Метрики классификации и регрессии • Metrics

Сегодня

Эльбрус буткемп

- как измерить, что алгоритм хорошо решает задачу?
- как сравнить несколько алгоритмов между собой и выбрать лучший?
- как выбрать правильный способ оценки и от чего это зависит?

Метрики классификации

Эльбрус буткемп

		Actual Class	
		Cat	Non-Cat
Predicted Class	Cat	90 *TP*	60 *FP*
	Non-Cat	10 *FN*	940 *TN*

- *True Positive / **TP** / Истинно-положительное*: классификатор отнес объект к классу `cat` и объект действительно относится к классу `cat`
- *False Positive / **FP** / Ложно-положительное*: классификатор отнес объект к классу `cat`, но на самом деле объект не относится к этому классу
- *False Negative / **FN** / Ложно-отрицательное*: классификатор отнес объект к классу `non cat`, но на самом деле объект относится к классу `cat`
- *True Negative / **TN** / Истинно-отрицательное*: классификатор отнес объект к классу `non cat`, и объект действительно относится к этому классу

Метрики классификации

Эльбрус буткемп

Accuracy

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

> Если выборка несбалансирована, то эта метрика ничего не покажет

Метрики классификации

Эльбрус буткемп

Precision

$$precision = \frac{TP}{TP + FP}$$

Метрики классификации

Эльбрус буткемп

Recall

$$recall = \frac{TP}{TP + FN}$$

Метрики классификации

Эльбрус буткемп

F1-score

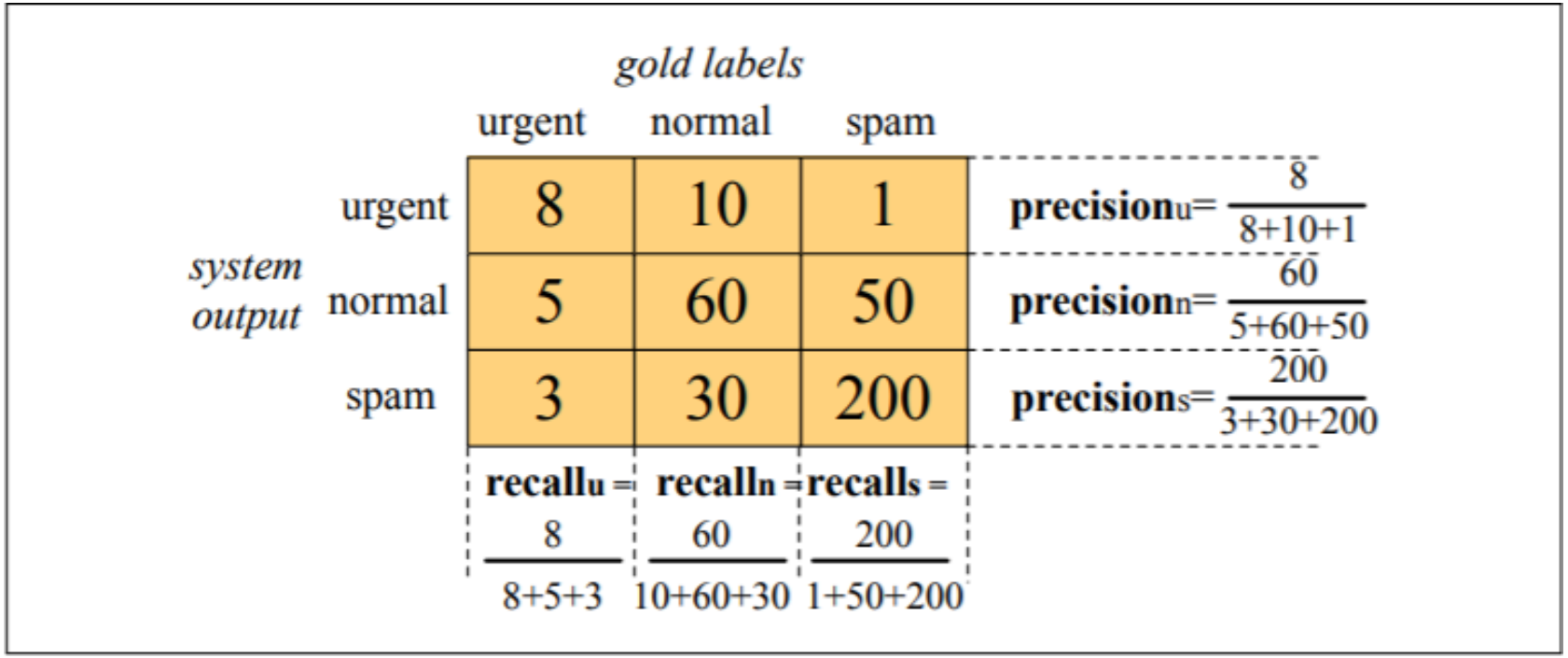
$$f1 = 2 \times \frac{precision \times recall}{precision + recall}$$

> гармоническое среднее между точностью и полнотой

Метрики классификации

Эльбрус буткемп

- Часто метрики *precision*, *recall* и *f1* можно встретить с приставкой *micro* или *macro*
- Посмотрим на примере *precision*



📝 Micro Average vs Macro average Performance in a Multiclass classification setting

Метрики классификации

Эльбрус буткемп

Class 1: Urgent

	true urgent	true not
system urgent	8	11
system not	8	340

$$\text{precision} = \frac{8}{8+11} = .42$$

Class 2: Normal

	true normal	true not
system normal	60	55
system not	40	212

$$\text{precision} = \frac{60}{60+55} = .52$$

Class 3: Spam

	true spam	true not
system spam	200	33
system not	51	83

$$\text{precision} = \frac{200}{200+33} = .86$$

$$\text{macroaverage precision} = \frac{.42+.52+.86}{3} = \mathbf{.60}$$

Pooled

	true yes	true no
system yes	268	99
system no	99	635

$$\text{microaverage precision} = \frac{268}{268+99} = \mathbf{.73}$$

📝 Micro Average vs Macro average Performance in a Multiclass classification setting

ROC-AUC

Для вычисления ROC-AUC используются две дополнительные метрики:

- True positive rate: $TPR = \frac{TP}{TP + FN}$ **(a.k.a Recall)**

- False positive rate: $FPR = \frac{FP}{FP + TN}$

Метрики классификации

Эльбрус буткемп

ROC-AUC

id	оценка	класс
1	0.5	0
2	0.1	0
3	0.2	0
4	0.6	1
5	0.2	1
6	0.3	1
7	0.0	0

Табл. 1

id	оценка	класс
4	0.6	1
1	0.5	0
6	0.3	1
3	0.2	0
5	0.2	1
2	0.1	0
7	0.0	0

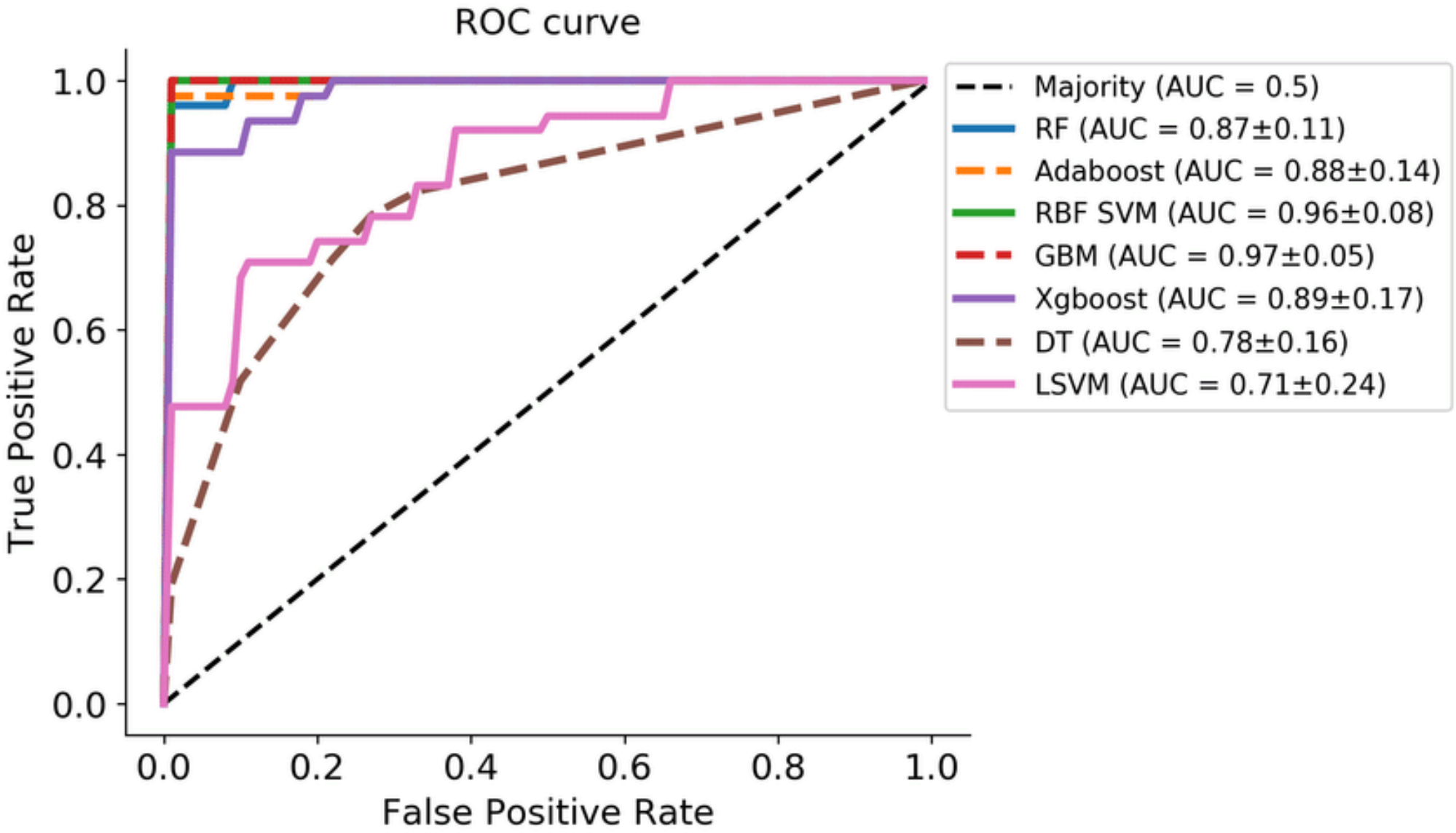
Табл. 2

id	> 0.25	класс
4	1	1
1	1	0
6	1	1
3	0	0
5	0	1
2	0	0
7	0	0

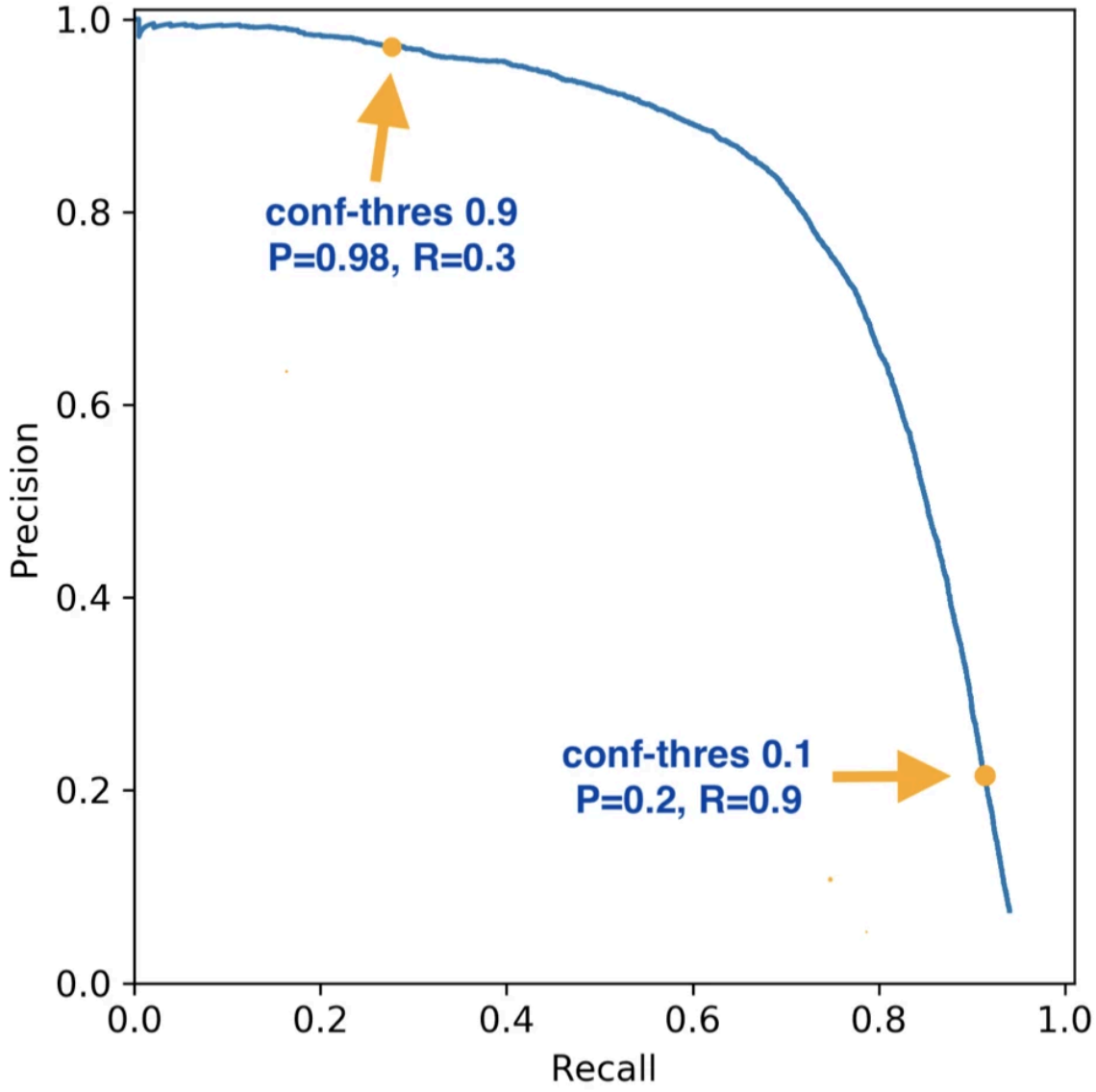
Табл. 3

📝AUC ROC (площадь под кривой ошибок) 📊ROC, AUC, precision, and recall visually explained

ROC-AUC пример графика



PR-AUC


1.0
0.8
0.6
0.4
0.2
0.0
Precision
0.0
0.2
0.4
0.6
0.8
1.0
Recall
conf-thres 0.9
P=0.98, R=0.3
conf-thres 0.1
P=0.2, R=0.9

Метрики регрессии

Эльбрус буткемп

Mean Squared Error

y – истинные значение целевой переменной

$\hat{y}$ – предикт алгоритма

N – число объектов в выборке

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

> MSE применяется в ситуациях, когда нам надо подчеркнуть большие ошибки и выбрать модель, которая дает меньше больших ошибок прогноза.

Метрики регрессии

Эльбрус буткемп

Root Mean Squared Error

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

> Легко интерпретировать, поскольку он имеет те же единицы, что и исходные значения (в отличие от MSE).

Метрики регрессии

Эльбрус буткемп

Mean Absolute Error

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

> Среднеквадратичный функционал сильнее штрафует за большие отклонения по сравнению со среднеабсолютным, и поэтому более чувствителен к выбросам.

Метрики регрессии

Эльбрус буткемп

Mean Absolute Percentage Error

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100$$

> Этот коэффициент можно интерпретировать в долях или процентах. Если получилось, например, что MAPE=11.4%, то это говорит о том, что ошибка составила 11,4% от фактических значений.

Метрики регрессии

Эльбрус буткемп

Symmetric mean absolute percentage error

$$SMAPE = \frac{100\%}{N} \sum_{i=1}^{N} \frac{|\hat{y}_i - y_i|}{|y_i| + |\hat{y}_i|}$$

📝Symmetric mean absolute percentage error

Метрики регрессии

Эльбрус буткемп

R2 score

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

Итоги

Эльбрус буткемп

- выбор метрики **всегда** зависит от задачи
- метрика $\neq$ функция потерь
  - функцию потерь мы минимизируем по параметрам модели
  - метрикой мы измеряем, насколько качественно работает модель
- `sklearn:` Metrics and scoring: quantifying the quality of predictions