

образовании: сб. науч. статей по материалам Междунар. науч. конф. ICIT 2014. Саратов, 2014. С. 16-22.

10. Сухоруков К.Ю. Информационная поддержка процессов разработки программ дополнительного образования: постановка задачи / К.Ю. Сухоруков, Л.К. Бобров // Вестник НГУЭУ. 2014. № 4. С. 288-303.

11. Коротченко, Е.А., Родионова З.В. Метод проверки профессиональной компетентности с помощью игровой имитации в технологии кадрового продюсирования // В мире научных открытий. 2014. № 2.1 (50). С. 670-681.

Власенко А.И., Шульга Т.Э.

О ВОЗМОЖНОСТЯХ ПРИМЕНЕНИЯ МОДЕЛИ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ

Регрессионные модели позволяют описать зависимость одной переменной (зависимой) от одной или нескольких независимых переменных. Регрессионные модели имеют вид:

$$y = f(x, w),$$

где y – зависимая переменная, $x \in X$ – множество независимых переменных, $w \in W$ – множество параметров.

Построение регрессионной модели заключается в определении значений параметров w .

Частным случаем регрессионных моделей является логистическая регрессия. Данная регрессионная модель имеет две отличительные особенности. Во-первых, зависимая переменная должна быть бинарной. Во-вторых, анализируются не только количественные, но и качественные данные, предварительно перекодированные определенным образом. Кодирование осуществляется следующим образом: каждому возможному значению переменной присваивается числовой эквивалент. Например, переменная пол может принимать два значения: мужской и женский. Кодирование состоит в замене значения «мужской» на 1 и «женский» на 2. Далее, при построении модели, используются только числовые значения.

Модель логистической регрессии имеет следующий вид:

$$P(x) = \frac{1}{1+e^{-z}},$$

где P – вероятность наступления события; e – основание натуральных логарифмов; z – уравнение линейной регрессии, вида $z = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n$.

Согласно проведенному анализу современных исследований связанных с использованием логистической регрессии было выявлено несколько особенностей. Во-первых, применение этой модели наиболее распространено в социально-экономических исследованиях, хотя есть опыт применения в работах технического характера. Во-вторых, с

помощью этой модели решают три типа задач: прогнозирование, классификация, оценка значимости не зависимых переменных в модели.

При использовании логистической регрессии возникает ряд трудностей:

При классификации объектов, которые близки к границе класса (вероятность близка к 0,5), возникает ошибка соотнесения с классом значения. Данная проблема представлена на рис. 1. Например, в [1] модель логистической регрессии используется для решения задачи кредитного скоринга. Задача заключается в классификации клиентов на «хороших» или «плохих». Автор отмечает, что клиентов, для которых вероятность возвращения кредита близка к 0,5, невозможно классифицировать как «хороших» или «плохих».

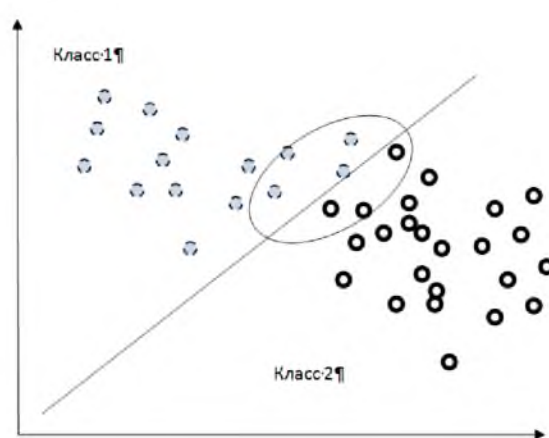


Рис. 1. Классификация объектов, находящихся на границе классов

Модель логистической регрессии не учитывает связи между не зависимыми переменными (мультиколлинеарность). В [2] отмечается, что пренебрежение зависимостями между не зависимыми переменными ведет к построению ошибочных моделей. Автор предлагает проводить предварительное статистическое исследование и исключать такие переменные из анализа.

Модель плохо обучается на выборках, имеющих неравномерное распределение классов значений зависимой переменной. Например, в работе [4] автор отмечает, что при обучении модели на выборке с неравномерным распределением классов значений зависимой переменной была получена неадекватная модель. После «выравнивания» выборки и повторного обучения модель показала 99% точность.

Проблема «границ чувствительности», например в кредитном скоринге клиенты в возрасте 30 и 31 год практически одинаковые группы, а клиенты с возрастом 60 и 61 год – весьма разные группы заемщиков [5]. Логистическая регрессия не может распознавать такого рода зависимости.

Тем не менее большинство исследований показывают эффективность модели логистической регрессии. Кроме высокой точности

прогнозирования, стоит отметить следующие достоинства: решать задачи различного масштаба: от 3 независимых [6] переменных до 230 [7], от 40 записей [8] до 20 млн [9]. Построенные модели достаточно просто интерпретировать. Несмотря на существующие ограничения, модель логистической регрессии показывает хорошие результаты для исследований в различных предметных областях.

Литература

1. Симонов П.М., Лазуков С.И. Оценка кредитного риска: актуальные практические вопросы // Вестник Пермского университета. Сер. Экономика. 2009. № 1. С. 61-67.
2. Мурадов Д.А. Logit-регрессионные модели прогнозирования банкротства предприятий // Труды Российского государственного университета нефти и газа им. И.М. Губкина. 2011. № 3. С. 160-172.
3. Сорокин А.С. Построение скоринговых карт с использованием модели логистической регрессии // Науковедение: Интернет-журнал. 2014. № 2. С. 82.
4. Середний С.С. Оценивание вероятности дефолта по кредитным операциям с использованием логистической регрессии и кластерного анализа // Системні дослідження та інформаційні технології. 2011. № 1. С. 126-132.
5. Сапонов Д.И. Опыт конкурентной борьбы как фактор академической успеваемости // Мониторинг общественного мнения: экономические и социальные перемены. 2013. № 6 (118). С. 113-126.
6. Симонова С.И. Интеллектуальный анализ данных для задач CRM // International Journal of Open Information Technologies. 2015. Т. 3. № 2. С. 17-22.
7. Осиков М.В., Ахматов К.В. Использование логистической регрессии в оценке изменений психологического статуса у больных хронической почечной недостаточностью, находящихся на гемодиализе // Вестник Южно-Уральского государственного университета. Сер. Образование, здравоохранение, физическая культура. 2010. № 19 (195). С. 34-37.
8. Богданов Л.Ю. Оценка эффективности бинарных классификаторов на основе логистической регрессии методом ROC-анализа // Вестник СГТУ. 2010. Т. 4. № 2с. С. 92-97.

Богданов В.Ю., Ивженко С.П.

ИСПОЛЬЗОВАНИЕ ОБЛАЧНЫХ ТЕХНОЛОГИЙ В ОБРАЗОВАНИИ

Информатизация процесса образования – это необходимость и уже свершившийся факт. В этой работе обосновывается применение в учебной работе сочетания платных и бесплатных облачных сервисов как относительно нового и перспективного для образовательных учреждений направления информатизации.

Углубление информатизации в образовании требует своевременного массового обновления техники и программного обеспечения (ПО),