

Матрица ошибок (Error matrix)

Синонимы: Матрица неточностей, Confusion matrix

Разделы: [Метрики](#)

Матрица ошибок представляет собой способ визуализации для оценки качества классификаторов. Обычно используется в машинном обучении с учителем как для бинарной, так и для многоклассовой классификации.

В случае бинарной классификации матрица ошибок представляет собой таблицу, состоящую из двух строк и двух столбцов, при этом строки соответствуют фактическим классам, а столбцы — предсказанным.

В процессе обучения классификатор делает предсказания на обучающих примерах, для которых метка класса известна. При этом он допускает ошибки I и II рода. Если предсказанный класс соответствует фактическому, то исход классификации считается **истинным**, а в противном случае — **ложным**. Примеры положительного и отрицательного классов, для которых исход предсказания истинный, называются **истинноположительными** (true-positive, TP) и **истинноотрицательными** (true-negative, TN) соответственно. Очевидно, что это правильно классифицированные примеры.

Примеры положительного и отрицательного классов, для которых исход предсказания является ложным, называются **ложноположительными** (false-positive, FP) и **ложноотрицательными** (false-negative, FN) соответственно. Считается, что на этих примерах классификатор допустил ошибку.

Тогда результаты работы бинарного классификатора могут быть представлены в матрице ошибок следующим образом.

P+N	Положительный (предсказано)	Отрицательный (предсказано)
Положительный (факт)	TP	FP
Отрицательный (факт)	FN	TN

Если ячейка матрицы ошибок расположена на пересечении строки и столбца для одного и того же класса (т.е. элемент на главной диагонали), то она соответствует истинным классификациям, и в ней ставится число правильно отсортированных примеров для

соответствующего класса. Если столбец и строка, на пересечении которых расположена ячейка, относятся к разным классам, то в ней окажется число ошибочно определенных примеров.

По результатам, представленным в матрице ошибок, могут вычисляться меры качества модели бинарной классификации.

Меткость (Accuracy, ACC или Overall classification rate, OCR) — доля правильно классифицированных примеров:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN}.$$

Точность (Precision) — отношение числа истинноположительных классификаций к общему числу положительных классификаций. Данная величина также известна как positive predictive value (PPV) или положительное прогностическое значение:

$$Pr = PPV = \frac{TP}{TP+FP}.$$

Полнота (Recall) — доля истинноположительных примеров (TPR — true positive rate). Упоминается еще как чувствительность. Определяется как число истинноположительных классификаций относительно общего числа положительных примеров:

$$Re = TPR = \frac{TP}{TP+FN}.$$

Полноту можно рассматривать как способность бинарного классификатора обнаруживать определенный класс.

Специфичность — доля истинноотрицательных (True Negative Rate — TNR) классификаций в общем числе отрицательных классификаций:

$$Sp = TNR = \frac{TN}{TN+FP}.$$

Данная величина показывает, насколько хорошо модель классифицирует отрицательные примеры.

Точностью отрицательного прогноза — доля верно классифицированных отрицательных примеров (Negative predictive value — NPV) от общего числа примеров классифицированных как отрицательные:

$$NPV = \frac{TN}{TN+FN}.$$

False positive rate (FPR или Fall-out) — доля неверно классифицированных положительных примеров от общего количества отрицательных:

$$FPR = 1 - TNR = \frac{FP}{FP+TN}.$$

False negative rate (FNR) — доля неверно классифицированных отрицательных примеров от общего количества положительных примеров:

$$FNR = 1 - TPR = \frac{FN}{FN+TP}.$$

F1-мера объединяет в себе информацию о точности и полноте, поэтому позволяет находить баланс между ними:

$$F1 = \frac{2 \cdot PPV \cdot TPR}{PPV + TPR} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}.$$

Матрица ошибок может применяться и для многоклассовой классификации. В этом случае число строк и столбцов в ней будет равно числу классов. Однако в этом случае понятия отрицательного и положительного классов в том виде, в котором они были сформулированы для бинарной модели, не работают. Поэтому в ячейках матрицы ставятся не величины TP, FP, TN и FN, а количества классифицированных соответствующим образом примеров.

	Класс 1 (предсказано)	Класс 2 (предсказано)	Класс 3 (предсказано)
Класс 1 (факт)	20		15
Класс 2 (факт)		10	
Класс 3 (факт)	5		50

Числа, стоящие в ячейках на пересечении строк и столбцов для одноименных классов (когда предсказанный класс соответствует фактическому) определяют число правильно классифицированных примеров. Очевидно, что такие ячейки будут располагаться на главной диагонали матрицы ([20, 10, 50]). Ячейки, расположенные вне нее будут содержать количества ошибочно классифицированных примеров.

Если некоторые ячейки матрицы остались пустыми, это указывает на отсутствие ошибок модели в соответствующих комбинациях фактических и предсказанных классов. В них можно внести нулевые значения.

Матрица ошибок для многоклассовой классификации также позволяет наглядно представлять результаты работы классификатора и оценивать его качество. Например, по ней просто вычислить точность модели как отношение суммы чисел по главной диагонали матрицы к общему числу элементов в ней, или ошибку, как отношение суммы чисел вне главной диагонали к общему числу элементов.

Узнать подробнее, какие меры качества моделей бинарной классификации вычисляются с использованием матрицы ошибок, можно в статье [«Метрики качества моделей бинарной классификации»](#).