Generalized Linear Rule Models

Dennis Wei
Sanjeeb Dash
Tian Gao
Oktay Günlük

IBM Research
Generalized Linear Models + Rules
Generalized Linear Models + Rules

- Linear regression
- Logistic regression
- Poisson regression
- ...

- Conjunctions of conditions on individual features
- Can capture nonlinearities and interactions

- Age > 50 AND Blood pressure > 140
Generalized Linear Models + Rules

- Linear regression
- Logistic regression
- Poisson regression
- ...

- Conjunctions of conditions on individual features
  - Age > 50 AND Blood pressure > 140
- Can capture nonlinearities and interactions

Model interpretation
Problem Formulation

\[
\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \text{Loss} \left( \sum_{k \in \mathcal{K}} a_{ik} \beta_k , y_i \right) + \sum_{k \in \mathcal{K}} \lambda_k |\beta_k|.
\]

- **Performance**
  - Conjunction values
  - Coefficients

- **Complexity**
  - Conjunction complexity
Problem Formulation

\[
\min_\beta \frac{1}{n} \sum_{i=1}^{n} \text{Loss} \left( \sum_{k \in \mathcal{K}} a_{ik} \beta_k, y_i \right) + \sum_{k \in \mathcal{K}} \lambda_k |\beta_k| 
\]

**Challenge:** Set of conjunctions \( \mathcal{K} \) is exponentially large

- e.g. age alone, age AND blood pressure, age AND blood pressure AND body mass index, ...
**Problem Formulation**

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \text{Loss} \left( \sum_{k \in \mathcal{K}} a_{ik} \beta_k, y_i \right) + \sum_{k \in \mathcal{K}} \lambda_k |\beta_k|$$

**Challenge:** Set of conjunctions $\mathcal{K}$ is exponentially large

- e.g. age alone, age AND blood pressure, age AND blood pressure AND body mass index, ...

We avoid limitations of existing methods

- Pre-select (large) subset e.g. [Friedman & Popescu, 2008]
- Boost but do not revise rules e.g. [Cohen & Singer, 1999; Dembczynski et al., 2010]
Column Generation

conjunction complexities $\lambda$

conjunction matrix $A$
Column Generation

Solve over small subsets $\mathcal{S}$

Restricted GLM

conjunction matrix $A$

conjunction complexities $\lambda$

$$\min_\beta \frac{1}{n} \sum_{i=1}^{n} \text{Loss} \left( \sum_{k \in \mathcal{S}} a_{ik} \beta_k, y_i \right) + \sum_{k \in \mathcal{S}} \lambda_k |\beta_k|$$
Column Generation

Solve over small subsets $\mathcal{S}$

Restricted GLM

Conjunction matrix $A$

Conjunction complexities $\lambda$

Check optimality

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \text{Loss} \left( \sum_{k \in \mathcal{S}} a_{ik} \beta_k , y_i \right) + \sum_{k \in \mathcal{S}} \lambda_k |\beta_k|$$
Column Generation

Solve over small subsets $\mathcal{S}$

$\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \text{Loss} \left( \sum_{k \in \mathcal{S}} a_{ik} \beta_k, y_i \right) + \sum_{k \in \mathcal{S}} \lambda_k |\beta_k|$

conjunction complexities $\lambda$

conjunction matrix $A$

Restricted GLM

Add improving columns

Pricing Problem

Check optimality

Integer program (solved using CPLEX) gives optimality guarantee
Heuristic also effective and more practical
Performance-Complexity Trade-Offs

Logistic and linear regression experiments

Logistic Rule Regression with CG (LRR, LRRN) obtains better trade-offs than (RuleFit, RuleFitN) on most of 16 classification datasets.

- **Pima**
- **FICO**
- **magic**
- **musk**
Logistic/Linear Rule Regression with CG (LRR, LRRN) is highly competitive when tuned to maximize performance.

<table>
<thead>
<tr>
<th>method</th>
<th>LRR</th>
<th>LRRN</th>
<th>RuleFit</th>
<th>RuleFitN</th>
<th>GBM</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>logistic regression mean rank</td>
<td>4.1</td>
<td>3.6</td>
<td>4.8</td>
<td>3.6</td>
<td>5.3</td>
<td>4.0</td>
</tr>
<tr>
<td>linear regression mean rank</td>
<td>4.9</td>
<td>3.0</td>
<td>4.5</td>
<td>3.5</td>
<td>3.4</td>
<td>5.0</td>
</tr>
</tbody>
</table>

and uses 2-4 times fewer rules than RuleFit [Friedman & Popescu, 2008]
GLRM: Generalized Linear Models + Rules

Flexible and interpretable models

Probabilistic classification and regression

Column generation to efficiently search space of rules without restrictions

Poster #264, today 6:30-9:00 PM