Green Seminar Team

Vera Shalaeva

AMA Team, Laboratoire d'Informatique de Grenoble

October 01, 2015

My background

Education

- 2014 - 2015 *Master degree* - Institut polytechnique de Grenoble - Industrial and Applied Mathematics,
  Data science
- 2009 - 2011 *Master degree* - State University of Management [Moscow],
  Analytical Support for Business Decision
- 2005 - 2009 *Bachelor degree* - Volgograd State University,
  Applied Mathematics and Informatics

Experience

- 2015 Graduate Program Intern - AMA Team, Laboratoire d'Informatique de Grenoble
- 2012 - 2014 Data Analyst - Samsung Russia, Moscow
- 2011 - 2012 Leading specialist-expert - The Ministry of Economic Development of the Russian Federation, Moscow
- 2010 - 2011 EDI administrator - Procter & Gamble, Moscow

My last work

Master thesis

- *Inferring hierarchy structures in large-scale classification problems*,
  supervisors - Prof. Eric Gaussier, Prof. Massih-Reza Amini
  - Inferred the hierarchical structure that enough effective to yield the good results on large-scale classification problem.
  - Investigated and emphasized the PAC-Bayes bound for multi-class classification.

Introduction

- The amount of data has been growing rapidly.

- The demand on automatic classification of unseen data to one of tens of thousand target classes lead to *large-scale classification problems*.
- Applications :
  - text categorization according to the topic
  - image classification
  - classification a gene in bionformatics
- Conventional methods cannot be apply for large-scale problems :
  - one-versus-one $\Longrightarrow$ $\frac{K(K-1)}{2}$ classifiers need to be evaluated
  - one-versus-rest $\Longrightarrow$ K classifiers need to be evaluated

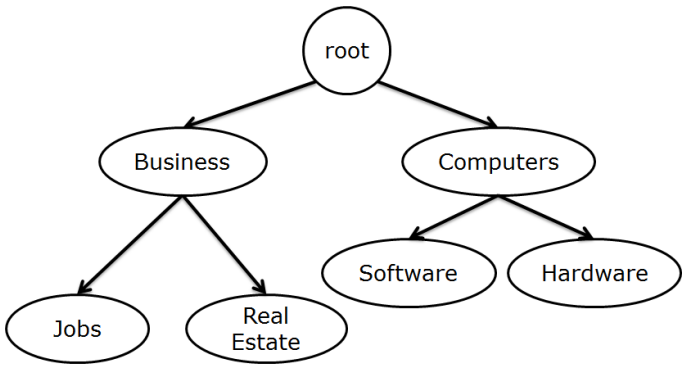
Introduction

Why ?

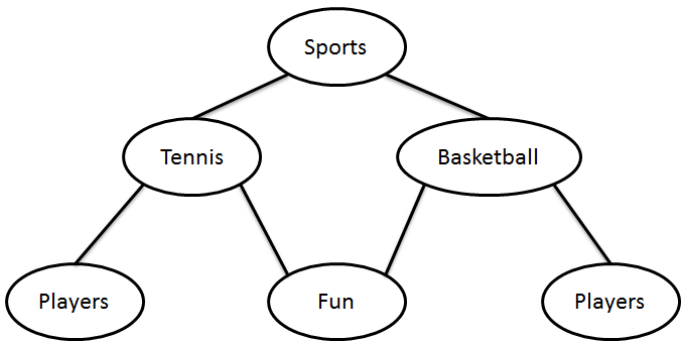
- computationally expensive
- impossible to apply it when immediate prediction is required

Solution : to organize classes in a hierarchy
Examples of large-scale taxonomies :

- DMOZ
- Wikipedia



(a) DMOZ Taxonomy



(b) Wikipedia Taxonomy

Strategy of inferring hierarchy

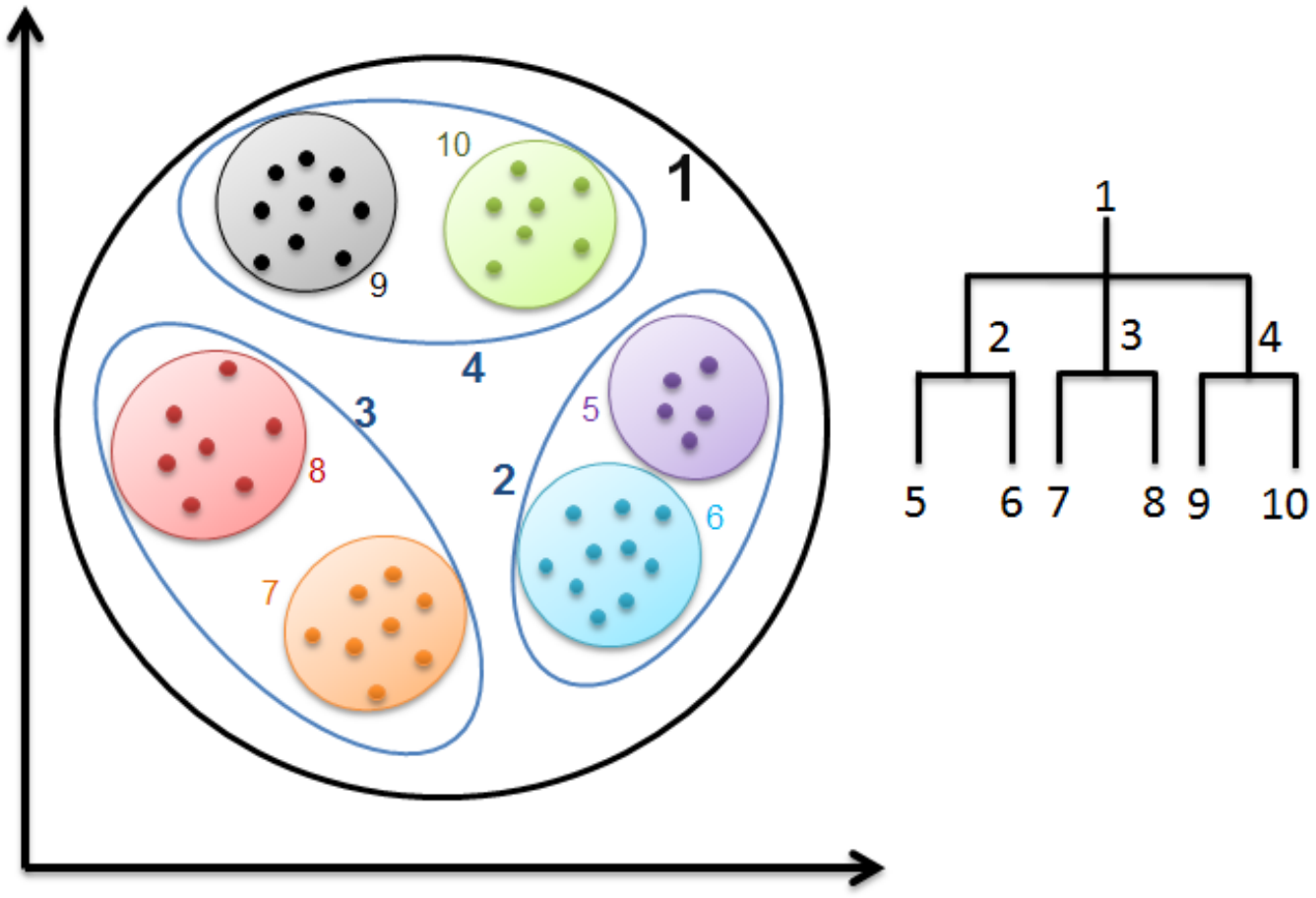


FIGURE : Building of hierarchy

Estimating the number of classes

Heuristic strategies :

- Square root

$$\underbrace{500}_{\text{leaves}} \rightarrow \underbrace{22}_{\text{1 level}} \rightarrow \underbrace{5}_{\text{2 level}} \rightarrow \underbrace{2}_{\text{3 level}} \rightarrow \underbrace{1}_{\text{root}}$$

- Cubic root

$$\underbrace{500}_{\text{leaves}} \rightarrow \underbrace{8}_{\text{1 level}} \rightarrow \underbrace{2}_{\text{2 level}} \rightarrow \underbrace{1}_{\text{root}}$$

Theoretical methods :

- Akaike Information Criterion (AIC)

$$\text{KMeans}_{AIC} : K = \arg\min_{K}[SSE(K) + 2VK]$$

V - the dimensionality of the dataset, $SSE(K)$ - Sum of Squared Errors

- Bayesian Information Criterion (BIC)

$$BIC = \frac{-2 * \ln(L)}{N} + \frac{K * \ln(N)}{N} = \frac{1}{N} \times \ln \frac{N^K}{L^2}$$

N - the number of points, L - likelihood function

Algorithm of construction a hierarchy

begin

- **Input**: training set S consisting of $(\mathbf{x}, y)$ pairs, $\mathbf{x} \in \mathcal{X}$ and $y \in \mathcal{Y}$, number of classes $= N$
- **Function** KMEANS$(\vec{s}_1, \ldots, \vec{s}_K)$
  - **for** $k \leftarrow$ *1 to K* **do**
    - $\vec{\mu}_k \leftarrow \vec{s}_k$
  - **while** *stopping criterion will not been met* **do**
    - **for** $k \leftarrow$ *1 to K* **do**
      - $x_k \leftarrow \{\}$
    - **for** $n \leftarrow$ *1 to N* **do**
      - j $\leftarrow$ *argmin*$_{j'} \|\vec{\mu}_{j'} - \vec{w}_n\|$
      - $x_j \leftarrow x_j \cup \{\vec{w}_n\}$ (reassignment of vectors)
    - **for** $k \leftarrow$ *1 to K* **do**
      - $\vec{\mu}_k \leftarrow \frac{1}{|x_k|} \sum_{\vec{w} \in x_k} \vec{w}$ (recomputation of centroids)
  - **return** $\{\vec{\mu}_1, \ldots, \vec{\mu}_K\}$
- **Function** HIERARCHY$(\{\vec{w}_1, \ldots, \vec{w}_N\}, N,$ *cl_criterion* $=$ *" square, cubic, AIC, BIC"* $)$
  - $K \leftarrow (N,$ *cl_criterion*$)$
  - $(\vec{s}_1, \ldots, \vec{s}_K) \leftarrow$ SELECT RANDOM SEEDS $(\{\vec{w}_1, \ldots, \vec{w}_N\}, K)$
  - **while** *K >= 1* **do**
    - $\{\vec{\mu}_1, \ldots, \vec{\mu}_K\} \leftarrow$ KMEANS$(\vec{s}_1, \ldots, \vec{s}_K)$
    - $K \leftarrow (K,$ *cl_criterion*$)$
    - $(\vec{s}_1, \ldots, \vec{s}_K) \leftarrow$ SELECT RANDOM SEEDS $(\{\vec{\mu}_1, \ldots, \vec{\mu}_K\}, K)$
  - **return** *Hierarchy of classes*
- $(\{\vec{w}_1, \ldots, \vec{w}_N\}, N) \leftarrow$ Train N different binary classifiers
- Hierarchy of classes $\leftarrow$ HIERARCHY$(\{\vec{w}_1, \ldots, \vec{w}_N\}, N,$ cl_criterion$)$

Results

- The experimental datasets - subsets of LSHTC training datasets. [1] **HK** - Hierarchical classification with structure that was built using K-means.

TABLE : Characteristics of the datasets used in the experiments

	number of instances			number of attributes
	total in dataset	training	testing	
DMOZ_500	7,343	5,818	1,525	68,268
DMOZ_1000	14,028	11,123	2,905	104,768
Wikipedia_100	2,207	1,815	392	346,293
Wikipedia_500	13,788	11,570	2,218	346,231

TABLE : The performance of the hierarchical and flat classifications (SVM).

SVM	**DMOZ_500**	**DMOZ_1000**	**Wikipedia_100**	**Wikipedia_500**
Flat	82.5	77.1	88.8	78.1
HK (square root)	79.3	74.0	84.9	74.0
HK (cubic root)	81.6	75.8	87.8	77.8
HK (AIC)	81.9	75.3	87.8	77.3
HK (BIC)	79.5	74.6	85.5	76.6

TABLE : Running time to predict labels of instances.

	DMOZ_500	DMOZ_1000	Wikipedia_100	Wikipedia_500
Flat - Rocchio	59 sec	1 min 53 sec	4.22 sec	2 min 24 sec
HK - Rocchio	28 sec	1 min 36 sec	1.78 sec	2 min 3 sec
Flat - SVM	49 sec	2 min 20 sec	5.24 sec	2 min 58 sec
HK - SVM	36 sec	1 min 59 sec	2.28 sec	2 min 32 sec

1. http ://lshtc.iit.demokritos.gr/

My future work

The subject :

Temporal decision trees.

THANK YOU