Cost-Sensitive Prediction: Applications in Healthcare

Daniel P. Robinson
Johns Hopkins University
Department of Applied Mathematics and Statistics

Collaborator: Suchi Saria (Computer Science)
Funding: IDIES Seed Funding Program

IDIES Annual Symposium
October 16, 2015

Outline

1. Cost-Sensitive Prediction: Applications in Healthcare
   - The application: prediction of sepsis
   - An exact cost-sensitive modeling formulation of the problem
   - Convex relaxations
   - Nonconvex relaxations and new optimization algorithms

Outline

1. Cost-Sensitive Prediction: Applications in Healthcare
   - The application: prediction of sepsis
   - An exact cost-sensitive modeling formulation of the problem
   - Convex relaxations
   - Nonconvex relaxations and new optimization algorithms

Outline

1 Cost-Sensitive Prediction: Applications in Healthcare
- The application: prediction of sepsis
- An exact cost-sensitive modeling formulation of the problem
- Convex relaxations
- Nonconvex relaxations and new optimization algorithms

Definition (sepsis)

Sepsis occurs when chemicals released into the bloodstream to fight an infection trigger inflammatory responses throughout the body.

Definition (septic shock)

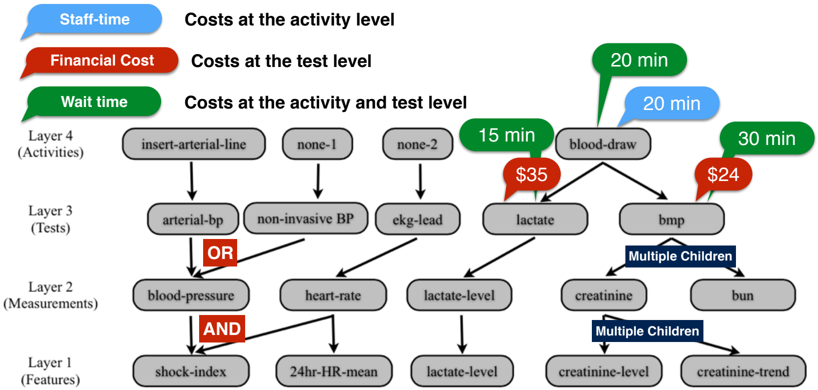
Widespread infection causing organ failure & dangerously low blood pressure

Facts:

- $200{,}000 - 3{,}000{,}000$ cases of sepsis in the USA each year.
- Sepsis is estimated to cost American hospitals $\$20$ billion each year.
- Most common and dangerous in older adults or those with weakened immune systems.
- Early treatment of sepsis with antibiotics and intravenous fluids improves chance of survival.
- 40% of the patients diagnosed with sepsis do not survive.
- Early detection is key to improve survival rates.

The symptoms are not always straightforward or agreed upon.

Cost structure graph



Research goals:

- prediction of septic shock
- model should capture the complex structure of the relevant costs
- evaluate convex/nonconvex relaxations of the exact formulation
- develop new algorithms to solve the resulting optimization problem
- evaluate performance on real healthcare data
- extensions to real-time setting, perhaps to aid in personalized healthcare

Outline

1. Cost-Sensitive Prediction: Applications in Healthcare
   - The application: prediction of sepsis
   - An exact cost-sensitive modeling formulation of the problem
   - Convex relaxations
   - Nonconvex relaxations and new optimization algorithms

Model prediction: Solve the problem

$$\min_{\beta} \ \mathcal{F}(\beta)$$

- $\beta \in \mathbf{R}^{|f|}$ is the weight vector
- For example, the logistic loss is

$$\mathcal{F}(\beta) = \sum_{n=1}^{N} \log\left(1 + \exp\left(-y^n \beta^T f^n\right)\right)$$

  - $f^n \in \mathbf{R}^{|f|}$ is the n^{th} observed feature vector
  - $y^n \in \{-1, 1\}$ the label of f^n
  - $|f|$ the number of features
  - N the number of observations

To balance predictive accuracy and model complexity:

$$\min_{\beta} \; \mathcal{F}(\beta) + \lambda C(\beta)$$

- λ is a weighting parameter
- Complexity/Cost function $C(\beta)$, e.g.,
  - $C(\beta) = ||\beta||_2^2$
  - $C(\beta) = ||\beta||_1$
  - $C(\beta) = \sum_{i \in G_j} ||\beta_{G_j}||_2$

Key: none of the above cost regularizers accurately model the relevant costs

- first nonzero feature to enter is most predictive
- the most predictive may be the most expensive
- no control over the relevant costs
- a regularizer based on the complicated cost structure is needed!

A cost-sensitive regularizer

- $\mathcal{A}$ nodes at activity layer
- $\mathcal{T}$ nodes at test layer
- $\mathcal{M}$ nodes at measurement layer
- $\mathcal{F}$ nodes at feature layer

Reduce 4 layer boolean circuit (cost graph) to a 3-layer boolean circuit:

- layer-1 contains the nodes $\mathcal{F}$
- layer-2 contains the nodes $\mathcal{Z} := \{f_{i,p} : 1 \leq i \leq n_f$ and $1 \leq p \leq w_i\}$
- layer-3 contain the nodes $\mathcal{A}$
- only **or** gate functions are used in layer-1
- only **and** gate functions are used in layer-2

A cost-sensitive regularizer

We now have that the cost for the lth care-giver activity is given by

$$\mathcal{C}_l^a \, I\left(\bigvee_{(i,p)\in\mathcal{S}_l^a} \beta_{i,p}\right)$$

where the index set $\mathcal{S}_l^a$ is defined as

$$\mathcal{S}_l^a := \{(i,p) : \text{ the output of } g_{f_{i,p}}(\cdot) \text{ depends on } a_l\}$$

Comments:

- The definition of our regularizer follows only from the knowledge of the 3-layer (reduced) boolean circuit
- This procedure generalizes to any cost graph that can be represented as a finite layer boolean circuit

We then have the following optimization problem based on the exact regularizer:

$$\min_{\beta} \ \mathcal{F}(\beta) + \lambda C_{exact}(\beta)$$

- λ is a weighting parameter
- Complexity/Cost function $C(\beta)$:

$$C_{exact}(\beta) := \sum_{l=1}^{|\mathcal{A}|} \mathcal{C}_l^a \, I\left(\bigvee_{(i,p)\in\mathcal{S}_l^a} \beta_{i,p} \right) + \sum_{l=1}^{|\mathcal{T}|} \mathcal{C}_l^t \, I\left(\bigvee_{(i,p)\in\mathcal{S}_l^t} \beta_{i,p} \right)$$

Comments:

- This cost-regularizer is exact, but not tractable for large-scale problems.
- Convex and nonconvex relaxations are possible.

Outline

1. Cost-Sensitive Prediction: Applications in Healthcare
   - The application: prediction of sepsis
   - An exact cost-sensitive modeling formulation of the problem
   - Convex relaxations
   - Nonconvex relaxations and new optimization algorithms

One could consider convex relaxations:

$$\min_{\beta} \; \mathcal{F}(\beta) + \lambda C(\beta)$$

where

$$C(\beta) := \sum_{l=1}^{|\mathcal{A}|} \mathcal{C}_l^a \, \| \bigvee_{(i,p)\in\mathcal{S}_l^a} \beta_{i,p} \|_\infty + \sum_{l=1}^{|\mathcal{T}|} \mathcal{C}_l^t \, \| \bigvee_{(i,p)\in\mathcal{S}_l^t} \beta_{i,p} \|_\infty$$

Comments:

- Sum of (overlapping) group regularizers.
- Can use software such as SPAMS to solve this problem.
- Tried this approach on the following data ...

Setup:

- Used MIMIC-II, a large publicly available dataset containing electronic health records from patients admitted to the ICU at the Beth Israel Deaconess Medical Center over a period of seven years.
- Processed the data from all adults (age > 15) with at least one measurement of blood urea nitrogen, hematocrit, and heart rate.
- This yielded data from $16{,}232$ patients.
- Septic shock onset was identified using the 2012 Surviving Sepsis Campaign definitions, which resulted in $2{,}291$ patients.
- Patients with severe sepsis who never developed septic shock but received treatment, were removed since their outcome was confounded.
- Split individuals into training (75%) and test (25%) sets.
- Since the dataset is highly unbalanced, during training, we subsample the negative pairs to generate a balanced training set.

Table : Various costs for different models obtained by using our structured regularizer.

Models	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
Sensitivity at $\mathbf{0.85}$ specificity	0.61	0.66	0.65	0.72
AUC	$\mathbf{82.79 \pm 0.55}$	$\mathbf{84.45 \pm 0.64}$	$\mathbf{84.75 \pm 0.55}$	$\mathbf{87.21 \pm 0.46}$
Financial Cost	**$0**	**$0**	**$72**	**$170**
Care-giver Time	0 minutes	**10** minutes	0 minutes	**30** minutes
Result Time	0 minutes	**10** minutes	50 minutes	**50** minutes
Tests Required	routine	routine, urine	abg, routine	abg, cbc, cmp, hct, hemoglobin, routine, urine
Activities Required	none	urine	arterial stick	arterial stick, blood draw, urine

Comments:

- Diverse models are easily obtained using this cost-sensitive regularizer.
- Easily can adjust model preferences.

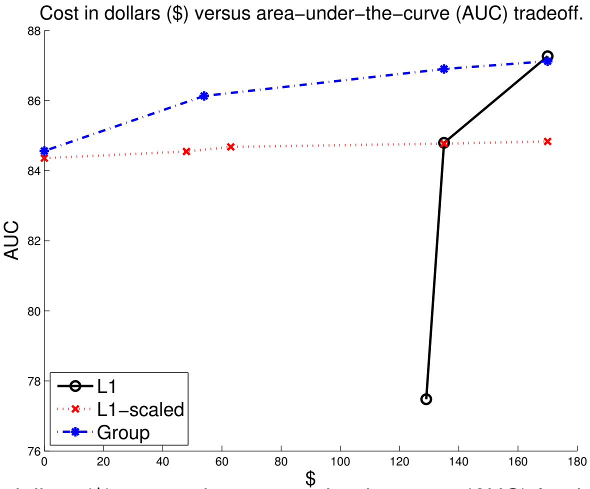


Figure : Cost in dollars ($) versus the area under the curve (AUC) for the ℓ_1 (L1), the scaled-ℓ_1 (L1-scaled), and group regularizer (Group).

Incorporating End-User Preferences in Predictive Models Via Structured Regularizers, submitted to AAAI, 2015.

Outline

1. Cost-Sensitive Prediction: Applications in Healthcare
   - The application: prediction of sepsis
   - An exact cost-sensitive modeling formulation of the problem
   - Convex relaxations
   - Nonconvex relaxations and new optimization algorithms

Solve the problem

$$\min_{\beta} \ \mathcal{F}(\beta) + \lambda C(\beta)$$

for choices such as

$$C(\beta) := \sum_{l=1}^{|\mathcal{A}|} \mathcal{C}_l^a \sqrt{\sum_{(i,p)\in\mathcal{S}_l^a} |\beta_{i,p}|} + \sum_{l=1}^{|\mathcal{T}|} \mathcal{C}_l^t \sqrt{\sum_{(i,p)\in\mathcal{S}_l^t} |\beta_{i,p}|}$$

Other choices are possible:

- Sigmoid functions.
- Min functions such as $\min\{0, \sum_{(i,p)\in\mathcal{S}_l} |\vec{\beta}_{i,p}|\}$.

The general problem formulation

$$\min_{\beta \in \mathbb{R}^{|f|}} \quad \mathcal{F}(\beta) + \sum_{j=1}^{|G|} g_j\Big(\sum_{i \in U_j} |\beta_i|\Big)$$

- $f : \mathbb{R}^{|f|} \to \mathbb{R}$ is convex
- $g_i : \mathbb{R} \to \mathbb{R}$ is concave and increasing on $[0, \infty)$

Comments:

- Difference of convex functions.
- Known algorithms when the gradient of $\mathcal{F}$ is cheap to compute.
- Methods become inefficient in the big data setting.
- We are developing methods that ...
  - only use stochastic or batch gradients of $\mathcal{F}$ (scalable)
  - utilize the difference of convex functions structure
  - utilize projections onto cones
  - inexact subproblem solvers (efficiency)
  - have convergence guarantees
  - details...hopefully next year!

Thank You