Mathematical optimization for scoring modelling

Guillermo Navas-Palencia

DBRS Morningstar

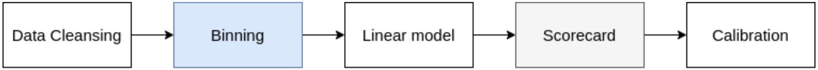
31st European Conference on Operational Research

Athens, 11 - 14 July 2021

Table of Contents

Scorecard introduction

- Scorecard modelling comprises the use of simple interpretable linear models to make predictions.
- Scorecard models are widely employed in sectors such as the financial industry or healthcare.
  - Interpretable models with a small number of variables.
  - Satisfy multiple behavioural and operational constraints.



Scorecard structure

- Use of binning techniques to discretize variables. Apply a data transformation.
- The score points are calculated using the coefficients of the linear model and the data transformation assigned to each bin and variable.
- Given a feature i with $\mathcal{B}_i$ bins, $j = 1, \ldots, \mathcal{B}_i$, and linear model coefficient c_i the corresponding score point s_{ij} is given by

  $$s_{ij} = c_i t_{ij},$$

  where t_{ij} is the data transformation.
- Given an instance x with n features, the corresponding total score is

  $$score(x) = \sum_{i=1}^{n} \sum_{j=1}^{\mathcal{B}_i} s_{ij} \mathbf{1}_{\{x_i \in I_{ij}\}}$$

  where I_{ij} is the interval corresponding to the j bin for feature i.

Scorecard example

Feature	Bin	Points
VariableA	[-∞, 59.5)	5.43
VariableA	[59.5, 63.5)	11.62
VariableA	[63.5, 65.5)	18.15
VariableA	[65.5, ∞)	25.44
VariableB	[-∞, 1.5)	0.47
VariableB	[1.5, 3.5)	3.12
VariableB	[3.5, ∞)	7.06

Table: Example of scorecard with two features.

- $n = 2$
- $\mathcal{B} = [4, 3]$
- $x = [60, 1]$

$$score(x) = 0 + 11.62 + 0 + 0 + 0.47 + 0 + 0 = 12.09$$

Types of binning algorithms

- Unsupervised binning techniques
  - Equal-width interval binning
  - Equal-size or equal-frequency interval binning

- Supervised binning techniques
  - Merging: Monotone Adjacent Pooling Algorithm
  - Decision trees: CART, Minimum Description Length Principle (MDLP), condition inference trees (CTREE)

- Fast and robust algorithms ✓
- Constraints ✗

Optimal binning

Definition: The optimal binning is the optimal discretization of a variable into bins given a discrete or continuous numeric target.

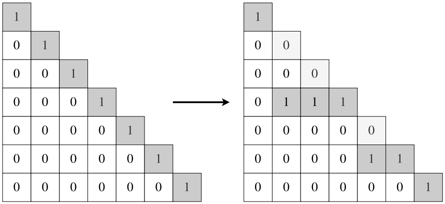
- Pre-binning
  - m split points, $s_1 < s_2 < \ldots < s_m$ to create $n = m + 1$ pre-bins. These pre-bins are defined by the intervals $(-\infty, s_1), [s_1, s_2), \ldots, [s_m, \infty)$.
- Optimization
  - Decision variable: $X_{ij} \in \{0, 1\}, \forall (i, j) \in \{1, \ldots, n : i \geq j\}$

- Each column must contain exactly one 1. For $j = 1, \ldots, n$

$$\sum_{i=1}^{n} X_{ij} = 1.$$

- Continuity by rows, no $0 - 1$ gaps are allowed. For $i = 1, \ldots, n;\ j = 1, \ldots, i - 1$

$$X_{ij} \leq X_{ij+1}.$$



Optimal binning: MIP formulation

Objective function with binary and multi-class target, maximize a divergence measure (Jeffrey's divergence). Given non-event ($y = 0$) and event ($y = 1$)

$$J(P||Q) = IV = \sum_{i=1}^{n} (p_i - q_i) \log \left(\frac{p_i}{q_i} \right).$$

$$V_{ij} = \left(\sum_{z=j}^{i} \frac{r_z^{NE}}{r_T^{NE}} - \frac{r_z^{E}}{r_T^{E}} \right) \log \left(\frac{\sum_{z=j}^{i} r_z^{NE} / r_T^{NE}}{\sum_{z=j}^{i} r_z^{E} / r_T^{E}} \right), \quad i = 1, \ldots, n; \; j = 1, \ldots, i.$$

$$V_{i\cdot} = V_{ii} X_{ii} + \sum_{j=1}^{i-1} (V_{ij} - V_{ij+1}) X_{ij}.$$

Basic MIP formulation:

$$\max_{X} \quad \sum_{i=1}^{n} V_{ii} X_{ii} + \sum_{j=1}^{i-1} (V_{ij} - V_{ij+1}) X_{ij} \tag{1a}$$

$$\text{s.t.} \quad \sum_{i=j}^{n} X_{ij} = 1, \qquad j = 1, \ldots, n \tag{1b}$$

$$X_{ij} - X_{ij+1} \leq 0, \qquad i = 1, \ldots, n; \; j = 1, \ldots, i-1 \tag{1c}$$

$$X_{ij} \in \{0, 1\}, \qquad \forall (i, j) \in \{1, \ldots, n : i \geq j\} \tag{1d}$$

Optimal binning: monotonic event rate constraint

For $i = 1, \ldots, n;\; j = 1, \ldots, i$

$$D_{ij} = \frac{\sum_{z=j}^{i} r_z^E}{\sum_{z=j}^{i} r_z}.$$

For $i = 1, \ldots, n$.

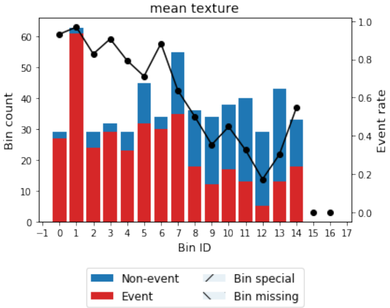
$$D_{i\cdot} = D_{ii}X_{ii} + \sum_{j=1}^{i-1}(D_{ij} - D_{ij+1})X_{ij}.$$

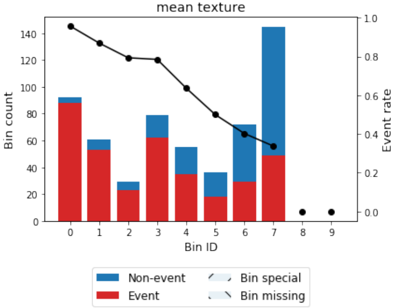
For $i = 2, \ldots, n;\; z = 1, \ldots i-1$

$$D_{ii}X_{ii} + \sum_{j=1}^{i-1}(D_{ij} - D_{ij+1})X_{ij} + \beta(X_{ii} + X_{zz} - 1)$$

$$\leq 1 + (D_{zz} - 1)X_{zz} + \sum_{j=1}^{z-1}(D_{zj} - D_{zj+1})X_{zj},$$

Auto monotonic trend using Machine Learning algorithm to determine the trend maximizing divergence.





Optimal binning: additional constraints

Reduction of dominant bins

$$\max_{X, p_{\min}, p_{\max}} \quad \sum_{i=1}^{n} V_{ii} X_{ii} + \sum_{j=1}^{i-1} (V_{ij} - V_{ij+1}) X_{ij} - \gamma (p_{\max} - p_{\min})$$

$$\text{s.t.} \quad \text{(1b - 1d)}$$

$$p_{\min} \leq r_T (1 - X_{ii}) + \sum_{j=1}^{i} r_j X_{ij}, \qquad i = 1, \ldots, n$$

$$p_{\max} \geq \sum_{j=1}^{i} r_j X_{ij}, \qquad i = 1, \ldots, n$$

$$p_{\min} \leq p_{\max}$$

$$p_{\min} \geq 0.$$

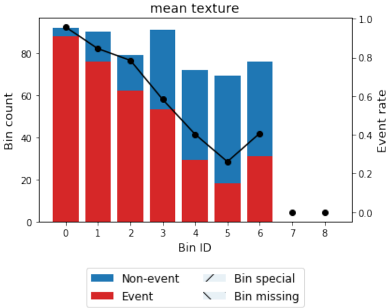
$$p_{\max} \geq 0.$$

Maximum p-value constraint

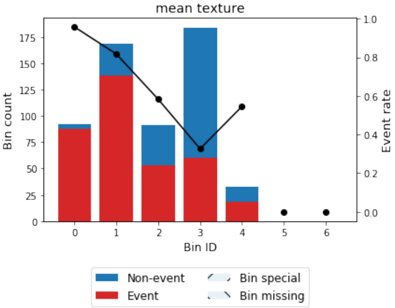
$$X_{ij} + X_{kl} \leq 1, \quad \forall (i, j, k, l) \in \mathcal{I}.$$

Optimal binning: additional constraints

Reduction of dominant bins



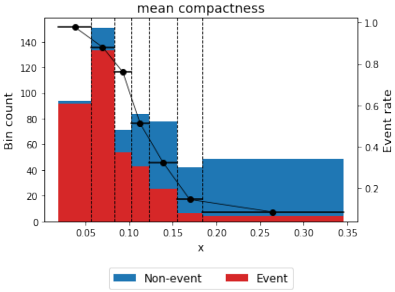
Maximum p-value constraint



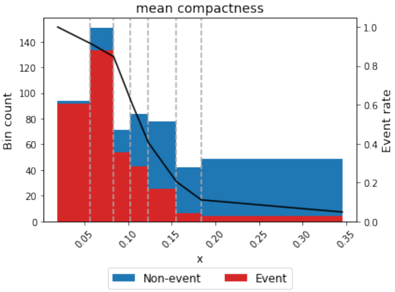
Optimal piecewise binning

- Classical binning algorithms: discontinuous piecewise constant function
- Constant value reduces the granularity of the transformation
- Discontinuities are a source of unfairness
- Good compromise between accuracy and explainability

Discontinuous piecewise constant



Continuous piecewise linear



Optimal piecewise binning: MIP formulation

Pre-binning produces mutually exclusive sets of indices satisfying

$$n = \sum_{j=1}^{B} |\mathcal{I}_j|$$

Basic MINLP formulation:

- Binary case $y = P[Y = 1|X = x]$ using ML estimator
- Polynomial degree K

$$\min \quad \sum_{j=1}^{B} \left(\sum_{i \in \mathcal{I}_j} \left| y_i - \sum_{k=1}^{K} c_{jk} x_i^{k-1} \right|^p \right)^{1/p} \tag{3a}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} c_{jk} s_j^{k-1} = \sum_{k=1}^{K} c_{j+1k} s_j^{k-1}, \qquad j = 1, \ldots, B-1 \tag{3b}$$

$$l_j \leq \sum_{k=1}^{K} c_{jk} x_i^{k-1} \leq u_j, \qquad j = 1, \ldots, B,\ i \in \mathcal{I}_j \tag{3c}$$

$$c_{jk} \in \mathbb{R}, \qquad j = 1, \ldots, B,\ k = 1, \ldots, K \tag{3d}$$

Objective with $p = \{1, 2\}$ and suitable for quantile regression.

Optimal piecewise binning: additional constraints

Ascending/descending monotonicity

$$\sum_{k=1}^{K} c_{jk}(k-1)x_j^{k-2} \geq 0, \quad j = 1, \dots, B, \ i \in \mathcal{I}_j$$

$$\sum_{k=1}^{K} c_{jk}(k-1)x_j^{k-2} \leq 0, \quad j = 1, \dots, B, \ i \in \mathcal{I}_j$$

Regularization: avoid coefficients for large K when unnecessary.

$$\min \sum_{j=1}^{B} \left(\sum_{i \in \mathcal{I}_j} \left| y_i - \sum_{k=1}^{K} c_{jk} x_i^{k-1} \right|^p \right)^{1/p} + \|c\|_{p'}$$

Implementation using SOCP solver in the RoPWR library [NP20c]

The OptBinning library

- Python library, open source (Apache-2.0)
- GitHub: `https://github.com/guillermo-navas-palencia/optbinning`
  - GitHub stars: 121
  - Downloads: > 70k
- Documentation: `http://gnpalencia.org/optbinning/`
- Support several solvers:
  - Google OR-Tools: CBC and CP-SAT
  - LocalSolver
  - Gurobi (coming soon)
- Functionalities
  - Optimal binning for binary, continuous and multi-class target [NP20b]
  - Optimal binning over data streams [NP20a]
  - Optimal binning with uncertainty
  - Scorecard development
  - Counterfactual explanations for scorecard models [NP21]

Bibliography

G. Navas-Palencia.
Optimal binning for streaming data.
http://gnpalencia.org/blog/2020/binning_data_streams/, 2020.

G. Navas-Palencia.
Optimal binning: mathematical programming formulation.
abs/2001.08025, 2020.

G. Navas-Palencia.
Robust piecewise regression.
https://github.com/guillermo-navas-palencia/ropwr, 2020.

G. Navas-Palencia.
Optimal counterfactual explanations for scorecard modelling.
abs/2104.08619, 2021.

Thank you!