Profit Maximizing Logistic Regression Modeling for Credit Scoring

Arnout Devos[1,2], Jakob Dhondt[3], Eugen Stripling[2], Bart Baesens[2,4], Seppe vanden Broucke[2], Gaurav Sukhatme[1]

[1]USC, Los Angeles, USA - [2]KU Leuven, Leuven, Belgium - [3]Switch, Zurich, Switzerland - [4]U of Southampton, Southampton, UK

1. Motivation

1. Credit scoring is an imbalanced classification problem in terms of *costs (% of credit given*) and *benefits (interest).*
2. Maximum profit = most important business requirement, so profit-based classifier evaluation metric has been developed: *Expected Maximum Profit (EMP)* [1].
3. EMP has been applied to customer churn model construction [2], but credit scoring lacks this and has regulatory restrictions (Basel II/III) on prediction models.

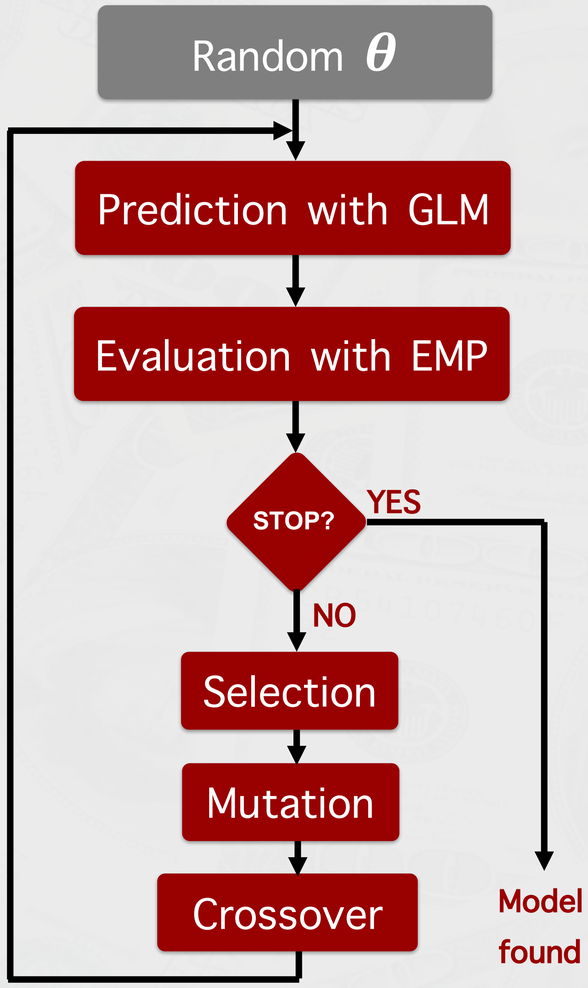
2. Idea

- Use a logistic regression model (GLM) to model applicant x_i defaulting ($y = 0$) or not ($y = 1$) + satisfy regulatory restrictions:

$$P(Y = 0|x_i) = \frac{e^{\beta_0 + \boldsymbol{\beta}^T x_i}}{1 + e^{\beta_0 + \boldsymbol{\beta}^T x_i}}$$

- Optimize parameters $\boldsymbol{\theta} = (\beta_0, \boldsymbol{\beta})$ using a Genetic Algorithm (GA) with as fitness function the parametrized Expected Maximum Profit (EMP).

GA EMP System Overview



PROFESSIONAL CONTRIBUTION

Classifier is more profitable than usual maximum likelihood logistic regression credit scoring models while satisfying regulatory restrictions.

3. Results

datasets

HMEQ

Model	AUC	Total Profit ($)	Extra Profit ($)
no model	0.5	1,851,022	0
GLM	**0.8077**	2,427,683	576,661
GA EMP	0.8068	**2,501,595**	**650,573**

GERM

Model	AUC	Total Profit ($)	Extra Profit ($)
no model	0.5	-53,015	0
GLM	**0.7819**	6,883	59,898
GA EMP	0.7705	**13,890**	**66,905**

4. Conclusion

- Proposed classifier optimizes for maximum profit in model construction.
- Considers both costs AND benefits, contrary to cost-only classifiers.
- Shows increased profit compared to maximum likelihood logistic regression.
- Future work: Tree-based, AUC, H measure

[1] T. Verbraken, C. Bravo, R. Weber, and B. Baesens, "Development and application of consumer credit scoring models using profit-based classification measures," European Journal of Operational Research , vol. 238, no. 2, pp. 505 513, 2014.

[2] E. Stripling, S. vanden Broucke, K. Antonio, B. Baesens, and M. Snoeck, "Profit maximizing logistic model for customer churn prediction using genetic algorithms," Swarm and Evolutionary Computation, 2017.

USC University of Southern California

KU LEUVEN