

УДК 330.43, 519.2, 519.86

Сорокин Александр Сергеевич

ФБГОУ ВПО «Московский государственный университет экономики, статистики и информатики»
Россия, Москва¹

Доцент кафедры Математической статистики и эконометрики
Московский финансово-промышленный университет «Синергия»

Россия, Москва²

Доцент кафедры Бизнес-статистики

Кандидат экономических наук

E-Mail: alsorokin@mail.ru

К вопросу валидации модели логистической регрессии в кредитном скоринге

Аннотация: Построение автоматизированных скоринговых систем позволяет банкам снизить индивидуальные кредитные риски. Использование скоринговых систем на базе статистических моделей наряду с экспертными оценками андеррайтеров является распространенной практикой. Метод логистической регрессии — наиболее часто используемый метод для построения скоринговых систем в банках.

В статье рассматриваются вопросы применения логистической регрессии для классификации заемщиков в кредитном скоринге. Важным заключительным этапом построения скоринговой системы является этап ее валидации, который заключается в проверке достоверности полученной модели на обучающей, контрольной выборке и реальных данных. Валидация скоринговой модели должна производиться на основе системы критериев качества.

В данной статье, во-первых, систематизируются методы оценки достоверности скоринговой модели. Во-вторых, проводится сравнительный анализ методов оценки эффективности ранжирования заемщиков по модели логистической регрессии, а также по скоринговой карте, построенной на основе логистической регрессии. В третьих, дается обзор стратегий валидации скоринговых моделей.

Кроме того, рассматриваются такие вопросы как анализ ROC-кривых, анализ распределения скоринговых баллов, расчет статистики Колмогорова-Смирнова, коэффициента Джини, коэффициента дивергенции коэффициента разделения.

Авторские выводы и рекомендации могут быть использованы специалистами по управлению рисками в коммерческих банках при построении скоринговых систем и проверки их работы.

Ключевые слова: Кредитный риск; кредитный скоринг; логистическая регрессия; коммерческий банк; управление рисками; валидация модели; статистика Колмогорова-Смирнова; классификация заемщиков; качество классификации; ROC-анализ.

Идентификационный номер статьи в журнале 173EVN214

¹ 119501, г. Москва, ул. Нежинская, 7, МЭСИ, кафедра Математической статистики и эконометрики

² 125190, г. Москва, Ленинградский пр-кт, д. 80, МФПУ «Синергия», кафедра Бизнес-статистики

Одна из ключевых задач при управлении индивидуальными кредитными рисками в банке — оценка вероятности дефолта по кредиту у потенциальных заемщиков. Для оценки этой вероятности могут применяться различные статистические модели, но наибольшее распространение на практике получила модель логистической регрессии:

$$\ln\left(\frac{p_i}{1-p_i}\right) = b_0 + b_1x_i^{(1)} + b_2x_i^{(2)} + \dots + b_kx_i^{(k)}, \quad (1)$$

где p_i — вероятность наступления дефолта по кредиту для i -го заемщика; $x_i^{(j)}$ — значение j -ой независимой переменной; b_0 — независимая константа модели, b_j — параметры модели.

Зависимой переменной в модели логистической регрессии, как правило, выступает вероятность наступления просрочки по кредиту более 90 дней, а независимыми данные по кредиту, социально-демографические данные о заемщике, данные бюро кредитных историй. На основе полученных оценок коэффициентов логистической регрессии строится скоринговая карта, переводящая коэффициенты модели в скоринговые баллы. Именно по набранному суммарному скоринговому баллу происходит ранжирование заемщиков и принимается решение о выдаче кредита.

Заключительным этапом построения модели логистической регрессии является проверка ее достоверности и апробация на реальных данных. О степени валидации (от англ. от англ. validity — доказанность, обоснованность, пригодность) модели логистической регрессии говорит способность ее правильно классифицировать заемщиков, способность модели отличать «хороших» заемщиков от «плохих». Модель должна давать корректные прогнозы не только на обучающей совокупности, но и на практике при ее применении.

Одна из стратегий валидации модели — формирование случайным образом двух выборок: обучающей — по ней строится модель, и тестовой — используется для проверки модели. Проверку достоверности модели, как правило, производят на обучающей и контрольной выборке в пропорциях примерно 70–80% и 30–20% соответственно от исходных данных для построения модели. Хорошая модель должны давать приемлемые результаты точности и на обучающей, и на контрольной выборке. Схожие показатели, полученные на обеих выборках — признак того, что на практике модель будет выдавать верные прогнозы.

Более сложная стратегия валидации модели может предполагать формирование трех и более выборок. Например, первая выборка используется для оценки параметров модели. Вторая выборка — для проверки модели. Если получаются значительные отклонения результатов по обучающей и тестовой выборке, то из них удаляются выбросы или переменные, влияющие на эти отклонения, а затем строится новая модель по объединенной первой и второй выборке. Результаты новой модели проверяются на заранее зарезервированной третьей выборке.

Оценка дискриминирующей способности модели логистической регрессии может быть выполнен на основе анализ *таблицы классификации*. Таблица классификации представляет собой таблицу сопряженности из двух строк и двух столбцов между переменными фактических и предсказанных значений зависимой переменной. Обычно в такой таблице сопряженности приводятся не суммы по строкам и столбцам, а процент корректных предсказаний по категориям зависимой переменной и общий процент всех предсказаний по модели (см. таблицу 1).

Таблица 1

Пример таблицы классификации

Прогноз дефолта по модели	Фактическое наличие дефолта		Процент корректных предсказаний
	Да	Нет	
Да	770	250	75.5%
Нет	224	1220	84.5%
Итого	80.8%		

Результаты классификации по модели логистической регрессии иногда представляют графически в виде *гистограммы оценок классификации* (рисунок 1)³. На гистограмме буквами обозначаются градации предсказанной зависимой переменной: «П» — «плохой» заемщик, если прогнозное и фактическое значение зависимой переменной не совпадают; «Х» — «хороший» заемщик, если прогнозное и фактическое значение зависимой переменной не совпадают. Каждый столбик на гистограмме соответствует определенной предсказанной вероятности, а его высота — количеству наблюдений, для которых предсказана данная вероятность. На гистограмме классификации букве «Х» в правой части, а букве «П» в левой части соответствуют правильные предсказания. В случае идеальной классификации все буквы «Х» должны быть левее буквы «П», а разделять их будет вероятность отсечения «плохих» и «хороших» заемщиков⁴. Чем лучше качество прогнозирования построенной модели, тем теснее наблюдения на гистограмме будут сгруппированы у соответствующих концов левой и правой оси.

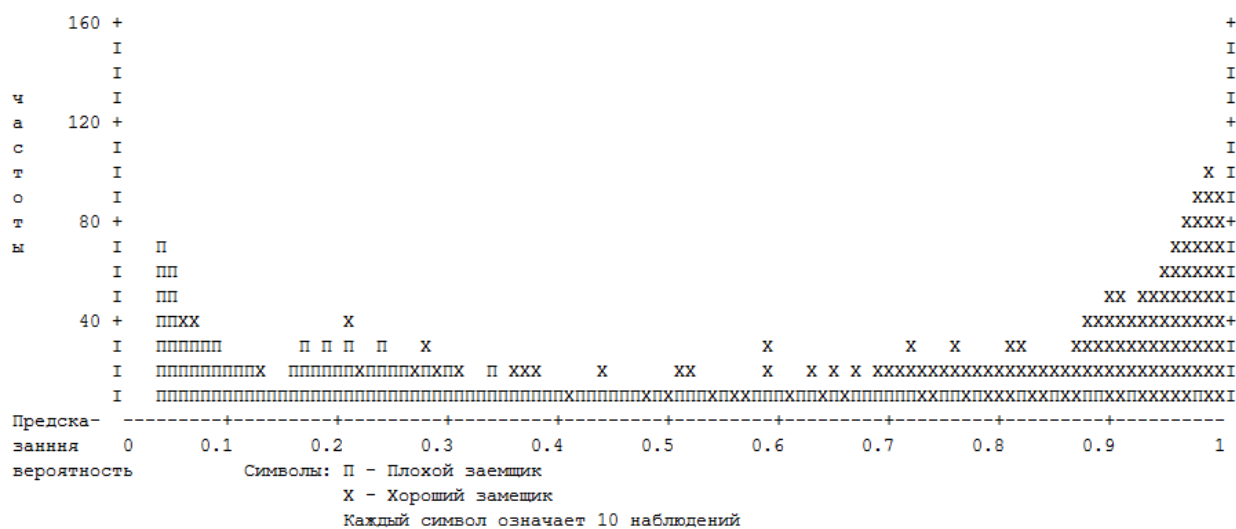


Рис. 1. Пример построения гистограммы классификации

Для оценки качества классификации часто прибегают к анализу *ROC-кривых* (от англ. *Receiver Operator Characteristic*). Этот анализ также позволяет выбрать оптимальный порог вероятности разделения «плохих» и «хороших» заемщиков для достижения приемлемого уровня чувствительности и специфичности модели.

³ Данная гистограмма построена с помощью статистического пакета IBM SPSS Statistics в модуле построения логистической регрессии.

⁴ На данной гистограмме вероятность отсечения равна 0.5.

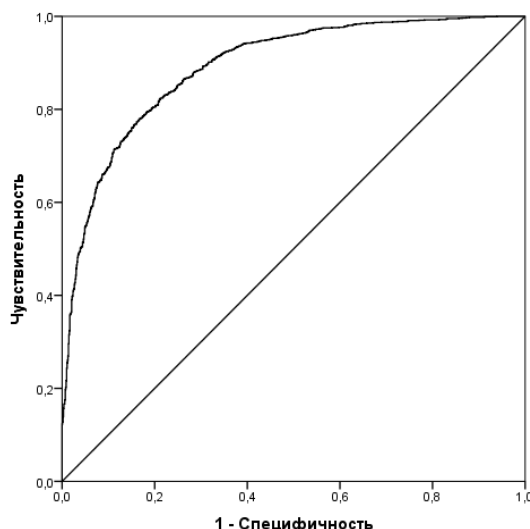


Рис. 2. Пример построения ROC-кривой

ROC-кривая показывает зависимость доли верно классифицированных положительных исходов (чувствительности) от доли неверно классифицированных отрицательных исходов (единица минус специфичность). ROC-кривая может строиться по значениям спрогнозированных вероятностей дефолта по модели логистической регрессии или по результатам классификации на основе построенной скоринговой карты. В первом случае ROC-анализ позволяет выбрать порог вероятности разделения «плохих» и «хороших» заемщиков для достижения приемлемого уровня чувствительности и специфичности модели. А во втором — оптимальный скоринговый балл отсечки «плохих» и «хороших» заемщиков.

Модель с высокой чувствительностью будет консервативной с ужесточенным отбором клиентов — максимальным предотвращением пропуска «плохих» заемщиков. Задача анализа чувствительности — минимизировать кредитный риск, связанный с выдачей кредита. Модель с высокой специфичностью будет менее тщательно выявлять «плохих» заемщиков и будет являться рискованной. Задача анализа специфичности в кредитном скоринге — минимизировать риск упущенной выгоды, связанной с отказом в выдаче кредита. Для определения оптимального порога классификации существуют множество критериев, например:

- 1) уровень плохих кредитов в портфеле одобренных заявок;
- 2) минимизация ошибок классификации и др.

Выбор оптимального значения порога отсечения зависит от цены совершения ошибки первого и второго рода при классификации. Модель должна точнее классифицировать «плохих» заемщиков, т.к. в кредитном скоринге цена ошибки первого рода выше. При снижении порога отсечения в модели будет увеличиваться чувствительность, т.е. способность модели правильно выявлять тех заемщиков, у которых будет просрочка платежа. За оптимальный порог отсечения можно взять точку баланса между чувствительностью и специфичностью.

Для анализа классификационной способности скоринговой модели используют показатель *площади под ROC-кривой* — AUC (от англ. *Area Under Curve*). Площадь под кривой AUC изменяется от 0.5 (нет разделения) до 1 (идеальное разделение). Обычно считают, что значение площади от 0.9 до 1 соответствует отличному качеству модели, от 0.8–0.9 — очень хорошему, 0.7–0.8 — хорошему, 0.6–0.7 — среднему, 0.5–0.6 — неудовлетворительному. Следует отметить, что показатель площади под кривой предназначен

только для сравнительного анализа моделей между собой. Значение площади под кривой не содержит никакой информации о чувствительности и специфичности модели.

При анализе качества модели по значению площади под ROC-кривой часто вычисляют *индекс Джини*. Этот показатель переводит значение площади под кривой в диапазон от 0 до 1, чем выше его величина, тем выше дискриминирующая способность модели. Рассчитывается индекс Джини по формуле:

$$D = 2 \cdot (AUG - 0.5), \quad (2)$$

где *AUG* — площадь по ROC-кривой.

Качество классификации разработанной на основе логистической регрессии скоринговой карты можно оценить, проанализировав *распределение скоринговых баллов* заемщиков (см. рисунок 3). При ранжировании заемщиков по хорошей скоринговой карте «плохие» и «хорошие» заемщики должны иметь максимально различающиеся скоринговые баллы. Например, «плохим» заемщикам присваивался меньший балл, чем «хорошим». В итоге группа «плохих» кредитов должна иметь в сумме меньше баллов, чем группа «хороших». Чем более явно будут разделены распределения скоринговых баллов для «плохих» и «хороших» кредитов, тем эффективнее будет работать скоринговая карта. Идеальной считается скоринговая карта, при которой распределения баллов у «плохих» и «хороших» заемщиков не пересекаются, а находятся рядом друг с другом.

При анализе качества скоринговой модели также анализируют значения *процентилей*, полученных по *распределению скоринговых баллов*. Обычно анализируют значения от 10 до 20% в нижней части распределения скорингового балла. Анализ процентилей важен для определения балла, ниже которого кредиты не выдаются. При анализе процентилей обращают внимание, какой процент «плохих» кредитов от их общего числа располагается в нижней части распределения скорингового балла. Например, если общий кумулятивный процент «плохих» кредитов равен 20% при интервале полученных скоринговых баллов 200-220, то пороговый скоринговый балл для отсеечения «плохих» заемщиков выбирают 220. Чем больше «плохих» потенциальных заемщиков будет в нижних процентах распределения скорингового балла, тем больше «плохих» кредитов может быть исключено при помощи соответствующего данному процентилю значения отсеечения скорингового балла.

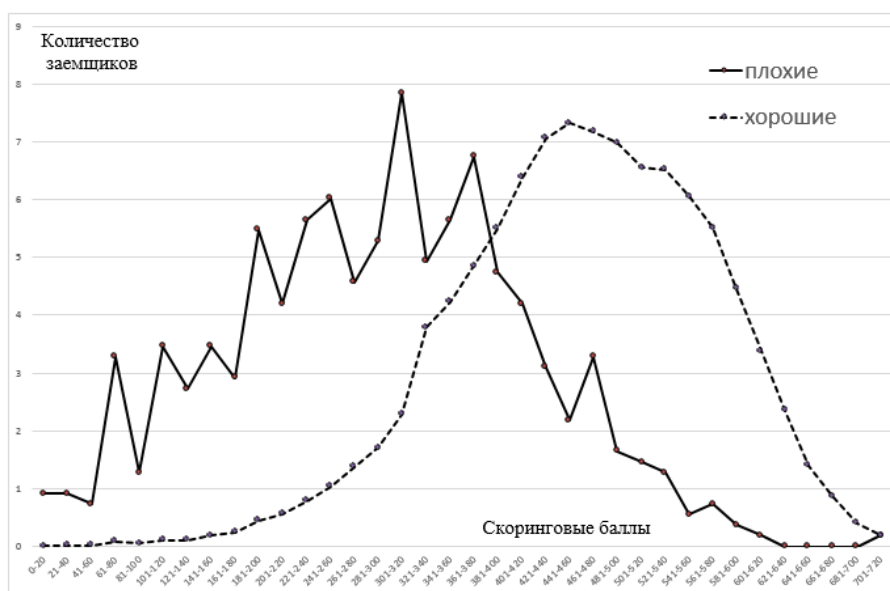


Рис. 3. Пример распределения скоринговых баллов

Для оценки качества прогнозирования модели логистической регрессии в кредитном скоринге используют тест или статистику Колмогорова-Смирнова. В тесте Колмогорова-Смирнова проверяется статистическая гипотеза, что две произвольные выборки принадлежат одной генеральной совокупности. Применительно к скорингу сравниваются два кумулятивных распределения скоринговых баллов «хороших» и «плохих» заемщиков. Статистика Колмогорова-Смирнова вычисляется как максимальная разница между кумулятивными функциями распределения «плохих» и «хороших» заемщиков:

$$KS = \max_x |F_m(x) - G_n(x)| \cdot 100, \quad (3)$$

где $F_m(x)$ и $G_n(x)$ — эмпирические кумулятивные распределения скорингового балла для «плохих» и «хороших» заемщиков;

n, m — количество «плохих» и «хороших» заемщиков.

Для вычисления статистики Колмогорова-Смирнова и проверки гипотезы о равенстве двух функций распределений необходимо использовать следующий алгоритм. Заемщиков проранжировать в порядке увеличения скорингового балла и провести группировку. Группировочным признаком выступает набранный скоринговый балл. Затем в каждой полученной группе заемщиков необходимо рассчитать следующие показатели:

- количество «хороших» заемщиков;
- количество «плохих» заемщиков;
- отношение шансов «плохих» к «хорошим» заемщикам;
- процент «плохих» и «хороших» кредитов;
- кумулятивную сумму «плохих» и «хороших» кредитов;
- кумулятивный процент «плохих» и «хороших» кредитов;
- общий кумулятивный процент плохих кредитов от их общего числа;
- разницу между кумулятивными процентам плохих и хороших кредитов.

После чего нужно найти максимальную разность между кумулятивным процентом «хороших» и «плохих» кредитов и вычислить по формуле (3) статистику Колмогорова-Смирнова (см. рисунок 4). Полученное значение статистики необходимо сравнить с табличным значением по таблице распределения Колмогорова-Смирнова с выбранным уровнем значимости или при числе «плохих» и «хороших» заемщиков соответственно больше 80 можно взять приближенное пороговое значение, вычисляемое по формуле:

$$z(\alpha) \sqrt{\frac{m+n}{mn}}, \quad (4)$$

где $z(\alpha)$ — значение соответствующее выбранному уровню значимости.

Если расчетное значение статистики по формуле (3) меньше порогового по таблице или по формуле (4), то гипотезу о равенстве двух функций распределений отвергают. Диапазон изменения значений статистики Колмогорова-Смирнова может быть от 0 до 100. Чем выше значение статистики Колмогорова-Смирнова, тем лучше классифицирует модель. Крайние значения статистика Колмогорова-Смирнова, как правило, не принимает, и ее значения сосредоточены обычно от 20-25 до 75-80. Можно придерживаться следующей шкалы для оценки качества модели по статистике Колмогорова

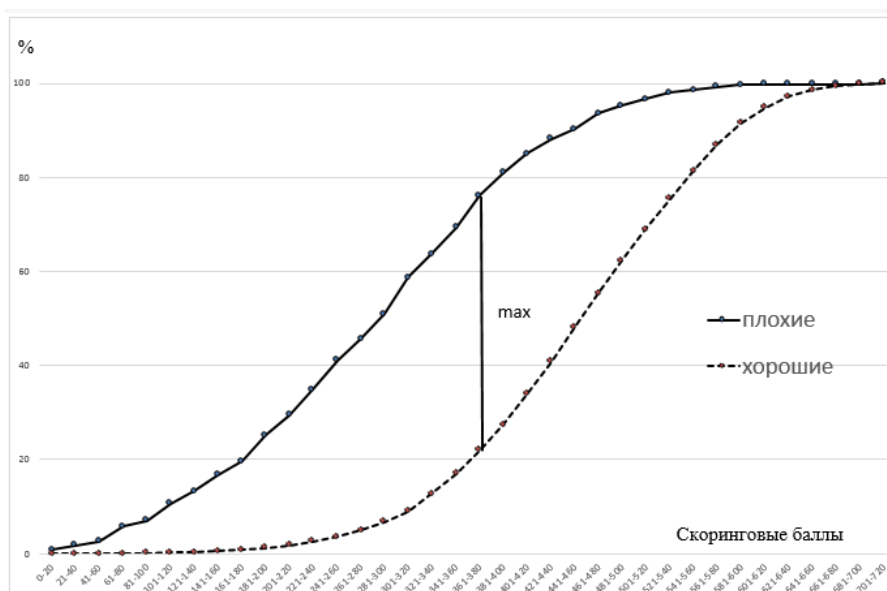


Рис. 4. Графическая иллюстрация расчета статистики Колмогорова-Смирнова

Альтернативная мера оценки валидации модели является *коэффициент дивергенции* (от англ. *Divergence*). Коэффициент дивергенции представляет собой оценку разницы математических ожиданий распределений скоринговых баллов для «плохих» и «хороших» заемщиков, нормализованную дисперсиями этих распределений, и рассчитывается по формуле⁵:

$$D = \frac{2 \cdot (\bar{x}_1 - \bar{x}_2)^2}{(s_1^2 + s_2^2)}, \quad (5)$$

где \bar{x}_1 и \bar{x}_2 — средние значения скорингового балла для «плохих» и «хороших» заемщиков; s_1^2 и s_2^2 — дисперсии скорингового балла для «плохих» и «хороших» заемщиков.

Чем больше значение коэффициента дивергенции, тем лучше качество модели с точки зрения ее классификационной способности. В большинстве случаев на практике используют все же статистику Колмогорова-Смирнова. Однако этот показатель не всегда может давать адекватную оценку валидности модели. Возможность применения статистики Колмогорова-Смирнова, а также коэффициента дивергенции, иллюстрирует график распределения скоринговых баллов, а именно их симметричность и наложение друг на друга.

Значение коэффициента дивергенции чувствительно к асимметрии распределений скоринговых баллов и может давать заниженную или завышенную оценку в зависимости от направления асимметрии. Статистика Колмогорова-Смирнова, наоборот, устойчива к асимметрии распределений скоринговых баллов. Однако статистика Колмогорова-Смирнова может давать ошибочную оптимистичную оценку при наложении кривых распределений скоринговых баллов друг на друга. Когда обе кривые распределения скоринговых баллов для «плохих» и «хороших» заемщиков нормально или приблизительно нормально распределены — можно использовать и статистику Колмогорова-Смирнова, и коэффициент дивергенции. Если кривые распределения скошены внутрь — дивергенция будет очень большой и не ее использование будет не корректно. Если обе кривые распределения будут скошены наружу — дивергенция будет очень маленькой и также не будет являться адекватной мерой качества

⁵ Этот показатель еще называют мерой энтропии Кульбака–Лейбера.

валидации модели. В таблице 2 приведено сравнение возможности применения критериев валидации в зависимости от формы распределения скоринговых баллов.

Таблица 2

Сравнение возможности применения показателей валидности модели

Распределение скоринговых баллов	Статистика Колмогорова-Смирнова	Коэффициент дивергенции
Симметричны	подходит	подходит
Скошены внутрь	подходит	не подходит (будет завышен)
Скошены наружу	подходит	не подходит (будет занижен)
Вложены друг в друга	не подходит (будет завышена)	подходит

Еще одной мерой качества построенной модели может быть *коэффициент разделения*. Алгоритм расчета коэффициента разделения следующий. Сначала рассчитывают эмпирические кумулятивные распределения скорингового балла для группы «плохих» и «хороших» заемщиков. Затем для каждого значения скорингового балла или интервала скоринговых баллов принимают две кумулятивные эмпирические вероятности — координаты по оси абсцисс и оси ординат. После чего необходимо начертить координаты под прямым углом и найти площадь между начерченной кривой и прямой, расположенной под углом 45° (см. рисунок 5). Если начерченная кривая будет лежать частично выше или частично ниже прямой, необходимо будет найти абсолютную разность между площадью выше этой прямой и площадью ниже ее. Такой случай будет наблюдаться, когда одна кривая распределения баллов будет находится внутри другой. После чего нужно абсолютную разность площадей разделить на площадь треугольника, находящуюся под прямой, расположенной под углом 45° — полученная величина будет являться коэффициентом разделения. Данный алгоритм похож на построение кривой Лоренца и требует вычисление площадей под кривыми. Эта задача может быть выполнена, например, аппроксимации площади методом трапеций.

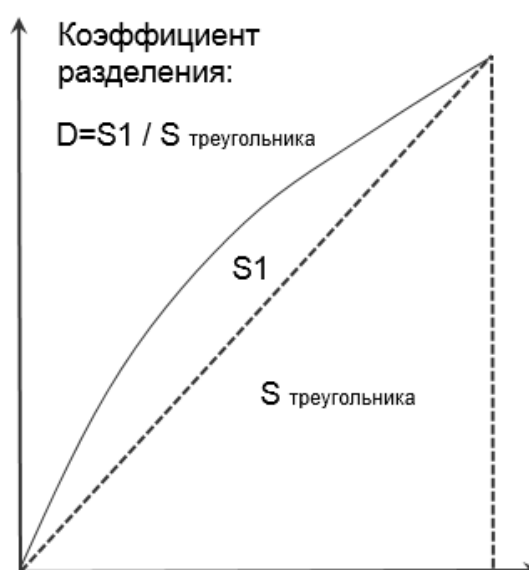


Рис. 5. Иллюстрация расчета коэффициента разделения

Итак, процент ошибок верных или неверных предсказанный по полученной скоринговой модели не может являться исчерпывающей характеристикой ее валидности. На практике валидность модели можно оценить по совокупности критериев. В дополнение к

таблице классификации для полученной модели рассчитывают статистические показатели ее качества. Первое направление анализа валидности скоринговой модели может быть основано на построении ROC–кривой и расчете показателей, производных от нее (коэффициент Джини, AUC). Второе направление — анализ распределения скоринговых баллов и расчет показателей на их основе (статистика Колмогорова-Смирнова, коэффициент разделения и коэффициент дивергенции). Статистика Колмогорова-Смирнова является не надежной оценкой валидности модели в случае сильного наложения распределений скоринговых баллов «плохих» и «хороших» кредитов. Рассмотренные показатели валидности скоринговых моделей наиболее часто используемые на практике. Однако могут и применяться другие критерии и методы такие, как построение диаграммы выигрышей, Байесовский критерий Шварца, критерий Акаике, сравнение распределений скоринговых баллов критерием Манна-Уитни Вилкоксона и др.

ЛИТЕРАТУРА

1. Сорокин, А.С. Применение законов распределения случайных величин для моделирования экономических явлений и процессов [Текст] : монография. / Н.Я. Бамбаева Н.Я., А.С. Сорокин – М.: МЭСИ, 2010. – 156 с. – ISBN 978-5-7764-0612-6
2. Руководство по кредитному скорингу [Текст] / под. ред. Элизабет Мэйз ; пер. с англ. И. М. Тикота ; науч. ред. Д. И. Вороненко. – Минск: Гревцов Пабlishер, 2008. – 464 с. – ISBN 978-985-6569-34-3
3. Сиддики, Н. Скоринговые карты для оценки кредитных рисков. Разработка и внедрение интеллектуальных методов кредитного скоринга [Текст] / Наим Сиддики; пер. с англ. Евгения Ильичева. – М.: Манн, Иванов и Фабер, 2014. – 268 с. – ISBN 978-5-91657-899-7
4. Смирнов, Н.В. Оценка расхождения между эмпирическими кривыми распределения в двух независимых выборках [Текст] / Н.В. Смирнов // Бюллетень МГУ, серия А. – Т.2. – №2. – 1939. – с.3-14.
5. Сорокин, А.С. К вопросу оценки согласованности мнений экспертов при использовании методов экспертного оценивания в кредитном скоринге. [Текст] /А.С. Сорокин // Роль бизнеса в трансформации общества – 2014: Сб. ст. по мат. IX междунар. научн. конгр. – М.: «Эдитус», 2014. – с. 281-283. – ISBN 978-5-00058-089-9
6. Улитина, Е.В. Статистика: учебное пособие [Текст] / Е.В. Улитина, О.В. Леднева, О.Л. Жирнова – М.: Московский финансово-промышленный университет «Синергия», 2013. – 320 с. – ISBN: 978-5-4257-0107-7
7. Улитина, Е.В. Статистика: учебное пособие [Текст] / Е. В. Улитина, О. В. Леднева, О. Л. Жирнова; под ред. Е. В. Улитиной. - 3-е изд., стер. - Сер. Университетская серия. – М: МФПА, 2011. – 320 с. – ISBN: 978-5-902597-30-8
8. Улитина, Е.В. Применение метода анализа иерархий при согласовании результатов оценки [Текст] / С.В. Харитонов, Е.В. Улитина, В.В. Дик // Прикладная информатика. – 6 (42). – 2012. – с. 108-113
9. Allison, P.D. (1999). Logistic regression using the SAS system: theory and application. [Text] – Cary, NC: SAS Institute. – 303 p. – ISBN 1580253520
10. Anderson R. (2007). The credit scoring toolkit: theory and practice for retail credit risk management and decision automation. [Text] – New York: Oxford University press. – 790 p. – ISBN 0199226407
11. Harrell, Frank. (2001). Regression modeling strategies. [Text] – NY: Springer. – 608 p. – ISBN 0387952322, 9780387952321
12. Hosmer D., Lemeshow S. (1989, 2000, 2013). Applied logistic regression. [Text] – New York: John Wiley and Sons. – 528 p. – 3rd ed. – ISBN 0470582472, 9780470582473
13. Jaccard J. (2001). Interaction effects in logistic regression. [Text] – Thousand Oaks: Sage Publications. – 70 p. – ISBN 0761922075
14. Kleinbaum D. G. (1994). Logistic regression: A Self-Learning Text. [Text] – New York: Springer-Verlag. – 282 p. – ISBN 0387941428

15. Lewis E. M. (1992). An introduction to credit scoring. [Text] – San Rafael: The Athena Press. – 172 p. , – ISBN 9995642239, 978-9995642235
16. Lyn C. Thomas. (2009). Consumer credit models: pricing, profit, and portfolios. [Text] –New York: Oxford University press. – 400 p. – ISBN 9780199232130
17. Mays E. (ed.) (2001). Handbook of credit scoring. [Text] – Chicago: Glenlake Publishing Company Ltd/Fitzroy Dearborn Publishers. – 382 p. – ISBN 1888988010, 978-1888988017
18. Naeem S. (2006). Credit risk scorecards: developing and implementing intelligent credit scoring. [Text] – New Jersey: John Wiley and Sons. – 208 p. – ISBN: 9780471754510

Рецензент: Емельянов Александр Анатольевич, главный редактор научного журнала «Прикладная информатика», Московский финансово-промышленный университет «Синергия».

Alexander Sorokin

Moscow state university of economics, statistics and informatics (MESI)
Moscow university for industry and finance «Synergy»
Russia, Moscow
E-Mail: alsorokin@mail.ru

On the question of validation logistic regression model in credit scoring

Abstract: Construction of automated scoring systems allows banks to reduce individual credit risks. Use of scoring systems based on statistical models along with expert estimates underwriters is a common practice. Logistic regression is the most commonly used method for constructing scoring systems in banks.

The article considers the question of the application of logistic regression to classify borrowers in credit scoring. Its validation phase is important final stage of construction of the scoring system, it consists in the verification of the resulting model for the training sample, the control sample and real data. Validation of the scoring model should be based on a quality criteria system.

Firstly, this article systematizes methods for assessing the reliability of the scoring model. Secondly, it presents a comparative analysis of methods for assessing the effectiveness of ranking borrowers on a logistic regression model, as well as scorecards, which was constructed on the basis of logistic regression. In third place, the article gives an overview of strategies for the validation of scoring models.

Keywords: Credit risk; credit scoring; logistic regression; commercial bank; risk management; validation of the model; Kolmogorov-Smirnov statistics; classification of prospective borrowers; quality classification; ROC-analysis.

Identification number of article 173EVN214

REFERENCES

1. Sorokin, A.S. Primenenie zakonov raspredelenija sluchajnyh velichin dlja modelirovanija jekonomicheskikh javlenij i processov [Tekst] : monografija. / N.Ja. Bambaeva N.Ja., A.C. Sorokin – M.: MJeSI, 2010. – 156 c. – ISBN 978-5-7764-0612-6
2. Rukovodstvo po kreditnomu skoringu [Tekst] / pod. red. Jelizabet Mjejjz ; per. s angl. I. M. Tikota ; nauch. red. D. I. Voronenko. – Minsk: Grevcov Publisher, 2008. – 464 s. – ISBN 978-985-6569-34-3
3. Siddiki, N. Skoringovyje karty dlja ocenki kreditnyh riskov. Razrabotka i vnedrenie intellektual'nyh metodov kreditnogo skoringa [Tekst] / Naim Siddiki; per. s angl. Evgenija Il'icheva. – M.: Mann, Ivanov i Faber, 2014. – 268 s. – ISBN 978-5-91657-899-7
4. Smirnov, N.V. Ocenka rashozhdenija mezhdru jempirichesкими kривыми raspredelenija v dvuh nezavisimyh vyborkah [Tekst] / N.V. Smirnov // Bjulleten' MGU, serija A. – T.2. – №2. – 1939. – s.3-14.
5. Sorokin, A.S. K voprosu ocenki soglasovannosti mnenij jekspertov pri ispol'zovanii metodov jekspertnogo ocenivaniya v kreditnom skoringe. [Tekst] /A.C. Sorokin // Rol' biznesa v transformacii obshhestva – 2014: Sb. st. po mat. IX mezhdunar. nauchn. kongr. – M.: «Jeditus», 2014. – s. 281-283. – ISBN 978-5-00058-089-9
6. Ulitina, E.V. Statistika: uchebnoe posobie [Tekst] / E.V. Ulitina, O.V. Ledneva, O.L. Zhirnova – M.: Moskovskij finansovo-promyshlennyj universitet «Sinergija», 2013. – 320 c. – ISBN: 978-5-4257-0107-7
7. Ulitina, E.B. Statistika: uchebnoe posobie [Tekst] / E. V. Ulitina, O. V. Ledneva, O. L. Zhirnova; pod red. E. V. Ulitinoj. - 3-e izd., ster. - Ser. Universitetskaja serija. – M: MFPA, 2011. – 320 s. – ISBN: 978-5-902597-30-8
8. Ulitina, E.V. Primenenie metoda analiza ierarhij pri soglasovanii rezul'tatov ocenki [Tekst] / S.V. Haritonov, E.V. Ulitina, V.V. Dik // Prikladnaja informatika. – 6 (42). – 2012. – c. 108-113
9. Allison, P.D. (1999). Logistic regression using the SAS system: theory and application. [Text] – Cary, NC: SAS Institute. – 303 p. – ISBN 1580253520
10. Anderson R. (2007). The credit scoring toolkit: theory and practice for retail credit risk management and decision automation. [Text] – New York: Oxford University press. – 790 p. – ISBN 0199226407
11. Harrell, Frank. (2001). Regression modeling strategies. [Text] – NY: Springer. – 608 p. – ISBN 0387952322, 9780387952321
12. Hosmer D., Lemeshow S. (1989, 2000, 2013). Applied logistic regression. [Text] – New York: John Wiley and Sons. – 528 p. – 3rd ed. – ISBN 0470582472, 9780470582473
13. Jaccard J. (2001). Interaction effects in logistic regression. [Text] – Thousand Oaks: Sage Publications. – 70 p. – ISBN 0761922075
14. Kleinbaum D. G. (1994). Logistic regression: A Self-Learning Text. [Text] – New York: Springer-Verlag. – 282 p. – ISBN 0387941428

15. Lewis E. M. (1992). An introduction to credit scoring. [Text] – San Rafael: The Athena Press. – 172 p. , – ISBN 9995642239, 978-9995642235
16. Lyn C. Thomas. (2009). Consumer credit models: pricing, profit, and portfolios. [Text] –New York: Oxford University press. – 400 p. – ISBN 9780199232130
17. Mays E. (ed.) (2001). Handbook of credit scoring. [Text] – Chicago: Glenlake Publishing Company Ltd/Fitzroy Dearborn Publishers. – 382 p. – ISBN 1888988010, 978-1888988017
18. Naeem S. (2006). Credit risk scorecards: developing and implementing intelligent credit scoring. [Text] – New Jersey: John Wiley and Sons. – 208 p. – ISBN: 9780471754510