$	$0.4839_{0.0787}$	$1.0954_{0.0670}$
SVORIM	$0.5230_{0.1050}$	$1.1139_{0.2330}$	$0.5889_{0.2590}$	$0.4198_{0.0429}$	$0.0550_{0.0418}$	$0.3333_{0.1203}$	$0.4270_{0.1481}$	$0.3667_{0.1404}$	$\mathit{0.4588_{0.0810}}$	$1.0931_{0.0723}$
SVR-PCDOC	$\mathit{0.4404_{0.1277}}$	$\mathbf{0.9692_{0.2244}}$	$0.4204_{0.0978}$	$\mathit{0.4001_{0.0429}}$	$\mathbf{0.0451_{0.0401}}$	$0.3481_{0.1041}$	$\mathit{0.3596_{0.1838}}$	$0.3963_{0.1583}$	$\mathbf{0.4548_{0.0706}}$	$\mathbf{1.0400_{0.0963}}$

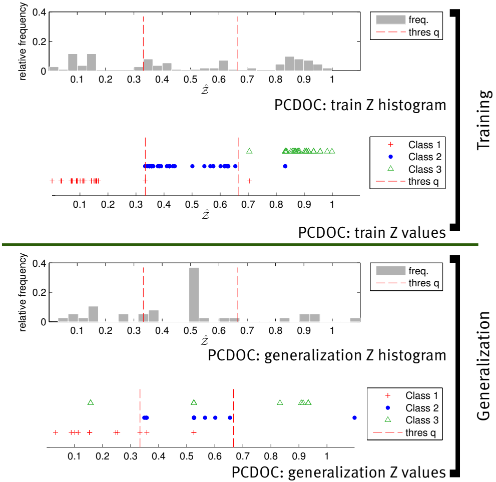
τ_b Mean$_{SD}$

Method/DataSet	automobile	bondrate	contact-lenses	eucalyptus	newthyroid	pasture	squash-stored	squash-unstored	tae	winequality-red
ASAOR(C4.5)	$0.7413_{0.0690}$	$0.1432_{0.1594}$	$\mathit{0.6041_{0.2155}}$	$\mathit{0.8024_{0.0254}}$	$0.8528_{0.0668}$	$\mathbf{0.7781_{0.1319}}$	$0.4153_{0.2447}$	$\mathit{0.6922_{0.1453}}$	$0.2432_{0.1766}$	$0.4962_{0.0358}$
GPOR	$0.5573_{0.1176}$	$0.0000_{0.0000}$	$0.3480_{0.3037}$	$\mathbf{0.8298_{0.0201}}$	$0.9375_{0.0449}$	$0.4609_{0.3143}$	$0.0748_{0.2110}$	$0.4201_{0.3313}$	$-0.0180_{0.1078}$	$0.5227_{0.0256}$
KLDOR	$\mathbf{0.7933_{0.0565}}$	$0.3564_{0.2571}$	$0.4484_{0.2730}$	$0.7858_{0.0171}$	$0.9484_{0.0338}$	$\mathit{0.7177_{0.1326}}$	$\mathbf{0.6461_{0.1603}}$	$\mathbf{0.7642_{0.1613}}$	$0.4770_{0.1137}$	$0.4601_{0.0277}$
POM	$0.4954_{0.2833}$	$0.2897_{0.3017}$	$0.4575_{0.3092}$	$0.0080_{0.0375}$	$\mathit{0.9494_{0.0402}}$	$0.4631_{0.2368}$	$0.1689_{0.3040}$	$0.1087_{0.3054}$	$0.3167_{0.1290}$	$0.4969_{0.0253}$
SVC	$0.6948_{0.0768}$	$0.1209_{0.1767}$	$0.6009_{0.3003}$	$0.7827_{0.0250}$	$0.9394_{0.0453}$	$0.6979_{0.1331}$	$0.5412_{0.2396}$	$0.5990_{0.1395}$	$0.3752_{0.1100}$	$0.5160_{0.0272}$
SVMRank	$\mathit{0.7512_{0.0539}}$	$0.2540_{0.2465}$	$0.5773_{0.2417}$	$0.7997_{0.0170}$	$0.9426_{0.0406}$	$0.7072_{0.1290}$	$0.6013_{0.1479}$	$0.6615_{0.1081}$	$0.4171_{0.1195}$	$0.5247_{0.0298}$
SVOREX	$0.7486_{0.0618}$	$\mathit{0.3685_{0.2160}}$	$0.4252_{0.3041}$	$0.7938_{0.0186}$	$0.9408_{0.0396}$	$0.6907_{0.1150}$	$0.5335_{0.2072}$	$0.5923_{0.2119}$	$0.4453_{0.1104}$	$0.5313_{0.0278}$
SVORIM	$0.7483_{0.0648}$	$0.2987_{0.2302}$	$0.3824_{0.2689}$	$0.7921_{0.0197}$	$0.9442_{0.0382}$	$0.7101_{0.1140}$	$0.5417_{0.1667}$	$0.6561_{0.1868}$	$\mathit{0.4819_{0.1182}}$	$\mathit{0.5328_{0.0297}}$
SVR-PCDOC	$0.7445_{0.0756}$	$\mathbf{0.4546_{0.2176}}$	$\mathbf{0.6202_{0.2168}}$	$0.7947_{0.0235}$	$\mathbf{0.9518_{0.0367}}$	$0.7124_{0.1015}$	$\mathit{0.6102_{0.2014}}$	$0.6022_{0.1327}$	$\mathbf{0.4934_{0.1007}}$	$\mathbf{0.5415_{0.0326}}$

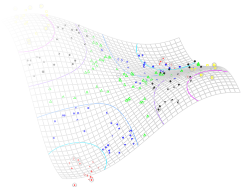
Problem of the highly non-linear transformations



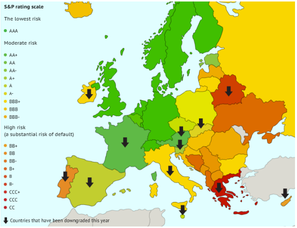
Relaxing of the non-linear transformations



Outline

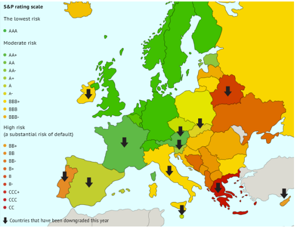


Sovereign Credit Rating



Sovereign Credit Rating has had an increasing importance since the beginning of the financial crisis

Sovereign Credit Rating



Sovereign Credit Rating has had an increasing importance since the beginning of the financial crisis

- Credit rating agencies **opacity has been criticised** by several authors, highlighting the suitability of designing more objective alternative methods
- Here we address the sovereign credit rating classification problem within an ordinal classification perspective

Work flow I

Input Variables

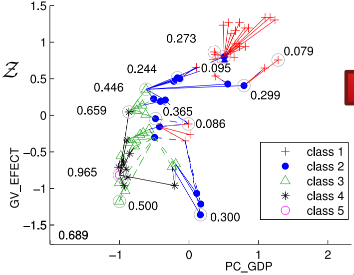
- Real GDP growth
- GDP per capita
- Government debt
- Fiscal balance
- External debt
- ...

Output Labels

AAA, AA-, BBB, ...

$\phi : \mathcal{X} \rightarrow \mathcal{Z}$

Pairwise Class Projection



Work flow II

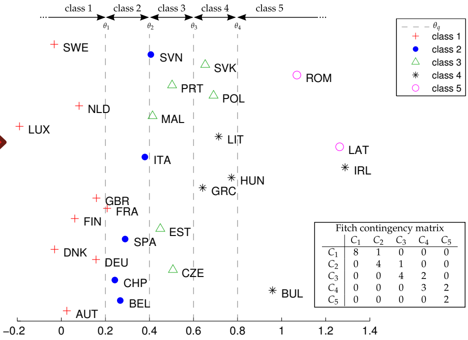
Regressor Training

Train a ϵ-SVR regressor to estimate the projection

$$g : \mathcal{X} \rightarrow \mathcal{Z}$$

$$\hat{z} = g(\mathbf{x})$$

Support vector regression prediction



Fitch contingency matrix

	C_1	C_2	C_3	C_4	C_5
C_1	8	1	0	0	0
C_2	0	4	1	0	0
C_3	0	0	4	2	0
C_4	0	0	0	3	2
C_5	0	0	0	0	2

Experimental Results I

Method/DataSet	Accuracy Fitch	Accuracy Moody's	Accuracy S&P	MAE Fitch	MAE Moody's	MAE S&P
C4.5	0.6296	0.6667	0.5926	0.4074	0.4074	0.4815
Mlogistic	0.4815	0.7778	0.3704	0.8889	0.3333	0.8889
MLP	0.6667	**0.8519**	0.6667	0.4074	0.2593	0.4444
Slogistic	*0.7407*	0.7778	*0.7037*	*0.2593*	0.2963	0.4074
ASAOR(C4.5)	0.5926	0.6296	*0.7037*	0.4815	0.4815	0.4074
RED-SVM	0.6667	*0.8148*	0.6667	0.3333	**0.2222**	0.4074
GPOR	*0.7407*	0.7037	0.6667	0.3704	0.4444	0.4444
SVOREX	0.7037	0.7778	0.5926	0.2963	0.2593	0.4444
SVORIM	0.6667	*0.8148*	0.6296	0.3333	**0.2222**	*0.3704*
SVR-PCDOC	**0.7778**	*0.8148*	**0.7407**	**0.2222**	**0.2222**	**0.2593**

The best result is in bold face and the second best result in italics

Experimental Results II

Method/DataSet	AMAE Fitch	AMAE Moody's	AMAE S&P	τ_b Fitch	τ_b Moody's	τ_b S&P
C4.5	0.4400	0.6800	0.5111	0.7621	0.7367	0.7655
Mlogistic	1.1600	0.6467	0.9333	0.5255	0.7719	0.5121
MLP	0.5267	**0.4067**	0.4000	0.7972	0.8097	0.7492
Slogistic	0.2667	0.6200	0.5111	*0.8951*	0.8151	0.8060
ASAOR(C4.5)	0.4533	0.7533	0.4222	0.6989	0.6655	0.7570
RED-SVM	0.2822	*0.5356*	0.4222	0.8835	0.8590	0.8052
GPOR	0.5133	0.9200	0.6222	0.7738	0.6869	0.7807
SVOREX	*0.2422*	0.5622	0.4444	0.8886	**0.8610**	0.7873
SVORIM	0.2756	*0.5356*	*0.3556*	0.8799	0.8525	*0.8370*
SVR-PCDOC	**0.2089**	0.5467	**0.2889**	**0.9224**	**0.8610**	**0.8849**

The best result is in bold face and the second best result in italics

Wind Forecasting

- Wind farm managers need forecasting of wind speed to manage the farm (e.g. wind turbines stop)
- Wind speed had been studied as a standard regression problem
- Managers need a general idea of the level of speed → **ordinal categories**
- Simplification of the problem can help to improve accuracy of the models

Wind Speed Categories

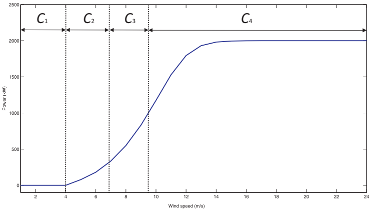


Figure: Wind speed classes ($C_1 \prec C_2 \prec C_3 \prec C_4$) and its relationship with the power curve of the wind turbines.

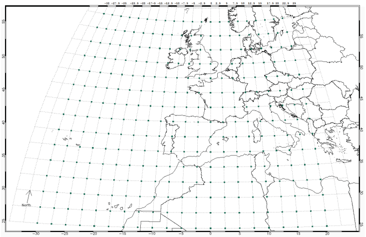


Figure: Synoptic pressure grid considered (Sea Level Pressure values have been used in this chapter).

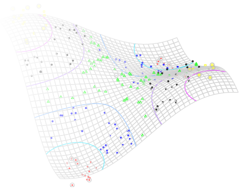
Experimental Results

Test Mean Absolute Error (*MAE*) results

Classifier	Wind farm H	M	P	U	Z	$\overline{M}$	$\overline{R}_M$
SVM	**0.242**	*0.267*	0.365	0.382	*0.300*	0.311	2.90
LMT	0.250	0.293	0.459	0.383	0.373	0.352	6.20
C45	0.335	0.310	0.540	0.434	0.487	0.421	10.90
Ada10(C45)	0.314	0.318	0.492	*0.381*	0.420	0.385	8.60
Ada100(C45)	0.260	0.281	0.419	0.389	0.354	0.341	6.00
MLogistic	0.258	0.288	0.514	0.433	0.405	0.379	7.80
SLogistic	0.250	0.293	0.495	0.434	0.400	0.374	7.70
ASAOR(C45)	0.293	0.299	0.465	0.438	0.463	0.392	9.40
RED-SVM	**0.242**	**0.261**	0.364	0.382	**0.295**	**0.309**	**2.20**
SVOREX	*0.245*	0.268	**0.354**	**0.378**	0.317	*0.312*	2.70
SVORIM	0.248	*0.267*	*0.355*	**0.378**	0.314	*0.312*	*2.60*
GPOR	0.289	0.316	0.526	0.472	0.513	0.423	11.00
HMM	0.301	0.322	0.646	0.525	0.535	0.466	12.60

The best result is in bold face and the second best result in italics

Outline



Conclusions and discussion I

1. State of the art in ordinal regression objectives: taxonomy and review, datasets and performance metrics
2. Class imbalance: linear combination of continuous error functions:
   - Improves computational time
   - Produces an unique candidate solution
   - Lessons for EELMOR

Conclusions and discussion II

3. Data ordering exploitation:
    - Considering patterns distribution through space can improve performance
    - Thresholds can be fixed so we reduce the number of free-parameters
    - Strong pressure in the projection is not always the best option
    - Not always the ordinal regression methods have the better performance

Future Work

Class imbalance techniques for ordinal regression

Future Work

Class imbalance techniques for ordinal regression

PCD projection is sensitive to **outliers**

Future Work

Class imbalance techniques for ordinal regression

PCD projection is sensitive to **outliers** $\rightarrow$ improve **robustness**

Future Work

Class imbalance techniques for ordinal regression

PCD projection is sensitive to **outliers** $\rightarrow$ improve **robustness**

Why ordinal methods are not achieving the best results in *ordinal* datasets?

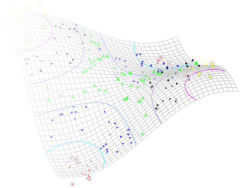
Future Work

Class imbalance techniques for ordinal regression

PCD projection is sensitive to **outliers** → improve **robustness**

Why ordinal methods are not achieving the best results in *ordinal* datasets? → How to **learn and evaluate data ordering**

Outline



8 Publications

Publications I

[1] P.A. Gutiérrez, M. Pérez-Ortiz, F. Fernandez-Navarro, **J. Sánchez-Monedero**, and C. Hervás-Martínez.
An Experimental Study of Different Ordinal Regression Methods and Measures.
In *7th International Conference on Hybrid Artificial Intelligence Systems*, pages 296–307, 2012.

[2] P.A. Gutiérrez, M. Pérez-Ortiz, **J. Sánchez-Monedero**, F. Fernández-Navarro, and C. Hervás-Martínez.
Ordinal regression methods: survey and experimental study.
Under Review, 2013.

[3] Manuel M. Cruz-Ramírez, C. Hervás-Martínez, **J. Sánchez-Monedero**, and P.A. Gutiérrez.
A Preliminary Study of Ordinal Metrics to Guide a Multi-Objective Evolutionary Algorithm.
In *Proceedings of the 11th International Conference on Intelligent Systems Design and Applications (ISDA 2011)*, pages 1176–1181, Cordoba, Spain, nov 2011.

[4] M. Cruz-Ramírez, C. Hervás-Martínez, **J. Sánchez-Monedero**, and P.A. Gutiérrez.
Metrics to guide a multi-objective evolutionary algorithm for ordinal classification.
Neurocomputing, Accepted, 2013.

[5] **J. Sánchez-Monedero**, C. Hervás-Martínez, F.J. Martínez-Estudillo, M. Carbonero Ruz, M. C. Ramírez Moreno, and M. Cruz-Ramírez.
Evolutionary learning using a sensitivity-accuracy approach for classification.
In *5th International Conference on Hybrid Artificial Intelligence Systems (HAIS2010)*, pages 0–0, 2010.

[6] **J. Sánchez-Monedero**, C. Hervás-Martínez, P.A. Gutiérrez, M. Carbonero-Ruz, M. C. Ramírez-Moreno, and M. Cruz-Ramírez.
Evaluating the Performance of Evolutionary Extreme Learning Machines by a Combination of Sensitivity and Accuracy Measures.
Neural Network World, 20:899–912, 2010.

Publications II

[7] **J. Sánchez-Monedero**, P.A. Gutiérrez, F. Fernández-Navarro, and C. Hervás-Martínez.
Weighting efficient accuracy and minimum sensitivity for evolving multi-class classifiers.
Neural Processing Letters, 34(2):101–116, 2011.

[8] **J. Sánchez-Monedero**, P.A. Gutiérrez, and C. Hervás-Martínez.
Evolutionary ordinal extreme learning machine.
In *International Conference on Hybrid Artificial Intelligence Systems (HAIS2013)*, volume Accepted, 2013.

[9] **J. Sánchez-Monedero**, M. Carbonero-Ruz, D. Becerra-Alonso, F. Martínez-Estudillo, P.A. Gutiérrez, and C. Hervás-Martínez.
Numerical variable reconstruction from ordinal categories based on probability distributions.
In *Proceedings of the 11th International Conference on Intelligent Systems Design andApplications (ISDA 2011)*, pages 1182–1187, Cordoba, Spain, Spain, nov 2011.

[10] **J. Sánchez-Monedero**, P. A. Gutiérrez, P. Tiňo, and C. Hervás-Martínez.
Exploitation of pairwise class distances for ordinal classification.
Neural Computation, 25(9), 2013.

[11] P.A. Gutiérrez, S. Salcedo-Sanz, C. Hervás-Martínez, L. Carro-Calvo, **J. Sánchez-Monedero**, and L. Prieto.
Ordinal and nominal classification of wind speed from synoptic pressure patterns.
Engineering Applications of Artificial Intelligence, 26(3):1008–1015, 2013.

[12] **J. Sánchez-Monedero**, P. Campoy-Muñoz, P. A. Gutiérrez, and C. Hervás-Martínez.
A guided data projection technique for classification of sovereign ratings: the case of european union 27.
Applied Soft Computing, Under Review, 2012.

¡Gracias!

¿Preguntas?

Questions?

Retos en clasificación ordinal: redes neuronales artificiales y métodos basados en proyecciones

Challenges in ordinal classification: artificial neural networks and projection-based methods

Tesis Doctoral

Universidad de Granada
Doctorado en Tecnologías de la Información y la Comunicación

Javier Sánchez Monedero
jsanchezm at uco dot es

Directores:
César Hervás Martínez y Pedro Antonio Gutiérrez Peña

Dept. of Computer Science and Numerical Analysis, University of Córdoba.
Learning and Artificial Neural Networks (AYRNA) research group.
http://www.uco.es/ayrna/

4th September 2013