Discussion of "Financial Crisis Prediction with Machine Learning"

Modelling with Big Data and Machine Learning
26-27 November 2018

Contribution

- The paper tackles an important issue, hence the prediction model is an interesting contribution per se
- The application of the Shapley regression is also very interesting as it helps us understand what drives the prediction
- I appreciated the choice of using an outer 5-fold cross-validation (to evaluate) and an inner 10-fold one (to choose ther regularization parameters)

Comments

1. The logistic regression does particularly well in this context
   - Why not try with a logistic LASSO where you also include interactions, polynomials/splines, etc.?
2. The inner and outer cross-validations are very useful in reducing the variability of the estimates
   - However, the main purpose is to forecast future crises. Would it be possible to check what happens using data up to some point to predict more recent events?
3. Should we use the Shapley regression to highlight a single variable (e.g. GDP growth)?
   - You also discuss the relevance of interactions, but the level of complexity might be higher
   - I see this more as a way to understand whether the black box gives us something transferable to other contexts