Efficient Learning of Classification Models from Soft-label Information by Binning and Ranking

Yanbing Xue Milos Hauskrecht
Department of Computer Science, University of Pittsburgh

Abstract

- Building of classification models requires labeled examples
- Real-world data are often not labeled
- Human annotation/labeling process may be:
  - Time consuming & Costly
- **Challenge:**
- Find ways of reducing the annotation/labeling effort
- **Solution:**
- Learning with Soft-labels
  - Soft labels reflect the degree of belief of an annotator in the class labels
- **This work:**
- New method based on soft-label binning
- Reduces the number of constraints

Introduction

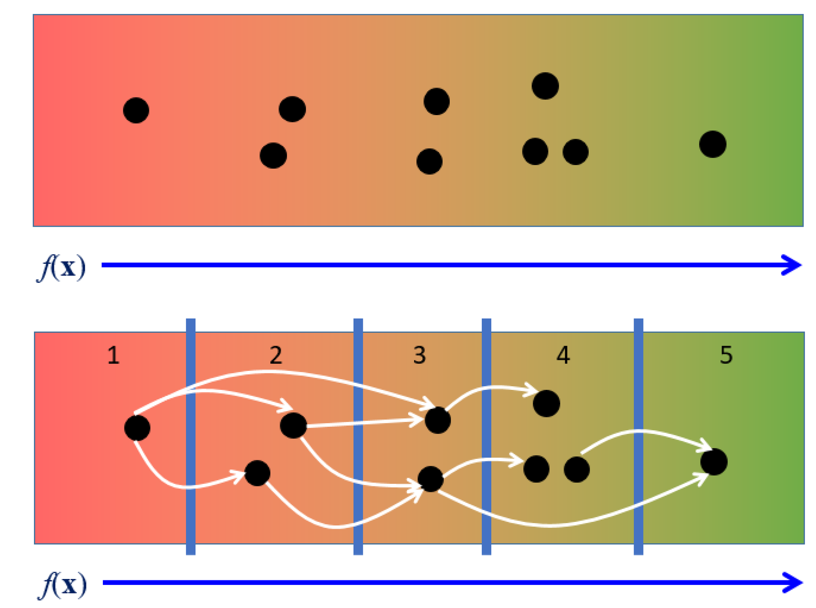
- **Soft labels:**
- represent the degree of belief of an annotator in the class label using a probabilistic score
- can facilitate learning of classification models

Class 0  Class 1

- **Caveat:**
- Humans are not good in providing consistent probabilistic assessments
  - Soft labels may be noisy
  - Learning methods should be robust to noisy soft labels

Related Work

- Existing works in literature show that there are two ways to learn with soft-label information: (1) Regression on the exact soft labels p_i, and (2) Respecting pairwise orderings rather than exact soft labels.
- **Regression on the exact soft labels p_i**
- **Idea:**
- Find a regression function f(x) that fits a logistic or linear regression model.
- **Disadvantage:**
- does not work well when soft-labels are noisy
- **Respecting pairwise orderings rather than exact soft labels**
- **Idea:**
- Find a discriminative projection f(x) that respects pairwise orderings of samples as much as possible
- **Disadvantage:**
- number of constraints is $O(N^2)$



Choosing the number of bins

How to choose the number of bins?

N bins: equivalent to all pairwise constraints
2 bins: equivalent to binary classification
The optimal number is somewhere between

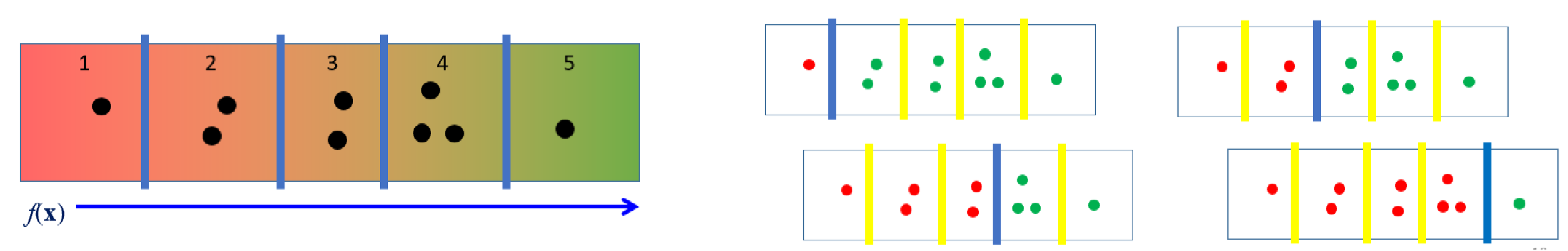
Heuristic function:

Inspired by optimal binning for discretization of continuous values

$$\#\text{of bins} = floor(\sqrt[3]{N})$$

Reducing constraints via binning

- Reformulate as ordinal regression
- **Idea:**
- solve jointly m-1 binary classifiers



- w: the set of parameters defining f(x)
- b_j: jth bin boundary
- **Advantage:**
- number of constraints is O(mN)

regularization term

errors for violating bin/sample constraints

$$\min_{\mathbf{w},\mathbf{b},\xi} \frac{\mathbf{w}^T\mathbf{w}}{2} + C\sum_{j=1}^{m-1}\sum_{i=1}^{N}\xi_{j,i}$$

$$\mathbf{w}^T\mathbf{x}_i - b_j \le \xi_{j,i} - 1 \quad \forall bin(p_i) \le j$$

$$\mathbf{w}^T\mathbf{x}_i - b_j \ge 1 - \xi_{j,i} \quad \forall bin(p_i) > j$$

$$\xi_{j,i} \ge 0 \quad \forall i, j$$

constraints defining errors between samples and bin boundaries

Experiments and Results

We test our approach on both synthetic and real-world data.

Noise generation:
Weak noise
Strong noise

BinarySVM
Labels: Binary only
Model: Support vector machine

SoftLogReg
Labels: Soft only
Model: Logistic regression

GPR
Labels: Binary + soft
Model: Gaussian process regression

SoftSvmRankPair
Labels: Binary + soft
Model: SVM + pairwise constraints

SoftSvmRankKN
Labels: Binary + soft
Model: SVM + mN random pairwise constraints

SoftSvmOrd- our method
Labels: Binary + soft
Model: SVM + bin/sample constraints

Figure 1. UCI Wine Red dataset (No noise)

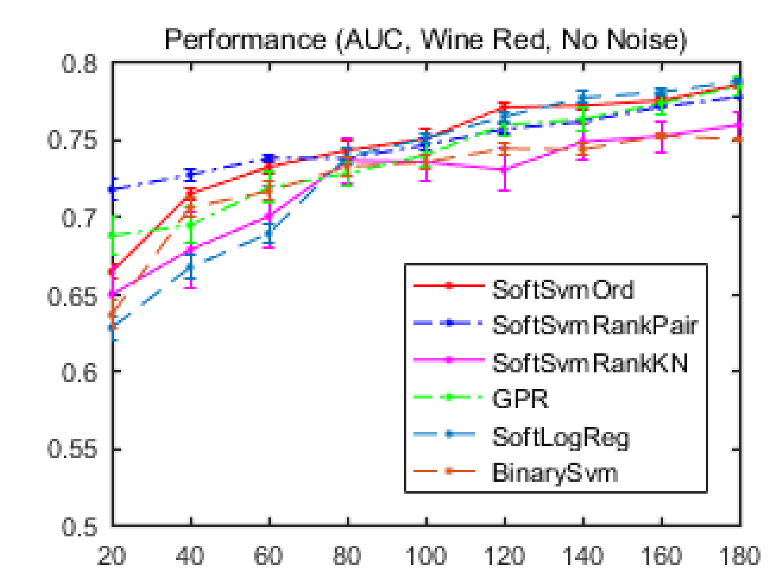


Figure 2. UCI Wine Red dataset (Weak noise)

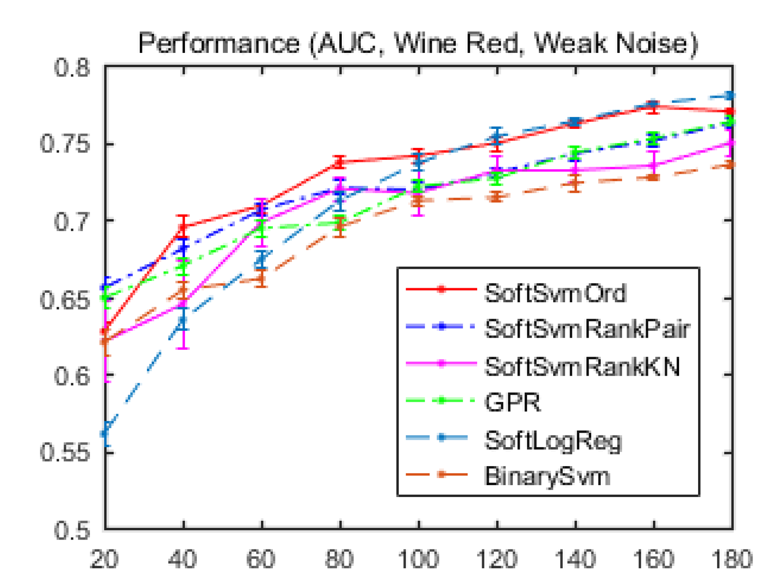
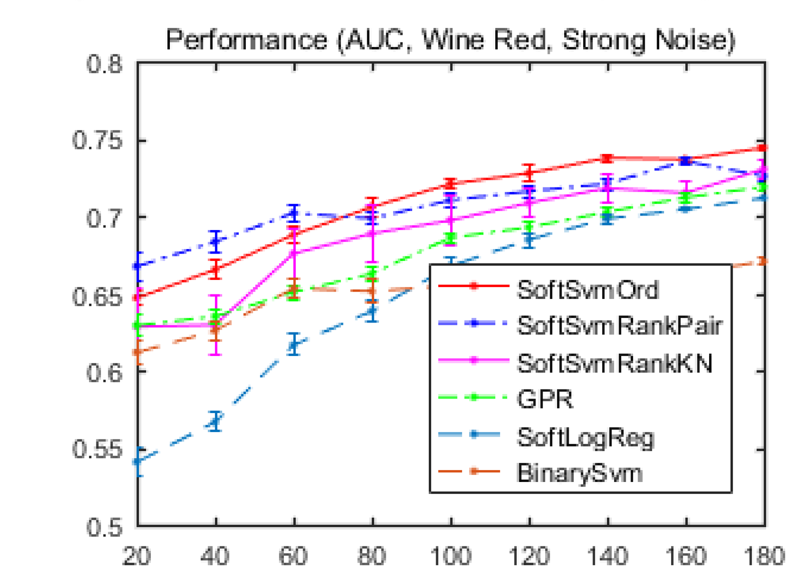


Figure 3. UCI Wine Red dataset (Strong noise)



Contact Information

Yanbing Xue, Milos Hauskrecht
University of Pittsburgh
Email: {yanbing, milos}@cs.pitt.edu
Website: http://people.cs.pitt.edu/~milos/
Phone: 412-624-8845

Conclusion

We develop a new robust method that uses soft-label information to reduce annotation effort.
Our method (1) can benefit greatly from soft-label information, and (2) is robust to different levels of soft-label noise.

Acknowledgements

The work presented in this paper was supported in part by grants R01GM088224 and R01LM010019 from the NIH. The content of the paper is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.