

Прогнозирование категорий российского экспорта и импорта в условиях проклятия размерности

Научный руководитель – Фокин Никита Денисович

Майорова Ксения Николаевна

Студент (бакалавр)

Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Экономический факультет, Москва, Россия

E-mail: mayorovak99@gmail.com

В связи с возросшим объёмом доступных данных, часто приходится оценивать регрессии с большим количеством объясняющих переменных, в том числе в условиях проклятия размерности, когда используется число предикторов (объясняющих переменных), намного превосходящее число наблюдений. Поэтому в последнее время интерес представляют именно те методы, которые могут быть использованы в таких условиях, так как классические эконометрические методы уже не могут быть применены.

Цель данной работы состояла в том, чтобы построить прогноз российского экспорта по товарным группам в условиях большого числа объясняющих переменных, опробовав различные модели для этого. В работе приведена методология выбранных моделей - Lasso, Ridge и ElasticNet (в качестве бенчмарка использовалась ARIMA). Данные модели используют методы регуляризации (L1-, L2-regularization) для решения проблемы проклятия размерности. Принцип метода регуляризации состоит в том, что вводится ограничение на параметры, которое штрафует коэффициенты за их избыточный рост, приводит к полному обнулению коэффициентов при переменных, объясняющая сила которых несущественна. Тем самым, это помогает снизить сложность модели, и, как следствие, снизить дисперсию оценки, её чувствительность к малым отклонениям.

Так же в работе были опробованы методы машинного обучения, которые позиционируют себя как устойчивые к проблеме проклятия размерности, а именно Random Forest и XGBoost, практика применения которых не так велика для прогнозирования российских макропоказателей.

Для проведения исследования были собраны квартальные данные с первого квартала 2005 года по второй квартал 2019 года по 96 категориям российского экспорта и импорта и по 18 биржевым ценам на экспортируемое из РФ сырьё. Для подбора наилучших гиперпараметров для каждой из модели машинного обучения была применена скользящая кросс-валидация, так как данные имеют тип временных рядов. Модели с выбранными наилучшими значениями гиперпараметров тестировались на конкретном наборе данных (тестовом) и сравнивались между собой и с бенчмарком по метрике ошибки прогноза (а нашем случае - RMSE).

Полученные сравнительные результаты показали, что методы машинного обучения хорошо показывают себя в условиях проклятия размерности и дают более точный прогноз в сравнении с бенчмарком, то есть с простейшей моделью.

Источники и литература

- 1) Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media, 2009.
- 2) Sims, Christopher A. "Macroeconomics and reality." *Econometrica: journal of the Econometric Society* (1980): 1-48.

- 3) Tibshirani, Robert. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (1996): 267-288.
- 4) Zou, Hui, and Trevor Hastie. "Regularization and variable selection via the elastic net." *Journal of the royal statistical society: series B (statistical methodology)* 67.2 (2005): 301-320.