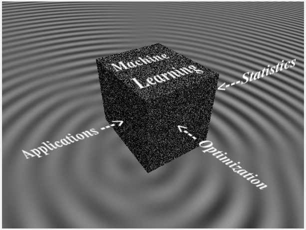
The Graphtron
An Online Algorithm for Learning a Labeling of a Graph

Kristiaan Pelckmans and Johan Suykens

ESAT - SCD/sista - KULeuven, Leuven, Belgium
CSML - UCL - London, UK

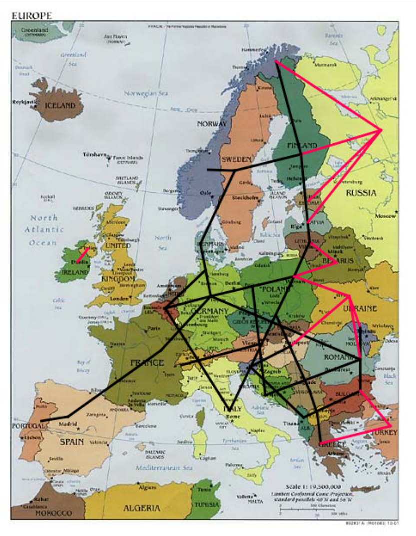
4 July 2008



I. - The Graphtron

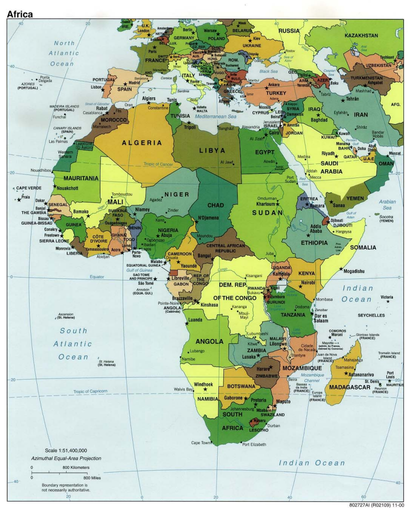
The Graphtron: Motivation

The Graphtron: Motivation



cut(Europe) = 14

The Graphtron: Motivation



cut(Africa) = 1

1. Given Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $n > 0$ nodes;
2. Given weights a_{ij} on edges e_{ij}, with $a_{ij} = a_{ji} \geq 0$ and $a_{ii} = 0$;
3. Given Laplacian
$$\mathbf{L} = \mathbf{A} - \mathbf{D}$$
and $\mathbf{D} = \mathrm{diag}(\mathbf{A}1_n)$ (degree matrix);
4. To each node v_i, a fixed but unknown label $y_i \in \{-1, 1\}$ is associated;

The Graphtron: Formal Setting

We adopt the assumption that the labeling to be learned is in some sense simple, quantified through the notion of GRAPHCUT induced by a labeling $y \in \{-1, 1\}^n$, or

$$\text{cut}(y) = \sum_{y_i \neq y_j} a_{ij} = \frac{1}{4} \sum_{i,j=1}^{n} a_{ij}(y_i - y_j)^2 = \frac{1}{4} y^T \mathbf{L} y. \quad (1)$$

The Graphtron: Formal Setting

The learning setting is as follows

1. An adversary (possibly non-random) asks for the label of a new node v_t;
2. Our algorithm makes a prediction
$$f_M(v_t) \in \{-1, 1\}$$
with M representing all needed information available prior to t.
3. The true label y_t becomes available, and we incur a loss ℓ
$$\ell(f_M(v_t), y_t)$$
4. After $T > 0$ iterations, we count the number of mistakes until t, or
$$m_T = \sum_{t=1}^{T} \ell(f_M(v_t), y_t)$$

The Graphtron: Algorithm

”Online learning over graphs, Mark Herbster, Massimiliano Pontil, Lisa Wainer; ICML 2005, Pages: 305 - 312 , 2005

$\rightarrow$ based on relation **L** and Kernels, margin vs. graphcut + perceptron,

But $O(n^3)$ and $\mathcal{G}$ fully given!

”On a theory of learning with similarity functions”, N. balcan, A. Blum, ICML 2006, p 73-80, 2006

The Graphtron: Algorithm

Definition (The Graphtron)

Input: initialize $M_0 = \{\}, m = 0$
repeat

1. An adversarial asks the label of node i.
2. We predict

$$f_M(v_i) = \text{sign}\left(\sum_{j \in M_m} a_{ij} y_j^*\right)$$

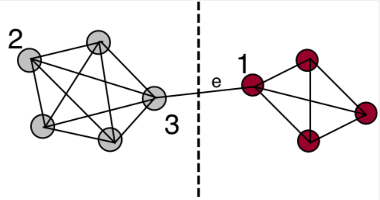
3. Nature provides the true label y_i^*

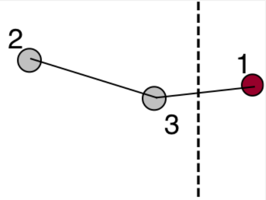
if $f_M(v_i) \neq y_i$ **then**
$M_{m+1} = M_m \cup \{i\}$ and $m = m + 1$
end if
until one is satisfied (computationally, accuracy)

III. - Some Examples

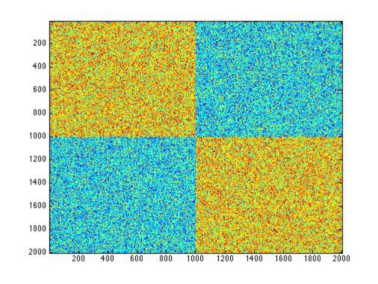
Examples

Sequence of 3 mistakes, degree of mistake graph 2.



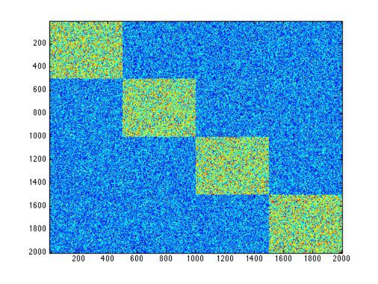


Examples (ct'd)



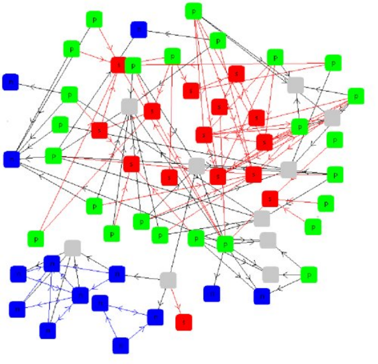
M: 1,2, 5, 7, 40, 73, 150, 262 n: 2000
surpr: 1.3888
m: 8

Examples (ct'd)



M: 1,2,3,4, 13, ...,402, 543,810,1202
n: 2000
surpr: 11.2876
m: 28

Examples (ct'd)



M: 1,2,3,4,5, ...
n: 5000
surpr: 200,9572
m: 202
prediction error 12.87%

modify adjacency matrix $A' = A * A + A$
m = 398
Prediction error 7.21%

The Graphtron

1. Simple
2. (Sub)linear in Time and Space Complexity
3. No need to know $\mathcal{G}$ in advance;
4. Stochastic Graphs $a_{ij} = s(X_i, X_j)$ with $X \sim P$;

III. - Some Analysis

Definition (Mistake Subgraph)

Let $M = M_M$ contain the indices of nodes where the algorithm incurs a mistake. Then the mistake subgraph $\mathcal{G}_M$ is the subgraph of $\mathcal{G}$ which only contains the nodes in M, and the present edges between them. Furthermore, let d_M be the degrees of the subgraph spanned by the nodes in M, or $d_{M,i} = \sum_{j \in M} a_{ij}$.

Lemma (Mistake Bound)

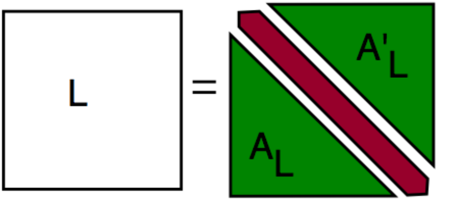
Let y^* be the true labeling. The above algorithm will incur at most $|M|$ mistakes where

$$\sum_{i \in M} d_{M,i} \leq 4\,\mathrm{cut}(y^*),$$

where $\sum_{i \in M} d_{M,i}$ equals twice the weight of all edges in the mistake graph $\mathcal{G}_M$.

Analysis: Proof

1. Re-index such that all mistakes first;
2. $y^T \mathbf{L} y \geq y_M^T \mathbf{L}_M y_M$
3. Decomposition



4. If $t > m$ (no mistake), then $y_t(A_L y_M) > 0$
5. Thus

$$4\,\mathrm{cut}(y) \geq y_M^T \mathbf{L}_M y_M \geq 2 degree(\mathcal{G}_M)$$

Graph Structure

"But what is the relation between 'degree($\mathcal{G}_M$)' and the number of mistakes m?"

1. If $\text{cut}(y) = 0$, tight upperbound

$$m \leq \alpha(\mathcal{G}) \leq \theta(\mathcal{G}) \leq \vartheta(\mathcal{G})$$

2. If nothing known on $\mathcal{G}$ apriori, implicit bound!
3. If the graph structure is known one can make the bound explicit

Conclusions

- ! The Graphtron as a paradigm for online learning without explicit representation;
- ! (sub) Linear Complexity
- ! Stochastic Graphs
- ! Multiclass Case
- ? Relation Mistake Subgraphs and graph Topology
- ? towards Optimal Teacher

  = *"Informed adversary making algorithm learn the labeling as fast as possible".*