Метрики классификации и регрессии • Metrics

### Сегодня



- 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?



		Actual Class				
		Cat	Non-Cat			
Predicted	Cat	90 TP	60 FP			
Class	Non-Cat	10 <i>FN</i>	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 саt, 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



# **Accuracy**

$$ACC = rac{TP + TN}{TP + FP + FN + TN}$$

Если выборка несбалансирована, то эта метрика ничего не покажет



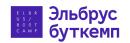
### **Precision**

$$precision = rac{TP}{TP + FP}$$



### Recall

$$recall = rac{TP}{TP + FN}$$



#### F1-score

$$f1 = 2 imes rac{precision imes recall}{precision + recall}$$

гармоническое среднее между точностью и полнотой



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

		old labels		
	urgent	normal	spam	
urgent	8	10	1	$\mathbf{precision_u} = \frac{8}{8+10+1}$
system output normal	5	60	50	$\mathbf{precision}_{n} = \frac{60}{5+60+50}$
spam	3	30	200	$precisions = \frac{200}{3+30+200}$
	recallu =	recalln=	recalls =	
	8	60	200	
	8+5+3	10+60+30	1+50+200	



	Cl	ass 1:	Urgent	t Cl	ass 2:	Norma	al C	lass 3:	Spam	ı	Poo	led	
		true	true		true	true		true	true		true	true	
		urgent	not		normal	not	_	spam	not		yes	no	
	ystem urgent	8	11	system normal	60	55	system spam	200	33	system yes	268	99	
S	ystem not	8	340	system not	40	212	system not	51	83	system no	99	635	
pı	ecision	$n = \frac{8}{8+1}$	$\frac{1}{11}$ = .42	precision =	60+5	- 5 = .52	precision	$=\frac{200}{200+}$	$\frac{0}{33} = .8$	microaverage precision	$e^2 = \frac{2}{268}$	68 3+99 =	= .73
	$\frac{\text{macroaverage}}{\text{precision}} = \frac{.42 + .52 + .86}{3} = .60$												



#### **ROC-AUC**

Для вычисления ROC-AUC используются две дополнительные метрики:

$$ullet$$
 True positive rate:  $TPR=rac{TP}{TP+FN}$  (a.k.a Recall)

$$ullet$$
 False positive rate:  $FPR=rac{FP}{FP+TN}$ 



### **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

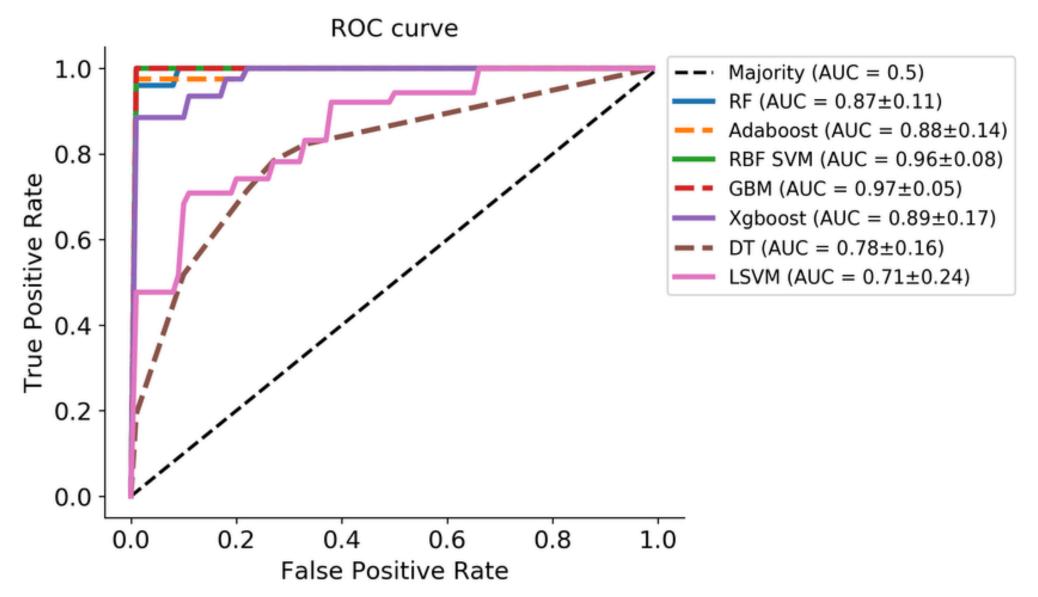
id	оценка	класс
4	0.6	1
1	0.5	0
6	0.3	1
3	0.2	0
5	0.2	1
2	0.1	0
7	0.0	0

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. 1 Табл. 2 Табл. 3

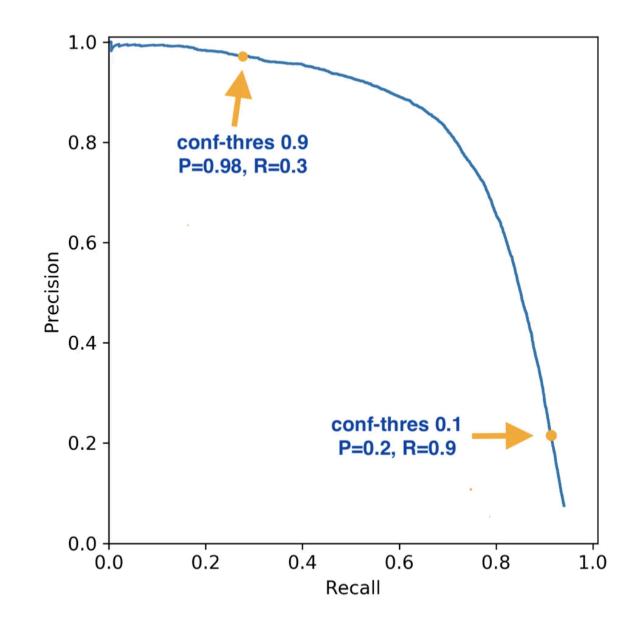
# ROC-AUC пример графика





# **PR-AUC**







### **Mean Squared Error**

y – истинные значение целевой переменной

 $\hat{y}$  – предикт алгоритма

N – число объектов в выборке

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

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



### **Root Mean Squared Error**

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

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



#### **Mean Absolute Error**

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

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



### Mean Absolute Percentage Error

$$MAPE = rac{1}{N} \sum_{i=1}^{N} \left| rac{y_i - \hat{y}_i}{y_i} 
ight| imes 100$$

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



### Symmetric mean absolute percentage error

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



#### R2 score

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

#### Итоги



- 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