



ОЦЕНКА КАЧЕСТВА БИНАРНЫХ КЛАССИФИКАТОРОВ С ПОМОЩЬЮ AUC-ROC

М. В. Галимова, К. В. Золотухин

Санкт-Петербургский государственный университет аэрокосмического приборостроения

Рассмотрена методика оценки качества бинарных классификаторов, представленных в виде модели бинарной логистической регрессии, с использованием метрики AUC-ROC (Area Under the Receiver Operating Characteristic Curve).

Предложен алгоритм вычисления оптимального порога отсечения, а также проведена оценка статистической значимости модели на основе AUC-ROC. Приведены результаты практического использования методики. Применён более точный и информативный подход к оценке качества модели, учитывающий весь спектр возможных порогов классификации, и позволивший оценить общую способность модели разделять классы.

Ключевые слова: бинарный классификатор, ROC-кривая, площадь под ROC-кривой, классификация; логистическая регрессия; AUC-ROC.

Для цитирования:

Галимова, М. В. Оценка качества бинарных классификаторов с помощью AUC-ROC / М. В. Галимова, К. В. Золотухин // Системный анализ и логистика. – 2024. – № 3(41). – с. 10-16. DOI: 10.31799/2077-5687-2024-3-10-16.

QUALITY ASSESSMENT OF BINARY CLASSIFIERS USING AUC-ROC

M. V. Galimova, K. V. Zolotukhin

St. Petersburg State University of Aerospace Instrumentation

The methodology of quality assessment of binary classifiers represented as a binary logistic regression model using the AUC-ROC (Area Under the Receiver Operating Characteristic Curve) metric is considered.

An algorithm for calculating the optimal cutoff threshold is proposed, and the statistical significance of the AUC-ROC-based model is evaluated. The results of practical use of the methodology are presented. A more accurate and informative approach to assessing the quality of the model, which takes into account the whole range of possible classification thresholds, was applied and allowed to evaluate the overall ability of the model to separate classes.

Keywords: binary classifier, ROC curve, area under the ROC curve, classification, logistic regression, AUC-ROC.

For citation:

Galimova, M. V. Quality assessment of binary classifiers using AUC-ROC / M. V. Galimova, K. V. Zolotukhin // System analysis and logistics. – 2024. – № 3(41). – p. 10-16. DOI: 10.31799/2077-5687-2024-3-10-16.

Введение

Регрессионный анализ представляет собой значимый инструмент для определения связи между зависимой переменной и независимыми факторами, предсказания значений зависимой переменной на основе предикторов и оценки их вклада в изменчивость выходных данных. Однако классический метод наименьших квадратов, используемый в классическом регрессионном анализе, требует выполнения множества условий, что может ограничивать его применение на практике.

Логистическая регрессия является дополнением к классической линейной регрессии и находит широкое применение, когда имеется бинарная зависимая переменная, такая как наличие или отсутствие определенных транспортных задержек, и набор независимых переменных, измеряемых в различных масштабах. Поскольку зависимая переменная бинарна, предположения, лежащие в основе классической линейной регрессии, не выполняются, и не всегда можно интерпретировать предсказанные значения, отличные от нуля или единицы. Вместо этого логистическая регрессия прогнозирует вероятность P того, что входная переменная относится к определенному классу [1].

В последние годы модели множественной логистической регрессии становятся все более



популярными в различных областях, таких как медицина, финансы, маркетинг, социология, управление и принятие решений.

Методы множественной логистической регрессии в бинарной классификации

При работе с задачами распознавания образов с применением методов обучения с учителем, часто встречается сценарий, когда результат эксперимента представлен в двоичной форме (1 – наличие признака, 0 – отсутствие признака). В таких случаях для решения задачи может применяться метод логистической регрессии, известный как логит-анализ. Формула этой модели может быть представлена следующим образом [2]:

$$P_j(B) = \frac{1}{1 + e^{-\left(\sum_{i=1}^m b_i X_{ji}\right)}}, \quad (1)$$

где $P_j(B)$, $j = 1, 2, \dots, n$ – выход модели; $B = \{b_i\}$, $i = 1, 2, \dots, m$ – вектор-столбец весовых коэффициентов; $X_j = \{b_i\}$, $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$ – вектор-строка параметров объекта j , измеренных в эксперименте; $X_j B$, $j = 1, 2, \dots, n$ – множественная линейная регрессия; m – количество измеряемых параметров объекта; n – количество объектов в обучающей выборке.

Значение $P_j(\cdot)$ интерпретируется как вероятность получения логитом значения 1 при подстановке в уравнение определенного вектора X_j , измеренного в эксперименте.

Путем максимизации логарифма функции максимального правдоподобия (ФМП) [3] могут быть вычислены оптимальные значения весовых коэффициентов:

$$L(B) = \sum_{j=1}^n \left[Y_j \ln(P_j(B)) + (1 - Y_j) \ln(1 - P_j(B)) \right], \quad (2)$$

где Y_j , $j = 1, 2, \dots, n$ – результат эксперимента, соответствующий измеренному вектору параметров X_j .

Для решения задачи используется метод Ньютона-Рафсона [4], который представляет собой итерационную схему:

$$B^{(k+1)} = B^k - \left(H(B^k) \right)^{-1} g(B^k), \quad (3)$$

где k – номер итерации; $H(\cdot)$ – матрица Гессе ФМП; $g(\cdot)$ – градиент ФМП.

В силу доказанной аналитически сходимости итерационной схемы, начальные значения весовых коэффициентов могут быть выбраны произвольно. Для удобства и точности расчетов часто применяется нулевой вектор.

Замечание: если в модели логистической регрессии присутствует свободный член, в массиве исходных данных добавляется столбец, состоящий из единиц, а при распознавании объектов необходимо соответствующим образом модифицировать вектор параметров [5].

Оценка производительности модели бинарного классификатора

Важно использовать различные показатели для оценки производительности модели бинарного классификатора. В данном контексте часто используется таблица 2x2, представляющая собой естественный способ оценки классификации для бинарных откликов. Пример представлен в табл. 1.



Таблица 1 – Таблица сопряженности

Предсказанные моделью результаты	Фактические результаты	
	Фактически положительный исход	Фактически отрицательный исход
Положительный исход	TP	FP
Отрицательный исход	FN	TN

Примечания: *T* (True, истинно) указывает на правильно предсказанный класс, в то время как *F* (False, ложно) означает ошибочное предсказание, тогда как *P* (Positive, положительный) и *N* (Negative, отрицательный) обозначают, что классификатор относит объект к положительному и отрицательному классам соответственно.

1. TP – верно классифицированные положительные примеры;
2. TN – верно классифицированные отрицательные примеры;
3. FN – положительные примеры, классифицированные как отрицательные (ошибка I рода);
4. FP – отрицательные примеры, классифицированные как положительные (ошибка II рода) [6].

Например, в контексте транспортной системы положительным исходом может быть определение задержки в пути, а отрицательным – ее отсутствие.

На основе таких показателей строится ROC-кривая (Receiver Operating Characteristic Curve). Она описывает соотношение между долей верно классифицированных положительных результатов (часто называемой чувствительностью Se) и долей неверно классифицированных отрицательных результатов (или количеством ложноположительных результатов, обозначаемым как $1 - Sp$, где Sp - специфичность теста, т.е. доля верно классифицированных тестом отрицательных случаев), при изменении порога отсечения (cut-off).

Каждая точка на кривой соответствует определенному порогу классификации, который определяет отношение между чувствительностью и специфичностью.

Чувствительность показывает долю истинно положительных случаев:

$$Se = \frac{TP}{TP + FN} \quad (4)$$

Специфичность показывает долю истинно отрицательных случаев:

$$Sp = \frac{TN}{TN + FP} \quad (5)$$

ROC-кривая строится на основе различных значений чувствительности и специфичности, представленных в процентах. Критерий Юдена [7], максимизирующий сумму чувствительности и специфичности $Se + Sp$, используется для выбора оптимального порога классификации.

Алгоритм построения ROC-кривой может быть представлен в виде последовательности действий:

1. Подготовка данных: объединить выборки в один массив диагностических параметров и упорядочить его по убыванию.
2. Выбор порогов: использовать значения массива диагностических параметров в качестве порогов классификации.
3. Построение таблиц 2x2: для каждого порога составить таблицу 2x2,
4. Вычисление чувствительности и специфичности: для каждой таблицы подсчитать чувствительность (Se) и специфичность ($1 - Sp$).



5. Построение ROC-кривой: нанести на график координаты, где ось X представляет собой $1 - Sp$, а ось Y – Se . Каждая точка на ROC-кривой соответствует определенному порогу классификации.
6. Вычисление площади под ROC-кривой (AUC): для оценки качества классификатора вычислить площадь под ROC-кривой с помощью метода трапеций или других численных методов интегрирования.

Этот алгоритм может быть эффективно выполнен в цикле для каждого порога классификации.

Оценка качества диагностического метода может быть выполнена с использованием площади под ROC-кривой, которая обозначается как AUC (Area Under Curve). Это численное значение оценивает способность классификатора правильно различать между положительными и отрицательными случаями при различных порогах классификации. AUC вычисляется с помощью формулы трапеций:

$$AUC = \sum_{i=1}^{n-1} \left(\frac{Se_i + Se_{i+1}}{2} \right) \times (Sp_i - Sp_{i+1}), \quad (6)$$

где Se и Sp представляют собой значения чувствительности и специфичности, соответственно, при различных порогах классификации. Значение AUC может находиться в диапазоне от 0,5 до 1. Приближение к 0,5 указывает на недискриминационную способность теста, в то время как значение 1 указывает на идеальную точность. Таким образом, чем ближе AUC к 1, тем выше прогностическая способность модели. Часто используется экспертная шкала для интерпретации значений AUC (см. таблицу 2), которая помогает оценить качество модели [8].

Таблица 2 – Экспертная шкала оценок AUC

Интервал AUC	Качество модели
0,9–1,0	Отличное
0,8–0,9	Очень хорошее
0,7–0,8	Хорошее
0,6–0,7	Среднее
0,5–0,6	Неудовлетворительное

Стандартная ошибка оценки AUC также может быть вычислена для оценки достоверности полученного значения AUC . Для этого используется следующая формула [4]:

$$SE(\widehat{AUC}) = \sqrt{\frac{2 \times (\widehat{AUC}) \times (1 - (\widehat{AUC})) + (\widehat{Q}_1 \times \widehat{Q}_2)}{n \times m}}, \quad (7)$$

где \widehat{Q}_1 и \widehat{Q}_2 обозначают сумму квадратов частот диагностических параметров.

Для сравнения двух ROC-кривых по их AUC используется статистика Z , которая имеет асимптотическое нормальное распределение. Это позволяет объективно оценить статистическую значимость различий между двумя классификаторами.

Наконец, для вычисления доверительного интервала оцениваемой AUC используется формула с использованием обратной функции стандартного нормального распределения и



заданного доверительного уровня.

Для вычисления AUC-ROC используются различные методы, включая метод трапеций [9], метод Симпсона или другие численные методы интегрирования.

Важно отметить, что ROC-кривая обычно имеет ступенчатый вид как для выборок, так и для распределений. Однако при использовании больших выборок и очень маленьких значений шага порога отсечения она может казаться более гладкой на графике. В случае, когда отсутствует порог отсечения (модель предоставляет бинарный ответ), кривая формируется из трех точек: $(0; 0)$, $(Se; 1 - Sp)$, $(1; 1)$.

Результаты применения предлагаемого подхода

Рассмотрим модель бинарного классификатора, которая предсказывает, вероятность увеличения времени в пути для улучшения управления транспортной системой на основе ряда признаков. Для оценки качества этой модели будет использована AUC-ROC.

Для оценки производительности модели классификации на практике использованы реальные данные о транспортной загруженности. Данные охватывают различные характеристики, такие как время начала и завершения поездки, тип транспорта, степень загруженности маршрута и другие факторы, влияющие на возможные задержки.

Для анализа данных и оценки качества модели был написан код на Python, используя библиотеки для машинного обучения, такие как scikit-learn [10]. Код включал в себя загрузку данных, их предварительную обработку, разделение на тренировочный и тестовый наборы, обучение модели логистической регрессии и оценку ее производительности с помощью метрик, включая AUC-ROC.

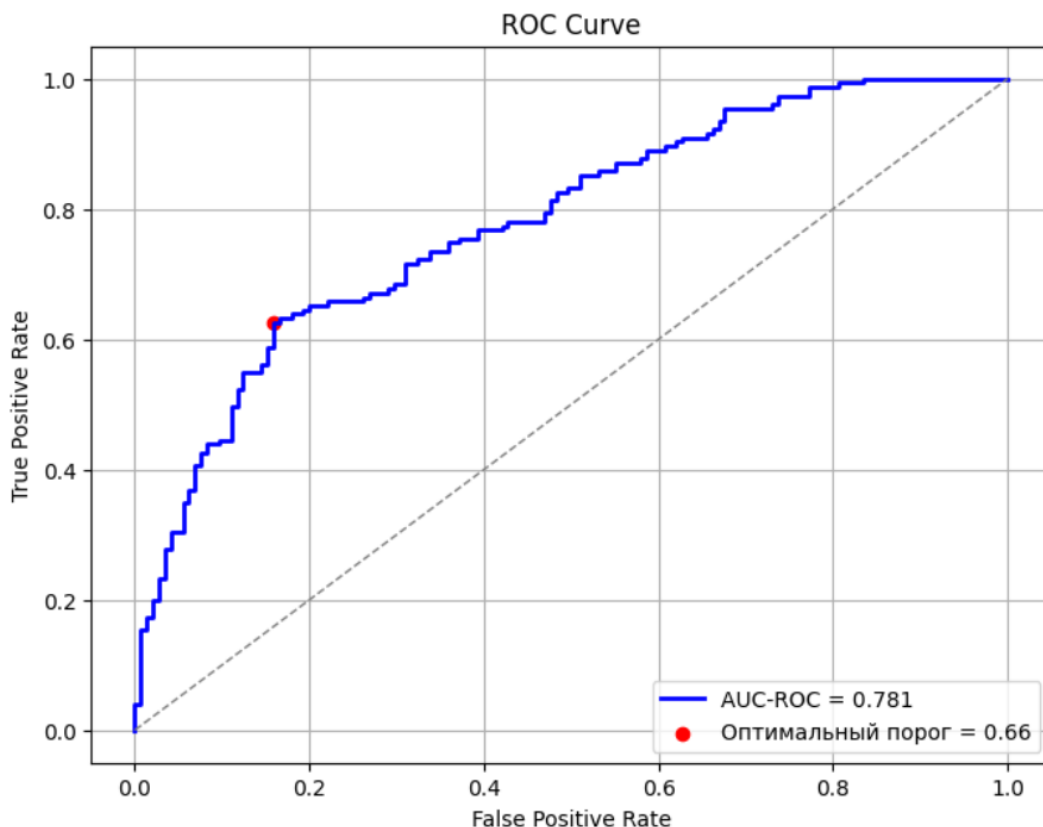


Рис. 1. Результаты ROC-анализа данных

На представленном графике (Рисунок 1) ось абсцисс (X) отображает долю ложноположительных результатов (False Positive Rate), а ось ординат (Y) отображает долю истинно положительных результатов (True Positive Rate). Каждая точка на кривой



соответствует различному порогу классификации.

Оптимальное пороговое значение, равное 0,66, найденное на графике, отображается точкой красного цвета. Это значение является оптимальной точкой компромисса между чувствительностью и специфичностью модели.

Чувствительность модели составляет 0,626, что означает, что модель правильно классифицирует 62,6% истинно положительных случаев. Специфичность модели равна 0,841, что показывает, что модель правильно классифицирует 84,1% истинно отрицательных случаев.

Таким образом, график ROC-кривой и связанные с ним метрики позволяют нам оценить качество модели бинарной классификации и определить оптимальные параметры для конкретной задачи.

Заключение

В данной статье рассмотрен важный аспект оценки моделей машинного обучения - показатель AUC-ROC. На основании всех описанных теоретических выкладок реализована модель логистической регрессии в Python с использованием библиотеки scikit-learn, вычислен показатель AUC-ROC.

Результаты исследования демонстрируют, что модель обладает хорошей прогностической способностью, что делает ее полезной для прогнозирования для прогнозирования времени в пути или вероятности задержки. Показатель AUC-ROC, равный 0.781, указывает на высокую дискриминирующую способность модели, что подтверждает ее пригодность для применения.

В целом, использование AUC-ROC в оценке моделей бинарной классификации является важным инструментом для анализа и сравнения производительности моделей. Правильная интерпретация ROC-кривой и ее метрик позволяет принимать обоснованные решения при выборе модели для конкретной задачи и оценке ее качества.

СПИСОК ЛИТЕРАТУРЫ

1. Богданов, Л. Ю. Оценка эффективности бинарных классификаторов на основе логистической регрессии методом ROC-анализа / Л. Ю. Богданов // Вестник Саратовского государственного технического университета. – 2010. – Т. 4, № 2(50). – С. 92-97.
2. Hosmer, D. W. Applied logistic regression / D. W. Hosmer, S. Lemeshow – John Wiley & Sons, 2000. – 375 p.
3. Цыплаков, А. А. Некоторые эконометрические методы. Метод максимального правдоподобия в эконометрии: методическое пособие / А. А. Цыплаков – Новосибирск: ЭФ НГУ, 1997. – 129 с.
4. Гайдышев, И. П. Оценка качества бинарных классификаторов / И. П. Гайдышев // Вестник Омского университета. – 2016. – № 1(79). – С. 14-17.
5. Макаров, Р. И. Анализ и синтез информационных систем: учеб. пособие / Р. И. Макаров, Е. Р. Хорошева; Владим. гос. ун-т им. А. Г. и Н. Г. Столетовых. – Владимир: Изд-во ВлГУ, 2019. – 251 с.
6. Андронов С. А. Введение в машинное обучение в среде Google Colab: лабораторный практикум / С. А. Андронов. – СПб: ГУАП, 2024. – 12 с.
7. Корнеев, А. А. Возможности ROC-анализа для категоризации переменных в модели прогноза исхода оперативного лечения у пациентов с болезнью Меньера / А. А. Корнеев, С. В. Лиленко, А. С. Лиленко [и др.] // Российская оториноларингология. – 2018. – № 4(95). – С. 62-68.
8. Ковалев А. А. Оценка качества бинарного классификатора в научных исследованиях / А. А. Ковалев, Б. К. Кузнецов, А. А. Ядченко, В. А. Игнатенко // Проблемы здоровья



- и экологии. – 2020. – № 4(66). – С. 105-113.
9. Бостанова, Ф. А. Приближенное вычисление интегралов методом Симпсона и методом трапеций / Ф. А. Бостанова, О. Б. Овезгельдиева // Современные проблемы математического образования: III Всероссийской научно-практической конференции. – Карачаевск: Карачаево-Черкесский государственный университет им. У.Д. Алиева, 2018. – С. 59-69.
 10. Саммерфилд, М. Программирование на Python 3: подробное руководство / М. Саммерфилд; Марк Саммерфилд; [пер. с англ. А. Киселева]. – Санкт-Петербург: Символ, 2009. – 607 с.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Галимова Мария Владимировна

Студент

Санкт-Петербургский государственный университет аэрокосмического приборостроения

190000, Санкт-Петербург, ул. Большая Морская, д. 67, лит. А

E-mail: kivy123@mail.ru

Золотухин Кирилл Вячеславович

Ассистент

Санкт-Петербургский государственный университет аэрокосмического приборостроения

190000, Санкт-Петербург, ул. Большая Морская, д. 67, лит. А

E-mail: zoloto.kv@mail.ru

INFORMATION ABOUT THE AUTHORS

Galimova Mariya Vladimirovna

Student

Saint-Petersburg State University of Aerospace Instrumentation

67, Bolshaya Morskaya str., Saint-Petersburg, 190000, Russia

E-mail: kivy123@mail.ru

Zolotukhin Kirill Vyacheslavovich

Assistant

Saint-Petersburg State University of Aerospace Instrumentation

67, Bolshaya Morskaya str., Saint-Petersburg, 190000, Russia

E-mail: zoloto.kv@mail.ru