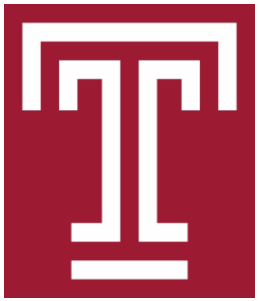
Supervised Clustering of Label Ranking Data

Mihajlo Grbovic, Nemanja Djuric, Slobodan Vucetic
{mihajlo.grbovic, nemanja.djuric, slobodan.vucetic}@temple.edu

SIAM SDM 2012, Anaheim, California, USA



Temple University
Department of Computer and Information Sciences
Center for Data Analytics and Biomedical Informatics
Philadelphia, USA

Outline

- Introduction
  - Label Ranking
    - Performance Measures
    - Related Work
  - Supervised clustering in context of Label Ranking
    - Motivation
    - Performance Measures
- Approaches
  - Baseline Approaches
  - Placket-Luce Mixture Model
- Empirical Evaluation
  - Experiments on Synthetic Data
  - Experiments on Real-world Data

Introduction

- Label Ranking Setup: L = 5 labels

	Costumer Features					Product
	Features (**x**)					*y*
1	0	21.2	-1.2	1	12	1
2	1	-1.8	-1.1	0	14	3
3	1	-2.5	0.3	1	8	3
4	0	7.5	-0.2	1	11	2

- Costumer Features: age, gender, how often they buy from us, how much on average they spend, etc.

Grbovic M., Djuric N., Vucetic S., Supervised Clustering of Label Ranking Data, SIAM SDM 2012

Introduction

- Label Ranking Setup: L = 5 labels

	Costumer Features (x)					Product Ranking (π)				
1	0	21.2	-1.2	1	12	1	4	2	5	3
2	1	-1.8	-1.1	0	14	3	2	5	1	4
3	1	-2.5	0.3	1	8	3	2	1	4	5
4	0	7.5	-0.2	1	11	2	1	5	4	3

$\pi = 1\ 4\ 2\ 5\ 3$, pairwise label preferences: $1 > 4 > 2 > 5 > 3$

- Goal: Learn a model that maps instances $\mathbf{x}$ to a total label order π

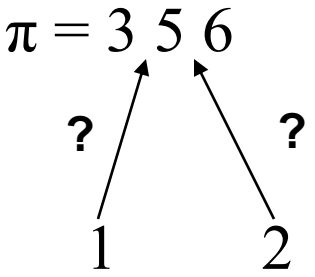
$$D = \{(\mathbf{x}_n, \pi_n), n = 1 \ldots N\} \qquad h : \mathbf{x}_n \rightarrow \pi_n$$

Introduction

- Label Ranking: Missing Information

	Features (x)					Label Ranking (π)				
1	0	21.2	-1.2	1	12	1	4			
2	1	-1.8	-1.1	0	14	3	5	4		
3	1	-2.5	0.3	1	8	2	1	4		
4	0	7.5	-0.2	1	11	1	5	3		

Partial Ranking



Introduction

- Label Ranking: Performance Measure

Notation: $\pi(i)$ the class label at i-th position in the order
$\pi^{-1}(j)$ the position of the y_j class label in the order

Distance between two rankings: true ranking (π) and predicted ranking (ρ):

$$d_\tau(\pi,\rho) = \left|\{(y_i, y_j) : \pi_n^{-1}(y_i) > \pi_n^{-1}(y_j)\} \mid \rho_n^{-1}(y_j) > \rho_n^{-1}(y_i)\}\right|$$

Kendall tau distance - counts the number of discordant label pairs

Given Data set: $D = \{(\mathbf{x}_n, \pi_n), n = 1 \ldots N\}$

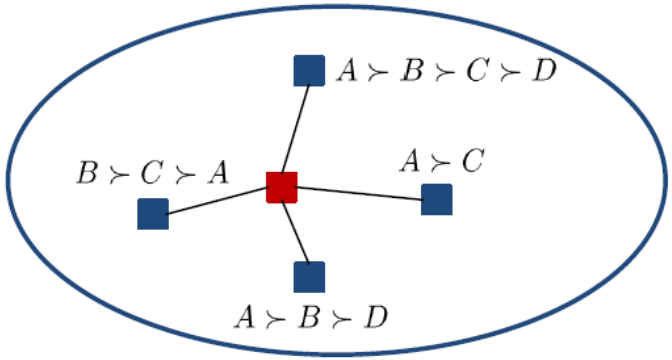
Label Ranking Loss: $loss_{LR} = \frac{1}{N}\sum_{n=1}^{N} \frac{2 \cdot d_\tau(\pi_n, \hat{\pi}_n)}{L \cdot (L-1)}$

Introduction

- Label Ranking: Related Work

1. Map into classification - $L(L\text{-}1)/2$ classifiers
   - 1 $(d \text{ x } L)$ dimensional problem

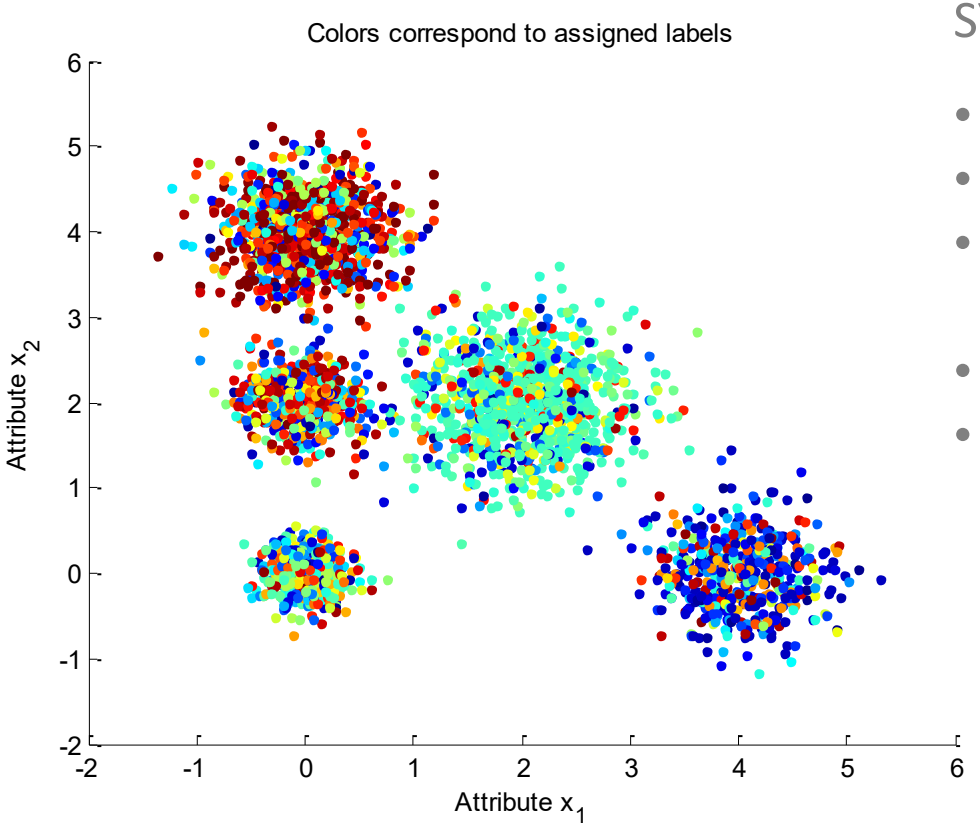
2. kNN based algorithms



3. Utility functions - Learn mappings $f_k : \mathbf{x} \rightarrow R, k = 1, \ldots, L$
   - Prediction: rank the utility scores

Introduction

- Label Ranking: Supervised Clustering



SYNTHETIC DATA

- 2 features
- 5 labels
- Each permutation represented with a color (similar color – similar rank)
- 5 natural clusters in feature space
- 3 natural clusters in label space

Introduction

- Label Ranking: Supervised Clustering

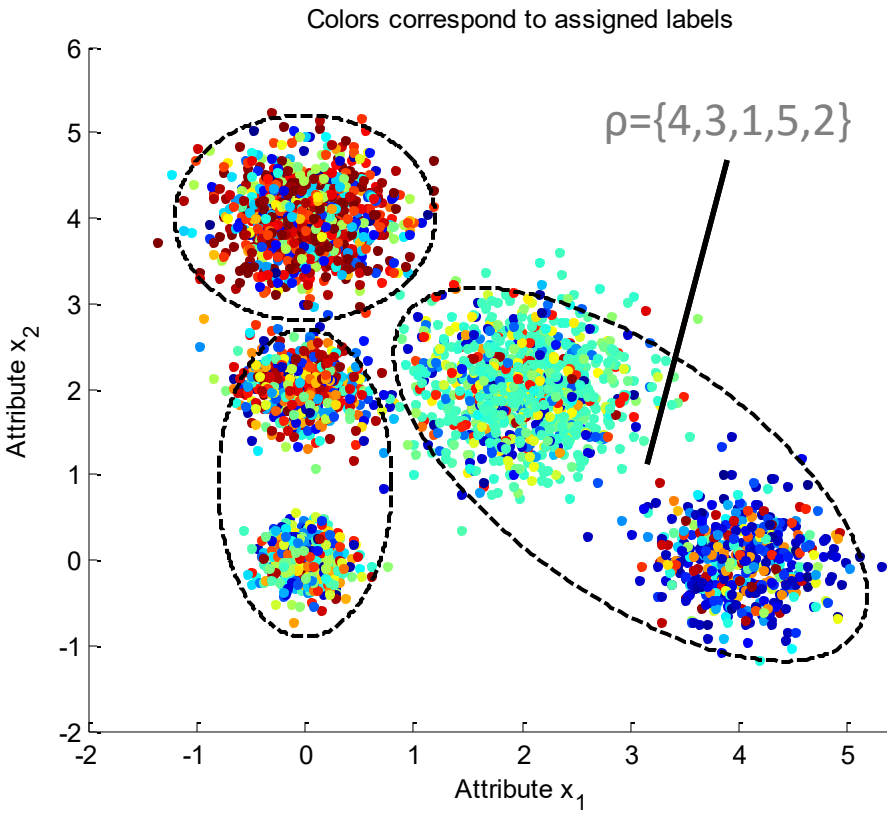
GOAL:

- Cluster data instances (customers) in the feature space by taking into consideration the assigned, potentially incomplete label rankings (product preferences)
- Such that the rankings of instances within a cluster are more similar to each other than to the rankings of instances in the other clusters
- Extract cluster centroid-rankings (preferences that represent each cluster uniquely)

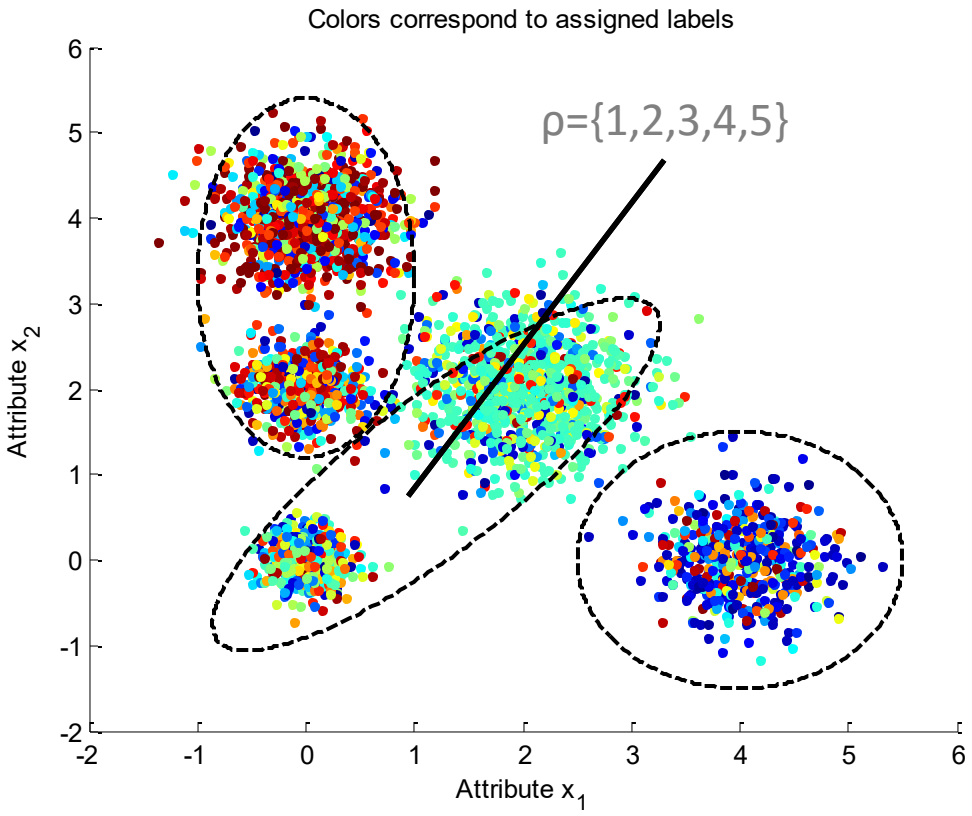
Introduction

- Label Ranking: Supervised Clustering

Traditional Clustering



Supervised Clustering



Introduction

- Label Ranking: Supervised Clustering

Example: **Target marketing**

- A company with several products would like to cluster its costumers (in feature space)
- Purpose: designing cluster-specific promotional material
- For each cluster, the company can make a different catalog, by promoting products in different order that best reflects the taste of its target costumers

Grbovic M., Djuric N., Vucetic S., Supervised Clustering of Label Ranking Data, SIAM SDM 2012

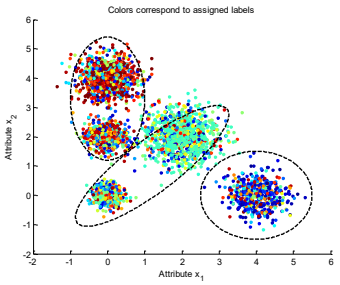
Introduction

- Label Ranking: Supervised Clustering

Performance Measures

- "Tightness" of clusters in label ranking space
  - How similar are the rankings of instances within the clusters
  - How far are cluster central ranking from cluster member rankings

$$\Delta_p = \frac{1}{K}\sum_{k=1}^{K}\frac{\sum_{n\in S_k} d_\tau(\pi_n, \rho_k)}{\mid S_k \mid},$$



- Happiness of new costumer when he receives the catalog by mail
  - How close is the cluster central ranking to true costumer ranking

$$loss_{LR} = \frac{1}{N}\sum_{n=1}^{N}\frac{2\cdot d_\tau(\pi_n, \hat{\pi}_n)}{L\cdot(L-1)}$$

Grbovic M., Djuric N., Vucetic S., Supervised Clustering of Label Ranking Data, SIAM SDM 2012

Approaches

- Heuristic Baselines

1. Cluster in Feature Space → Find Central Cluster Rankings

Kmeans → Mallows

2. Cluster in Label Ranking Space → Multi-Class Classification

Naïve → SVM
EBMS* → SVM

3. Add Label Rankings to Features → Unsupervised Clustering

Naïve Kmeans

4. 1-Rank (represent all data using one ranking)

* M. Meila and L. Bao, An exponential model for infinite rankings, Journal of Machine Learning Research, 11 (2010)

Approaches

- Plackett-Luce Mixture Model

The Plackett-Luce (PL) model is specified by a parameter vector $\boldsymbol{v} = (v_1, v_2, \ldots v_m) \in \mathbb{R}_+^m$:

$$\mathbf{P}(\pi \mid \boldsymbol{v}) = \prod_{i=1}^{m} \frac{v_{\pi(i)}}{v_{\pi(i)} + v_{\pi(i+1)} + \ldots + v_{\pi(m)}}$$

Reduces problem to learning a mapping $\boldsymbol{x} \mapsto \boldsymbol{v}$.

Example: $\boldsymbol{v} = (1, 4, 2)$, $\mathbf{P}(\pi \mid \boldsymbol{v}) = \frac{v_{\pi(1)}}{v_{\pi(1)}+v_{\pi(2)}+v_{\pi(3)}} \cdot \frac{v_{\pi(2)}}{v_{\pi(2)}+v_{\pi(3)}} \cdot 1$

1	2	3	0.0952
1	3	2	0.0476
2	1	3	0.1905
2	3	1	0.0571
3	1	2	0.3810
3	2	1	0.2286

Approaches

- Plackett-Luce Mixture Model (K clusters)

$$\mathbb{P}(\pi \mid \mathbf{x}) = \sum_{k=1}^{K} \mathbb{P}(k \mid \mathbf{x}) \cdot \mathbb{P}(\pi \mid k)$$

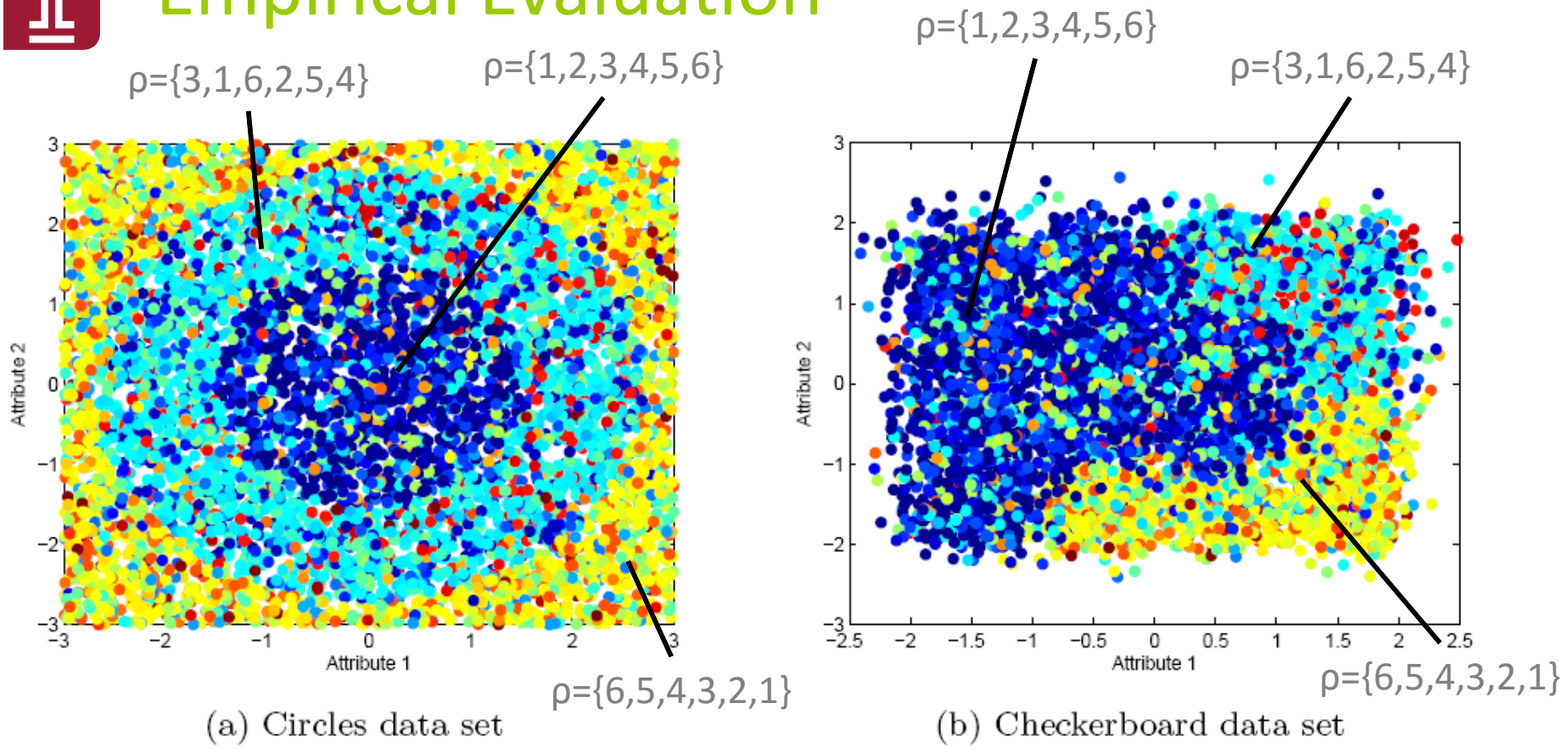
K clusters: $\lambda = \{\{\mathbf{m}_p\}_k, \mathbf{v}_k, k = 1, ..., K, \sigma_p\}$

Likelihood:

$$\mathbb{P}(\boldsymbol{\pi} \mid \lambda) = \prod_{n=1}^{N} \sum_{k=1}^{K} \mathbb{P}(k \mid \mathbf{x}) \cdot \prod_{i=1}^{L_n} \frac{\mathbf{v}_k(\pi_n(i))}{\sum_{j=i}^{L_n} \mathbf{v}_k(\pi_n(j))}$$

Grbovic M., Djuric N., Vucetic S., Supervised Clustering of Label Ranking Data, SIAM SDM 2012

Empirical Evaluation



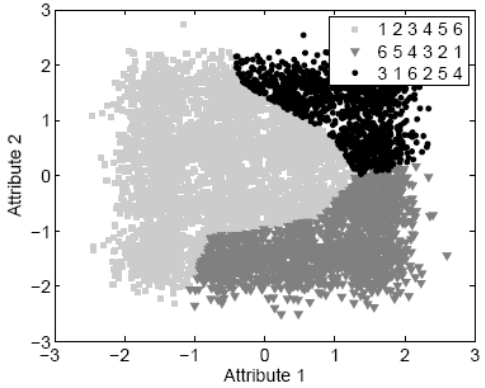
(a) Circles data set (b) Checkerboard data set

The resulting cluster central rankings

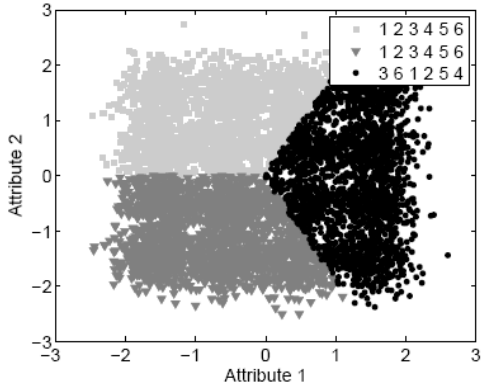
Data Set	complete ranking					30% missing labels					60% missing labels				
	Kmean *Mal*	Naive *Kmean*	Naive *SVM*	EBMS *SVM*	MM *PL*	Kmean *Mal*	Naive *Kmean*	Naive *SVM*	EBMS *SVM*	MM *PL*	Kmean *Mal*	Naive *Kmean*	Naive *SVM*	EBMS *SVM*	MM *PL*
checker	**123456**	**123456**	**123456**	132654	**123456**	**123456**	**123456**	**123456**	132546	**123456**	**123456**	**123456**	**123456**	132645	**123456**
$K = 3$	123456	**654321**	**654321**	653421	**654321**	123456	**654321**	**654321**	132645	**654321**	316254	**654321**	124356	132645	**654321**
	361524	423516	213456	653421	**316254**	361254	136254	124356	132645	**316254**	316452	123645	123546	132546	**316254**
circle	361254	**123456**	136254	**123456**	**123456**	361524	132654	312654	653421	**123456**	361524	635124	316524	316254	**123456**
$K = 3$	316254	**654321**	**654321**	132546	**654321**	361524	**654321**	**654321**	653421	**654321**	361254	316254	361254	316254	**654321**
	361524	**316254**	**316254**	312654	**316254**	361524	654321	**316254**	132654	**316254**	361524	**316254**	**316254**	**316254**	**316254**

Empirical Evaluation

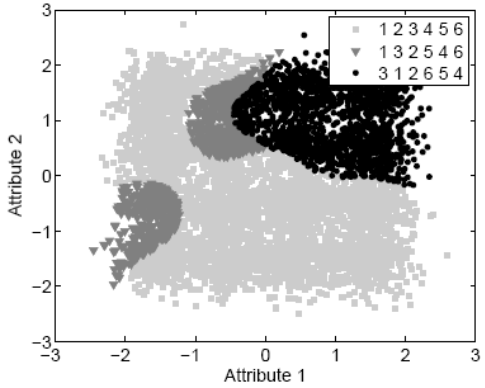
$$\Delta_p = \frac{1}{K}\sum_{k=1}^{K}\frac{\sum_{n\in S_k} d_\tau(\pi_n, \rho_k)}{|\,S_k\,|}$$



(a) $\underset{PL}{\text{MM}}$, $\Delta_p = .526$

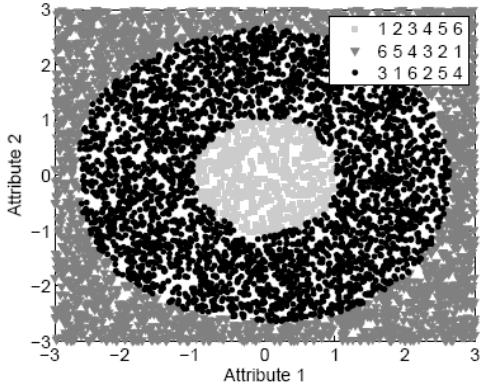


(b) $\underset{Mal}{\text{Km}}$, $\Delta_p = .771$

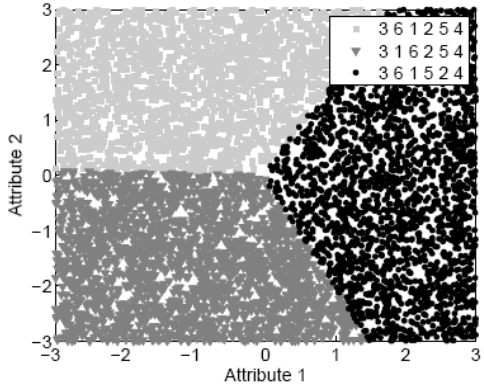


(c) $\underset{SVM}{\text{ebms}}$, $\Delta_p = .617$

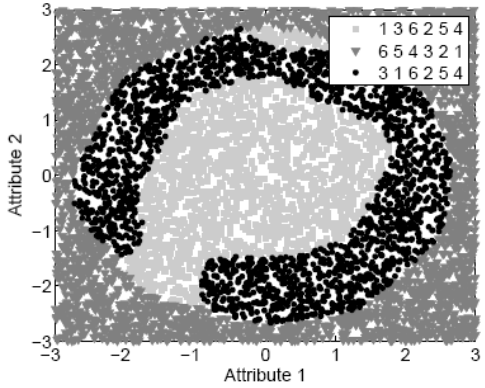
Clustering performance comparison on *checker* dataset ($K = 3$)



(a) $\underset{PL}{\text{MM}}$, $\Delta_p = .532$



(b) $\underset{Mal}{\text{Km}}$, $\Delta_p = .831$

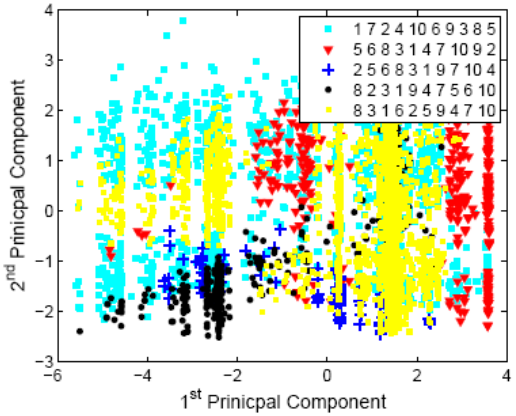


(c) $\underset{SVM}{\text{naive}}$, $\Delta_=.552$

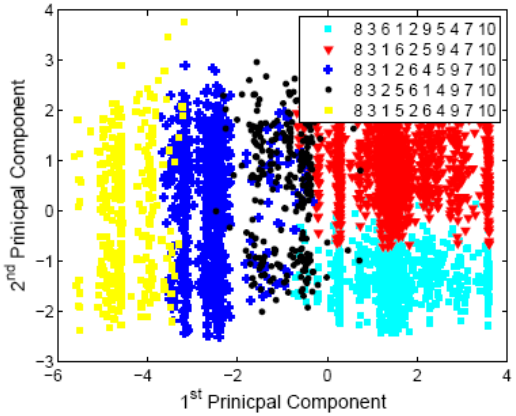
Clustering performance comparison on *circle* dataset ($K = 3$)

Empirical Evaluation

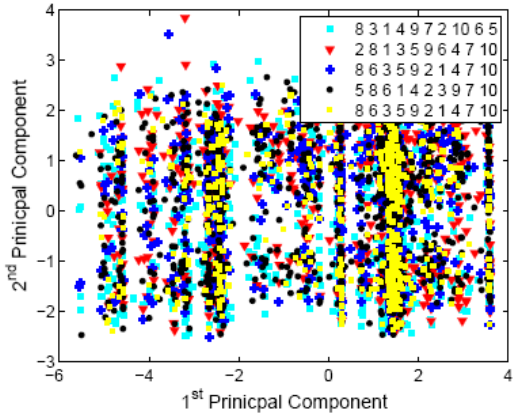
- Sushi Data Set (L=10)



(a) $\underset{PL}{\text{MM}}$, $\Delta_p = .641$



(b) $\underset{Mal}{\text{Km}}$, $\Delta_p = .695$



(c) $\underset{SVM}{\text{ebms}}$, $\Delta_p = .645$

Clustering performance comparison on *sushi* dataset ($K = 5$)

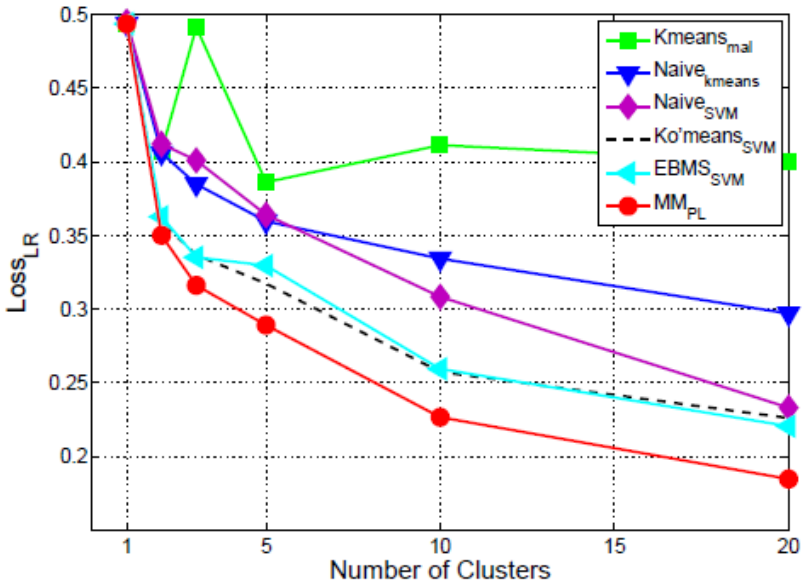
The resulting cluster central rankings for Sushi data set $K = 5$, no missing labels

1 Rank		Km $_{Mal}$	naive $_{Km}$	naive $_{SVM}$	ebms $_{SVM}$	MM $_{PL}$
		8 3 6 1 2 9 5 4 7 10	5 2 1 8 6 4 3 9 7 10	2 5 6 8 3 1 9 7 10 4	8 3 1 4 9 7 2 10 6 5	1 7 2 4 10 6 9 3 8 5
		8 3 1 6 2 5 9 4 7 10	1 7 4 2 3 10 9 8 6 5	5 6 8 3 1 4 9 2 7 10	2 8 1 3 5 9 6 4 7 10	5 6 8 3 1 4 7 10 9 2
8 3 1 6 2 5 9 4 7 10		8 3 1 2 6 4 5 9 7 10	8 3 9 1 2 6 4 7 10 5	8 3 6 5 9 2 1 4 10 7	8 6 3 5 9 2 1 4 7 10	2 5 6 8 3 1 9 7 10 4
		8 3 2 5 6 1 4 9 7 10	8 5 2 6 3 9 1 7 4 10	8 5 6 2 1 3 9 4 10 7	5 8 6 1 4 2 3 9 7 10	8 2 3 1 9 4 7 5 6 10
		8 3 1 5 2 6 4 9 7 10	8 5 3 6 1 4 9 2 7 10	1 8 3 9 2 4 7 10 6 5	8 6 3 5 9 2 1 4 7 10	8 3 1 6 2 5 9 4 7 10
Δ_p	.695	.778	.652	.645	**.641**	

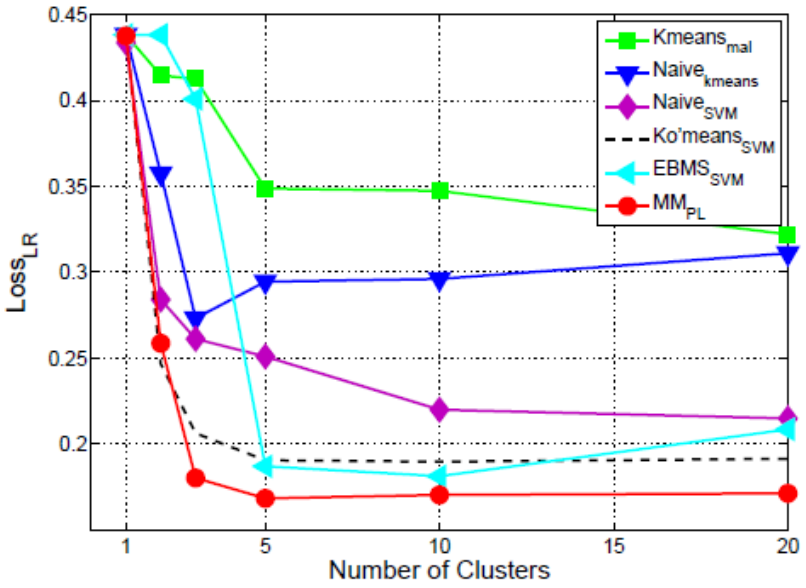
Empirical Evaluation

name	domain	N	d	L	1-Rank $loss_{LR}$	name	domain	N	d	L	1-Rank $loss_{LR}$
circles	S	7,000	2	6	.393	checker	S	7,200	2	6	.398
authorship	SS (A)	841	70	4	.269	iris	SS (B)	150	4	3	.451
bodyfat	SS (B)	252	7	7	.450	pendigits	SS (A)	10,992	16	10	.383
calhousing	SS (B)	20,640	4	4	.380	segment	SS (A)	2,310	18	7	.419
cpu-small	SS (B)	8,192	6	5	.431	stock	SS (B)	950	5	5	.445
elevators	SS (B)	16,599	9	9	.435	vehicle	SS (A)	846	18	4	.335
fried	SS (B)	40,769	9	5	.493	vowel	SS (A)	528	10	11	.334
glass	SS (A)	214	9	6	.171	wine	SS (A)	178	13	3	.234
housing	SS (B)	506	6	6	.412	wisconsin	SS (B)	194	16	16	.485
cold	biology	2,465	24	4	.401	heat	biology	2,465	24	6	.464
diau	biology	2,465	24	7	.348	spo	biology	2,465	24	11	.441
dtt	biology	2,465	24	4	.438	sushi	food	5,000	11	10	.392

Empirical Evaluation



(a) Fried Data $Loss_{LR}$



(b) Elevator Data $Loss_{LR}$

Empirical Evaluation

Label Ranking Supervised Clustering performance comparison in terms of $loss_{LR}$

	complete ranking					30% missing labels					60% missing labels				
Data Set	Km *Mal*	naive *Km*	naive *SVM*	ebms *SVM*	MM *PL*	Km *Mal*	naive *Km*	naive *SVM*	ebms *SVM*	MM *PL*	Km *Mal*	naive *Km*	naive *SVM*	ebms *SVM*	MM *PL*
checker	.340	.287	.287	.373	**.250**	.340	.300	.287	.377	**.249**	.342	.308	.413	.393	**.260**
circle	.392	.444	.260	.281	**.253**	.392	.410	.265	.280	**.258**	.392	.396	.374	.395	**.261**
fried	.413	.334	.272	.263	**.226**	.412	.342	.291	.340	**.236**	.422	.395	.317	.375	**.251**
calhousing	.364	.424	.450	.359	**.349**	.365	.463	.467	.379	**.352**	.365	.501	.469	.390	**.349**
elevators	.347	.300	.220	.182	**.171**	.334	.301	.272	.248	**.210**	.336	.308	.293	.271	**.214**
pendigits	.290	.308	**.148**	.314	.172	.289	.287	.197	.326	**.181**	.291	.337	.236	.349	**.202**
cpu-small	.362	.374	.407	.403	**.356**	.363	.389	.390	.399	**.356**	.365	.413	.407	.387	**.360**
segment	.214	.188	**.080**	.185	.146	.214	.197	**.115**	.184	.145	.213	.229	.202	.226	**.156**
wisconsin	.451	.443	.486	.453	.433	.455	.443	.487	.462	.435	.465	.461	.490	.469	**.444**
vowel	.253	.274	.237	.221	.209	.255	.276	.266	.239	**.211**	.265	.314	.305	.256	**.242**
vehicle	.132	.113	.095	.124	.090	.131	.165	.080	.155	**.095**	.136	.210	.142	.174	**.109**
stock	.213	.218	**.138**	.235	.140	.217	.227	.175	.241	**.148**	.222	.259	.273	.310	**.162**
iris	.091	.059	.080	.171	.061	.093	.142	.100	.174	**.088**	.136	.188	.196	.256	**.127**
glass	.110	.113	.092	.112	.108	.117	.173	**.098**	.125	.114	.139	.211	.189	.144	**.130**
authorship	.072	.066	.101	.175	.079	.073	**.072**	.109	.179	.075	.082	**.077**	.193	.221	.082
bodyfat	.451	.454	.461	**.419**	.430	.455	.459	.464	.453	.436	.469	.464	.491	.458	**.444**
wine	.050	.058	.051	.106	**.047**	.051	.163	.059	.122	**.045**	.087	.220	.104	.133	**.076**
housing	.376	.407	**.331**	.366	.339	.377	.421	.381	.412	**.357**	.397	.436	.384	.426	**.365**
cold	.395	.408	.416	.426	**.386**	.405	.404	.448	.437	**.393**	.404	.418	.434	.438	**.398**
heat	.438	.453	.477	.465	**.433**	.440	.451	.495	.472	**.437**	.447	.456	.477	.485	**.439**
diau	.336	.385	.379	.356	**.327**	.339	.376	.408	.366	**.338**	.345	.378	.362	.382	**.343**
dtt	.416	.432	.471	.447	**.410**	.422	.428	.464	.449	**.417**	.424	.432	.466	.453	**.422**
spo	**.431**	.440	.484	.463	**.431**	**.434**	.445	.487	.474	.435	.443	.471	.481	.478	**.440**
sushi	.372	.383	.379	.366	**.332**	.369	.374	.414	.372	**.335**	.379	.394	.382	.477	**.343**
average	.305	.307	.284	.303	**.257**	.306	.321	.301	.319	**.264**	.315	.345	.336	.348	**.276**
time (s)	10.1	9.3	1300	13K	234	13.5	10.9	3190	13.2K	228	13.8	11.6	3730	13.3K	188

Empirical Evaluation

Label Ranking Prediction performance comparison in terms of label rank loss $loss_{LR}$

	complete ranking							30% missing labels					60% missing labels				
Data Set	1NN	IB $_{CPS}$	IB $_{PL}$	IB $_{Mal}$	Lin $_{LL}$	Lin $_{PL}$	MM $_{PL}$	IB $_{PL}$	IB $_{Mal}$	Lin $_{LL}$	Lin $_{PL}$	MM $_{PL}$	IB $_{PL}$	IB $_{Mal}$	Lin $_{LL}$	Lin $_{PL}$	MM $_{PL}$
fried	.248	.201	.202	**.196**	.344	.347	.294	**.216**	.219	.345	.348	.300	**.257**	.312	.349	.350	.313
calhousing	.386	.344	.337	**.336**	.449	.435	.351	**.345**	.348	.450	.436	.353	**.376**	.415	.453	.437	.356
elevators	.261	**.224**	.234	.226	.306	.305	.244	**.238**	.244	.309	.306	.249	**.263**	.322	.311	.307	.265
pendigits	**.142**	.140	.143	.139	.364	.351	.211	**.152**	.159	.366	.354	.215	**.169**	.230	.368	.355	.219
cpu-small	.392	.347	.344	**.336**	.378	.361	.348	.351	**.350**	.379	.369	**.350**	.371	.408	.380	.370	**.355**
segment	**.073**	.077	.089	.089	.241	.227	.132	**.111**	.117	.242	.228	.135	**.124**	.154	.242	.226	.149
wisconsin	.458	.492	.435	.419	.403	**.401**	.413	.434	.427	.409	**.403**	.418	.445	.443	.422	**.412**	.430
vowel	**.110**	.117	.158	.148	.313	.317	.169	.217	.178	.316	.318	**.173**	.234	.240	.329	.327	**.231**
vehicle	.097	.087	.082	**.079**	.126	.115	.083	**.093**	.095	.129	.118	**.093**	.113	.131	.128	.125	**.102**
stock	.107	.118	.106	.108	.283	.268	**.105**	.134	.133	.284	.269	**.130**	.154	.169	.285	.272	**.151**
iris	.043	.049	.040	**.037**	.189	.171	.047	**.059**	.066	.199	.172	.064	**.103**	.118	.208	.174	.105
glass	**.075**	.102	.098	.094	.107	.106	.090	.107	.111	.111	.108	**.103**	**.115**	.140	.124	.118	.120
authorship	.080	.063	.063	**.062**	.182	.175	**.062**	.068	.072	.189	.178	**.065**	.085	.108	.200	.190	**.081**
bodyfat	.472	.436	.442	.438	.422	**.404**	.423	.463	.448	.430	**.406**	.430	.466	.460	.431	**.429**	.439
wine	.068	.039	.040	.036	.067	.059	**.034**	.041	.042	.071	.064	**.038**	.061	.077	.101	.079	**.059**
housing	.323	**.321**	.329	.334	.398	.385	.330	.355	.357	.404	.386	**.348**	.384	.385	.408	.391	**.364**
cold	.425	.389	.386	.387	.409	.399	**.385**	.398	.398	.410	.404	**.387**	.424	.423	.411	.405	**.397**
heat	.446	.434	.438	.435	.439	.438	**.430**	.445	.433	.441	.442	**.432**	.460	.455	.445	.443	**.435**
diau	.382	.336	.336	.332	.350	.349	**.329**	.338	.340	.353	.350	**.332**	.339	.365	.360	.352	**.338**
dtt	.440	.420	.418	.417	.421	.410	**.410**	.428	.430	.422	.417	**.414**	.447	.442	.423	**.419**	.423
spo	.440	.435	.438	.433	.437	.443	**.428**	.439	.444	.439	.445	**.434**	.441	.455	.444	.453	**.436**
sushi	.415	.347	.348	.349	.459	.462	**.337**	.363	.352	.461	.465	**.339**	.381	.374	.463	.468	**.342**
average	.267	.253	.250	**.247**	.322	.315	.255	**.263**	.264	.325	.318	**.263**	.282	.301	.331	.323	**.277**
avg. rank	4.73	3.23	3.41	**2.55**	6	5.27	2.82	2.36	2.77	4.32	3.77	**1.77**	2.45	3.64	3.95	3.14	**1.82**

Conclusion

- Conclusion

  - This paper presents the first attempt at supervised clustering of complex label rank data

  - We established several baselines for supervised clustering of label ranking data and proposed a Plackett-Luce (PL) mixture model specifically tailored for this application

  - We empirically showed the strength of the PL model by experiments on real-world and synthetic data

  - In addition to the supervised clustering scenario, we compared the PL model to the previously proposed label ranking algorithms in terms of predictive accuracy

THANK YOU