A new model averaging approach to credit risk default prediction.

Author: Paritosh Jha

Abstract

The literature suggests various approach of model averaging to reduce prediction error and therefore recommends some answer to model selection uncertainty. Despite the availability of large range of model averaging technique that exists, our effort in this paper is to introduce a new weighted model of averaging approach which is more tactical rather than bayesian or information-theoretic approach in reducing prediction error of credit default in general purely from application oriented perspective. The novel idea proposed here has the capacity to enhance better performance with compare to single best model (parametric, non-parametric and ensemble model) as evident in their empirical findings and offers the possibility for an effective credit lending process to decision makers at financial institution.

Methods

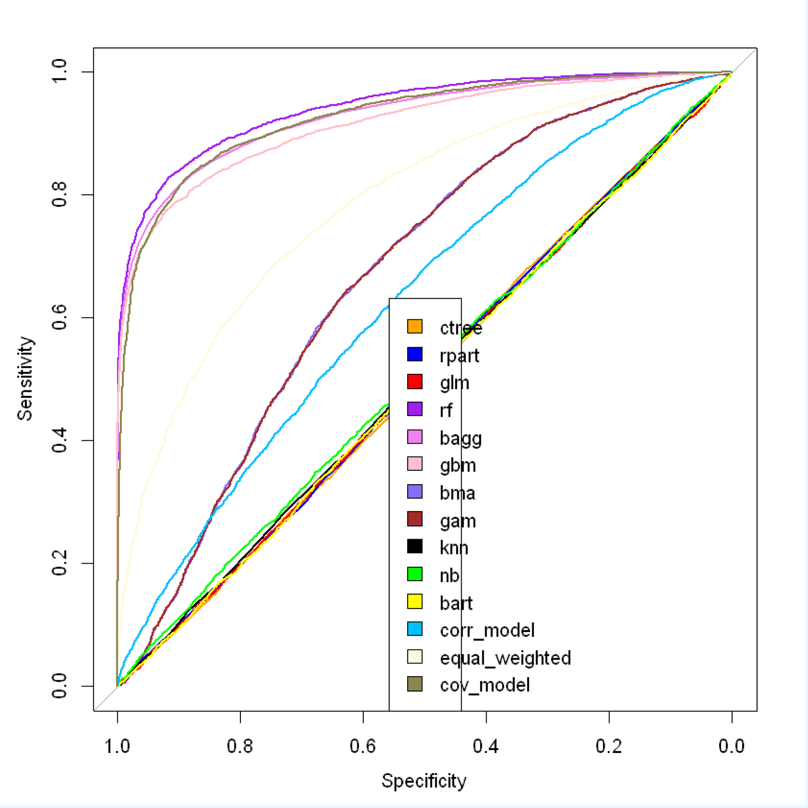
- **We formulate the optimization problem in the context of our research question with two possible extension to reduce prediction error and model selection uncertainity.**
- $\min w^T \sum w$ **such that** $w^T 1=1$, $w^T \bar{\tilde{y}}$ **, Solving analytically the optimization problem produces optimal vector of weights** $w^* = \sum^{-1} \times (1\ \bar{\tilde{y}}) \times A^{-1} \times \begin{bmatrix} 1 \\ \bar{y} \end{bmatrix}$ **where** $A = (1\ \bar{\tilde{y}})^T \times \sum^{-1} \times (1\ \bar{\tilde{y}})$

Results

- *In terms of H measure, RandomForest and proposed weighted model based on co-variance approach shows superior performance with compare to other models.*
- *With respect to AUC and AUCH values, RandomForest leads the performance and is very close to bagging and boosting, so, they are highly comparable to our proposed weighted model.*
- *Looking at the error metrics, best performing models are RandomForest, bagging, boosting and proposed weighted model.*

metrics	ctree		rpart	glm	rf	bagg	boost	bma	gam	knn	nb	bart	cor_m	ew_m	Cov_m
H	0.00		0.00	0.00	0.65	0.61	0.57	0.12	0.12	0.00	0.00	0.00	0.06	0.27	0.60
AUC	0.50		0.50	0.50	0.94	0.93	0.92	0.67	0.67	0.50	0.51	0.50	0.63	0.79	0.93
AUCH	0.51		0.51	0.50	0.94	0.93	0.92	0.68	0.68	0.51	0.52	0.51	0.63	0.79	0.93

metrics	ctree		rpart	glm	rf	bagg	boost	bma	gam	knn	nb	bart	cor_m	ew_m	Cov_m
MER	0.46		0.46	0.46	0.13	0.14	0.15	0.35	0.35	0.46	0.46	0.46	0.40	0.28	0.14
MWL	0.48		0.49	0.49	0.12	0.14	0.15	0.36	0.36	0.49	0.48	0.49	0.40	0.28	0.14



Discussion

- The main advantage of our proposed methodology is improvements in terms of performances compared to single models, closed form solution for the optimization and simple interpretation of the whole theoretical structure.

- There are few disadvantage to this approach of which one is as how to interpret negative weights. If the predicted value are bounded, the proposed approach does not guarantee that it will respect the bounds.

Conclusions

- ✓ The proposal advocates a new direction of model average to improve predictive performance.
- ✓ Benefits are earned when we are able to decrease co-variance and mean bias of contributing models.
- ✓ Non-parametric methods gain competitive advantage over parametric method with this proposed approach for model performance under uncertainity..