

Оценка на модели за класификация (Confusion matrix и Classification report)

Два важни критерия за оценка качествата на даден модел и прогнозната му сила са confusion matrix и classification report.

Ще опишем тяхната интерпретация в случая когато е известно множество от наблюдения, които са разделени в два класа (Class 1, Class 2). Наблюденията от Class 1 наричаме положителни, докато от Class 2 наричаме отрицателни. Построява се модел който моделира делението на два класа и прогнозира какво е разпределението по класове на наблюденията. В резултат са известни броя на наблюденията по класове, и също колко от как са разпознати от построенния модел. Тези числови стойности се записват в 2x2 матрица, която се нарича confusion matrix.

	Прогнозиран Class 1	Прогнозиран Class 2
Реален Class 1	TP	FN
Реален Class 2	FP	TN

Таблица 3.1. Confusion matrix.

<https://www.geeksforgeeks.org/confusion-matrix-machine-learning/> (януари, 2022)

Елементите на confusion matrix имат следното значение:

True Positive (TP): Брой на положителните реални наблюдения, които са прогнозирани като положителни наблюдения.

False Negative (FN): Брой на положителните реални наблюдения, които са прогнозирани като принадлежащи към класа на отрицателните наблюдения.

True Negative (TN) : Брой на отрицателни реални наблюдения, които са прогнозирани като отрицателни наблюдения.

False Positive (FP) : Брой на отрицателните реални наблюдения, които са прогнозирани като принадлежащи към класа на положителните наблюдения.

Елементите на confusion matrix дефинират няколко показателя, чрез които се оценява ефективността на построенния модел. Тези показатели са обединени в един доклад, наречен classification report. Събира в себе си параметрите Accuracy, **Recall**, **Precision** и F-score (F1-score). Те се дефинират както следва.

$$\text{Accuracy} = (TP+TN) / (TP+TN+ FP+FN)$$

Accuracy (Точност) = Score коефициент при класификационни задачи

Recall (Чувствителност). Дефинира се чрез формулите

$$\text{Recall (Class 1)} = TP / (TP+FN) ,$$

$$\text{Recall (Class 2)} = TN / (TN+FP) .$$

В допълнение, в научната литература двата параметъра Recall (Class 1) и Recall (Class 2) се използват и под имената Sensitivity и Specificity (Tripathi et al., 2018).

Висока стойност на Recall се интерпретира чрез факта, че съответният клас се разпознава добре от модела, т.е. броя на „сгрешените“ *реални* наблюдения е малък. За Class 1 този брой е числото FN, докато за Class 2 този брой е FP.

Параметърът Precision се дефинира чрез

Precision (Class 1) = $TP / (TP + FP)$,

Precision (Class 2) = $TN / (TN + FN)$.

Висока стойност на **Precision** се интерпретира чрез факта, че съответният клас се разпознава добре от модела, т.е. броя на „сгрешените“ *прогнозиранни* наблюдения е малък. За прогнозирания Class 1 (първа колона) това е числото FP, докато за прогнозирания Class 2 това е числото FN.

Често срещани случаи в практиката са следните:

Висока стойност за Recall и ниска стойност за Precision. Това означава, че повечето положителни наблюдения са правилно разпознати (нисък FN), но има много лъжливи положителни наблюдения (висок FP).

Ниска стойност за Recall и висока стойност за Precision. Това показва, че ни липсват много положителни наблюдения (висок FN), но тези, които прогнозираме като положителни, наистина са положителни (нисък FP).

Следващият параметър е F-мярката или F-score. Той се дефинира като

$F - score = 2 * Recall * Precision / (Recall + Precision)$.

F-мярката съчетава в себе си двете величини - Recall и Precision. Пресмята хармоничното средно на двете величини. F-мярката е по-близко до по-малката стойност от двете величини Recall и Precision.

Пример . Получени са следните резултати, а оценката на построения модел върху тестовото подмножество е следният:

confusion matrix for the TEST set

[[444 56]

[117 151]]

test set score : 0.8051948051948052

	precision	recall	f1-score	support
0	0.79	0.89	0.84	500
1	0.73	0.56	0.64	268
macro avg	0.76	0.73	0.74	768
weighted avg	0.77	0.77	0.77	768

Точността на построенния модел, изразена чрез score коефициента е почти 81% (score = 0.8051), докато стойността на recall за втория клас наблюдения е 56% (0.56), което означава че моделът не разпознава добре наблюденията от клас 1. Този факт се вижда и от втория ред на confusion matrix, т.е. 151 наблюдения от клас 1 са разпознати като такива, 117 са сгрешени от модела.

Пример от Python и Формули

[[444 56] [117 151]]				TP FN FP TN
	precision	recall	f1-score	Precision (Class 1) = $TP / (TP+FP) = 444 / (444+117) = 0.79$ Precision (Class 2) = $TN / (TN+FN) = 151 / (151+56) = 0.73$ *****
0	0.79	0.89	0.84	Recall (Class 1) = $TP / (TP+FN) = 444 / (444+56) = 0.88 =$ =Sensitivity Recall (Class 2) = $TN / (TN+FP) = 151 / (151+117) = 0.56 =$ =Specificity
1	0.73	0.56	0.64	
Confusion matrix от (Novriadi,2021): [[43 1] [11 2]]				Sensitivity = Recall (Class 1) = $TP / (TP+FN) = 43 / (43+1) = 0.98$ Specificity = Recall (Class 2) = $TN / (TN+FP) = 2 / (2+11) = 0.15$

Литература

- [1] Tripathi, D., Edla, D. R., Kuppili, V., Bablani, A., Dharavath, R., Credit Scoring based on Weighted Voting and Cluster based Feature Selection. Procedia Computer Science 132(2018) 22-31 (2018).
- [2] A.S. Novriadi, W. Sutarman, Sawaluddin, Dataset Weighting Features Using Gain Ratio To Improve Method Accuracy Naïve Bayesian Classification, IOP Conf. Series: Earth and Environmental Science , 2021, 748, N 012034.
- [3] Иван Иванов, Количествени методи за оценка на модели