



М. А. Широбокова

## ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ПРИ ПОСТРОЕНИИ СКОРИНГОВОЙ СИСТЕМЫ ОЦЕНКИ БАНКОВСКОГО КРЕДИТНОГО РИСКА

(Удмуртский государственный университет)

В условиях повышения уровня неопределенности на рынке, а также в соответствии с Базель II [1,2] управление кредитным риском становится одним из приоритетных направлений в банковской сфере. Одним из наиболее популярных подходов в оценке кредитоспособности заемщиков выступает построение индивидуальной скоринговой системы. Скоринговая система представляет собой систему принятия решения, базой которой является математическая модель присвоения рейтинга заемщикам на основе ключевых характеристик клиента. [7] Если ранее построение скоринговой системы основывалось преимущественно на экспертном подходе, то сейчас большие обороты набирает построение индивидуальной скоринговой системы на основе внутренних данных отдельных банков. Такой подход позволяет провести анализ кредитного портфеля и оценить его риски в соответствии с особенностями деятельности каждого банка в отдельности.

Вопрос выбора метода построения индивидуальной скоринговой системы остается открытым. В условиях общего уровня развития информационной среды с каждым годом происходит увеличение объема информации, улучшаются существующие возможности ее хранения, дорабатываются и появляются новые методы быстрой и качественной обработки данных. Информационные системы банковского сектора также набирают обороты: на основе собранной информации банки начинают использовать все более продвинутые методы анализа, в том числе методы машинного обучения, которые ранее немислимо было применить в финансовой сфере. Так потенциально возможными для использования методами построения моделей могут выступать логистическая регрессия, дерево решений, метод случайного леса (Random Forest), метод опорных векторов, нейронная сеть, бустинг.

Выбор метода построения модели одновременно зависит от исходных данных, которые зачастую не имеют явно выраженной зависимости (на разных участках данных лучше может работать тот или иной метод), их влияния на целевую функцию и поставленной задачи перед бизнесом. Если данные описываются некоторой линейной зависимостью (рис. 1а), то линейная модель или метод опорных векторов будут более качественными по сравнению с деревом (деревами) решений. Если же данные невозможно описать в виде некоторой линейной зависимости в явном виде (рис. 1б), то возникает обратная ситуация, и в данном случае дерево решений будет предоставлять более качественное решение. [6]

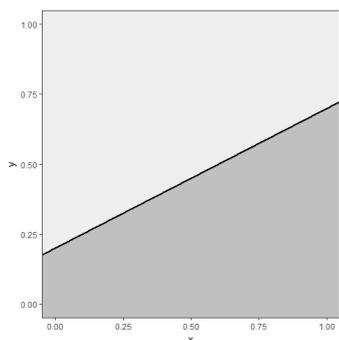


Рис. 1а

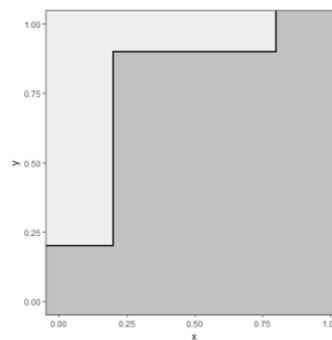


Рис. 1б

На основе имеющихся данных по договорам и дефолтам построим скоринговую модель следующими методами: с помощью метода логистической регрессии, дерева решений, метода «случайного леса», метода опорных векторов. Результирующим вектором (целевой функцией) является вероятность того, что клиент выйдет в дефолт.

Для логистической регрессии предсказанные вероятности основываются на вычислении логарифмов отношения шансов «хорошего» (не дефолтного) и «плохого» (дефолтного) кредитов на основе обучающей выборки. Формула расчета вероятности того, что клиент выйдет в дефолт, следующая:

$$p = \frac{1}{1 + e^{-z}}, \quad (1)$$

$$\text{где } z = b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n + a, \quad (2)$$

где  $x_1, x_2, \dots, x_n$  — значения атрибутов значимых характеристик,

$b_1, b_2, \dots, b_n$  — коэффициенты модели,

$a$  — некоторая константа.

Дерево решений представляет собой совокупность правил вида «если..., то...» в иерархической структуре, где каждому объекту соответствует единственный узел, дающий решение. [8]

Метод «случайного леса» на основе построения большого числа деревьев решений, каждое из которых строится не по всей, а по случайной выборке исходных данных, выбирает усредненный результат. Т.е. общая оценка регрессии — это усреднение оценок регрессии всех деревьев решений. [6]

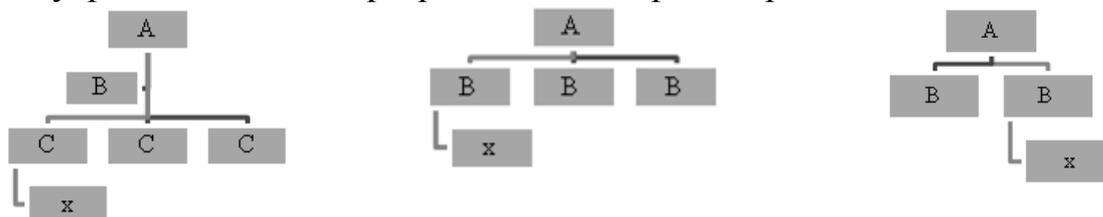


Рис. 3

Метод опорных векторов позволяет разделить данные на гиперплоскости таким образом, чтобы условно называемые все точки одного класса лежали по одну сторону от этой прямой, а все точки другого класса были на противоположной стороне. Для примера с двумя параметрами выглядит следующим образом:

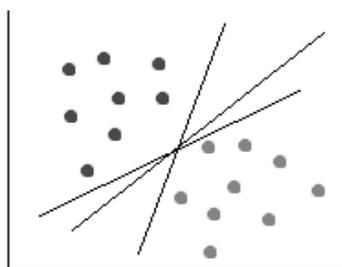


Рис. 4

Построив модели описанными методами, сравним их качество с помощью построения ROC-кривой и расчета коэффициента Gini для тестовой выборки.

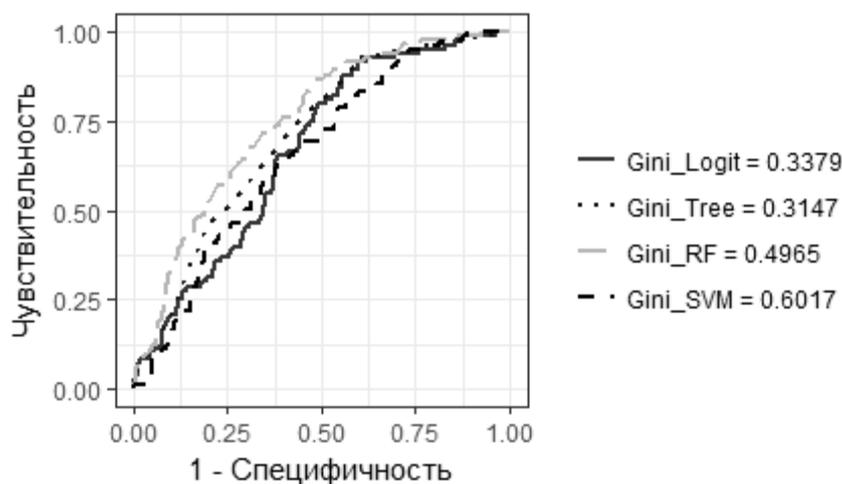


Рис. 4

Получаем, что для используемых данных по договорам и дефолтам метод опорных векторов является наиболее оптимальным, что обусловлено их структурой. В то время как базовые методы построения скоринговой модели — логистическая регрессия и дерево решений — дают результат хуже, отметим, что при использовании иных данных ситуация может быть и обратной. Стоит также обратить внимание, что с одной стороны более сложные методы построения модели позволяют увеличить точность предсказания скоринговой системы и за счет чего повысить экономическую эффективность банка. С другой стороны нельзя забывать о трудностях практической реализации модели, увеличению времени обработки информации и повышению сложности интерпретируемости результатов модели, которая становится «черным ящиком» при использовании все более сложного метода построения.

### Литература

1. Алескерев, Ф. Т. Анализ математических моделей Базель II / Ф. Т. Алескерев, И. К. Андриевская, Г. И. Пеникас, В. М. Солодков — 2-е изд., испр. — М.: ФИЗМАТЛИТ. 2013. — 296 с.
2. Международная конвергенция измерения капитала и стандартов капитала: Уточненные рамочные подходы / Базельский комитет по банковскому надзору. — Банк международных расчетов. 2004. — 266 с.



3. Джеймс Г., Уиттон Д., Хастис Т., Тибширани Р. Введение в статистическое обучение с примерами на языке R (пер. с англ. С. Э. Мастицкого). — М.: ДМК Пресс, 2016. — 460 с.
4. Чистяков С. П. Случайные леса: обзор / С. П. Чистяков. — Труды Карельского научного центра РАН, 2013. . — №1. . — С. 117 – 136.
5. Широбокова М. А. Построение скоринговой карты с использованием модели логистической регрессии / М. А. Широбокова ; науч. рук. А. В. Лётчиков. — Итоговая студенческая научная конференция (44; Апрель, 2016) XLIV итоговая студенческая научная конференция: материалы конф., Ижевск: [Удмуртский университет]. 2016. — С. 97-99.
6. Широбокова, М. А. Совмещение методов построения скоринговой модели: логистическая регрессия и метод «случайного леса» / М. А. Широбокова — Математические методы и интеллектуальные системы в экономике и образовании: Материалы Всероссийской заочной научно-практической конференции / УдГУ/ под. ред. А. В. Лётчикова. — Ижевск, 2016.
7. Siddiqi N. Credit risk scorecards: developing and implementing intelligent credit scoring. Canada: John Wiley & Sons, Inc. 1969. 196 с.
8. BaseGroup Labs — Режим доступа: <http://www.basegroup.ru/> — [Электронный ресурс].