Siminole kick-off meeting – Task 4

Simulation-based discriminative learning

Balázs Kégl

Linear Accelerator Laboratory, CNRS/University of Paris Sud

Orsay, September 27, 2010

SimInOLe

AGENCE NATIONALE DE LA RECHERCHE ANR

Discriminative learning

- Simulate $D = \{\mathbf{x}_i, y_i\}_{i=1}^n$
  - $\mathbf{x} \in \mathbb{R}^d$: observation vector
  - $y \in \{1, \ldots, K\}$: label
- Examples
  - $\mathbf{x}$: tank observables, y: number of muons
  - $\mathbf{x}$: images or films on the focal plane, y: shower / background
  - $\mathbf{x}$: ATLAS observables, y: Higgs / background

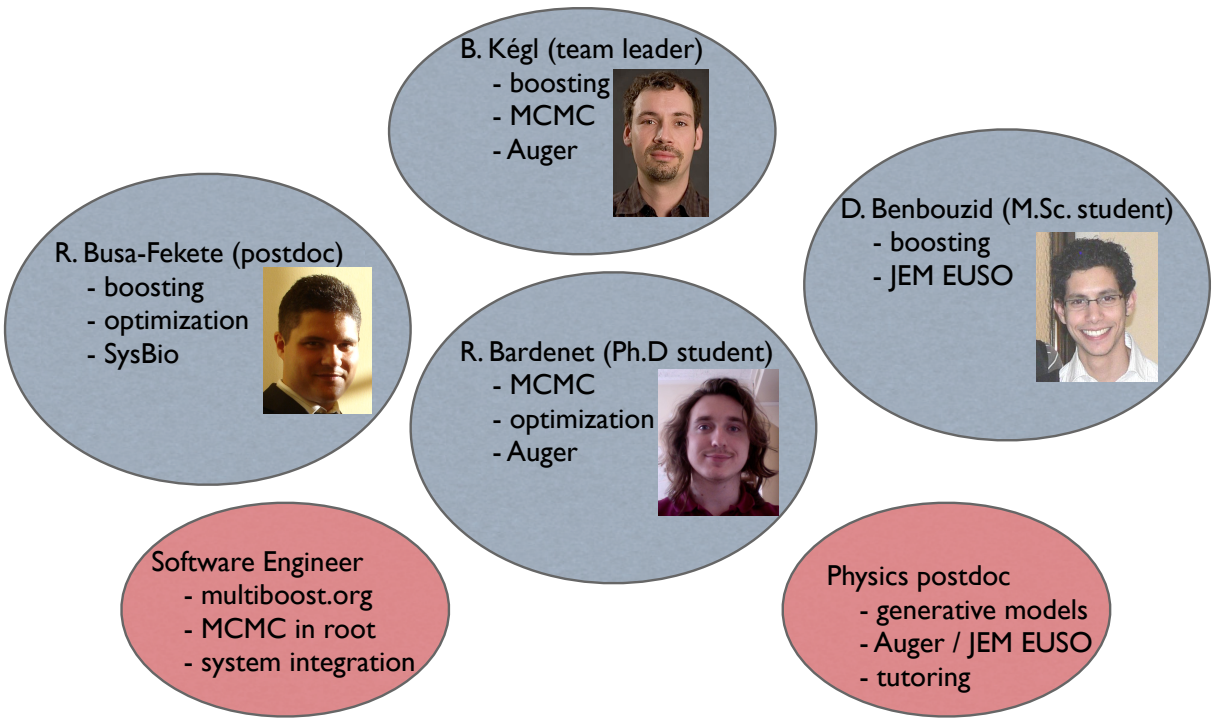
Discriminative learning

- Goal: learn $\widehat{y} = f(\mathbf{x})$ based on $D = \{\mathbf{x}_i, y_i\}_{i=1}^n$
  - minimize $\mathbb{P}\{f(\mathbf{x}) \neq y | D\}$
  - off-line analysis (reconstruction), on-line trigger, hypothesis testing
- Why AdaBoost?
  - we have the expertise
  - scales well for large problems
  - versatility (multiclass, non-differentiable weak learners)

Discriminative learning

- Research goals
  - large-scale issues (speed, memory, automatic hyperparameter optimization)
  - beyond classification (ranking, regression, conditional random fields, reinforcement learning)
  - unbalanced data: cascade architecture, Neyman-Pearson learning
- Deliverable
  - modular AdaBoost software (`multiboost.org`)

The AppStat team at LAL



- + M. Sebag and C. Germain (TAO/LRI)
- Formal ties to the TAO, Auger and JEM-EUSO teams