2. G. Amato, F. Carrara, F. Falchi, C. Gennaro, C. Meghini, and C. Vairo. Deep learning for decentralized parking lot occupancy detection. Expert Systems with Applications, 72:327?334, 2017. doi:10.1016/ j.eswa.2016.10.055.

3. Q. Tan, J. Ling, J. Hu, X. Qin and J. Hu, "Vehicle Detection in High Resolution Satellite Remote Sensing Images Based on Deep Learning,"in IEEE Access, vol. 8, pp. 153394-153402, 2020, doi: 10.1109/ ACCESS.2020.3017894.

4.Amato, Giuseppe Carrara, Fabio Falchi, Fabrizio Gennaro, Claudio Meghini, Carlo Vairo, Claudio. CNRPark+EXT DATASET DETAILS. 2017. http://cnrpark.it/

5. Jacob Solawetz. Vehicles-OpenImages Dataset. https://public.roboflow.com/object-detection/vehicles-openimages.

МАТЕМАТИЧЕСКАЯ КРИПТОГРАФИЯ И АНАЛИЗ ДАННЫХ

ОТБОР ИНФОРМАТИВНЫХ ПРИЗНАКОВ ДЛЯ КЛАССИФИКАЦИИ ЭНДОВАСКУЛЯРНЫХ ОПЕРАЦИЙ

Абрамович М.С.[1], Козловский В.В.[2], Стельмашок В.И.[3]

[1]НИИ прикладных проблем математики и информатики, БГУ, Независимости 4, 220050 Минск, Беларусь
abramovichms@bsu.by
[3]ООО “СК хайникс мемори солюшенс Восточная Европа”, Немига 5, 220030 Минск, Беларусь
kozlovskivlad@yandex.ru
[3]РНПЦ “Кардиология”, Р. Люксембург 110Б, 220036 Минск, Беларусь
stelval@yandex.by

При проведении эндоваскулярных операций существует вероятность перфорации коронарных артерий. В связи с этим представляется целесообразным разработать систему прогнозирования риска развития перфораций в ходе выполнения оперативного вмешательства, что позволит предотвратить данные осложнения. Построение системы прогнозирования риска проводилось с использованием методов машинного обучения на основании рентгенанатомических и клинических признаков.

Проведено исследование, которое включало данные 397 пациентов, которым проводилась эндоваскулярная операция. Все пациенты были разделены на 2 группы: лица у которых была зарегистрирована перфорация из 28 пациентов и 369 пациентов, у которых не было перфорации. Рассматривалось 77 признаков. Так как число признаков существенно превышает объем одной из групп, то при построении решающих правил классификации будет наблюдаться эффект переобучения. Поэтому проводился отбор информативных признаков с использованием точного критерия Фишера [1] и метода LASSO [2].

Выбор критерия Фишера обусловлен фактом наличия ожидаемой частоты значений хотя бы в одной из ячеек таблицы 2×2 менее 10%. Такая ситуация имеет место для группы меньшего объема. Таким образом, для всех бинарных признаков сравнивались частоты групп с возникшей и не возникшей перфорацией, и в число информативных признаков включались те, которые имели существенное различие на уровне значимости не менее чем 0.05.

Регрессия LASSO направлена на выявление переменных и соответствующих им коэффициентов регрессии, которые приводят к модели минимизирующей ошибку предсказания. Это достигается путем наложения ограничения на параметры модели, которые «сжимают» коэффициенты регрессии до нуля, то есть заставляют сумму абсолютных значений коэффициентов регрессии быть меньше фиксированного значения λ. Переменные с нулевым коэффициентом регрессии после сжатия исключаются из модели.

Выбор λ производился с использованием 3-кратной перекрестной проверки. Преимущество этой техники заключается в том, что она уменьшает переобучение без ограничения подмножества набора данных.

Так как присутствует дисбаланс классов обучающей выборки, то в модели использовались сбалансированные веса для классов, то есть вес каждого класса обратно пропорционален количеству экземпляров класса в выборке.

Следует отметить, что применение точного критерия Фишера и метода LASSO привело к различным наборам информативных признаков.

Литература

1. Афифи А., Эйзен С. *Статистический анализ. Подход с использованием ЭВМ.* М: Мир. 1982. Т. 488.

2. Ranstam J., Cook J.A. *LASSO regression* // Journal of British Surgery. 2018. V. 105. №. 10. P. 1348-1348.

ОЦЕНКА СВЕРХУ ДЛЯ БИНОМИАЛЬНЫХ КОЭФФИЦИЕНТОВ В ФОРМЕ МУАВРА — ЛАПЛАСА

С.В. Агиевич

Белгосуниверситет, НИИ прикладных проблем математики и информатики
Независимости 4, 220030 Минск, Беларусь agievich@bsu.by

Теорема Муавра — Лапласа применительно к симметричному биномиальному распределению может быть записана в виде следующей оценки биномиальных коэффициентов:

$$\binom{n}{k} = \frac{2^n}{\sqrt{\pi n/2}} \exp\left(-\frac{2(k-n/2)^2}{n}\right)(1+O(1/\sqrt{n})).$$

Оценка справедлива при $n \to \infty$ и $|k-n/2| = O(\sqrt{n})$, т. е. в так называемой центральной области изменения параметров.

То, что оценка носит асимптотический характер и справедлива только в центральной области, затрудняет ее применение в ряде случаев. Известны неасимптотические оценки, которые справедливы в более широких областях, например,

$$\left(\frac{n}{k}\right)^k \le \binom{n}{k} \le \left(\frac{en}{k}\right)^k, \quad 1 \le k \le n,$$

или, обозначив $H_2(x) = -x\log_2 x - (1-x)\log_2(1-x)$,

$$\frac{2^{nH_2(k/n)}}{\sqrt{8k(1-k/n)}} \le \binom{n}{k} \le \frac{2^{nH_2(k/n)}}{\sqrt{2\pi k(1-k/n)}}, \quad 1 \le k \le n-1$$

(см. соответственно [2] и [1; глава 10, лемма 7]). Однако эти оценки либо недостаточно точны, либо их форма оказывается недостаточно удобной.

Мы нашли оценку сверху для биномиальных коэффициентов, в которой сохраняется форма Муавра — Лапласа и которая справедлива во всей области изменения параметров.

Теорема. Для натурального n и $k \in \{0,1,\ldots,n\}$ справедлива оценка

$$\binom{n}{k} \le \frac{2^n}{\sqrt{\pi n/2}} \exp\left(-\frac{2(k-n/2)^2}{n} + \frac{23}{18n}\right).$$

При построении оценки мы следовали подходу работы [3], в свою очередь основанному на ряде предшествующих работ.