

Метрики классификация

$$y_i \in \{0, 1\}$$

$$\hat{p}_i = \frac{1}{1 + \exp(-(w_0 + w_1 x_{i1} + w_2 x_{i2} + \dots + w_k x_{ik}))} \in [0, 1]$$

$$= P(y_i = 1 | x_i)$$

$$\hat{y}_i \in \{0, 1\}$$

$$\hat{p}_i \geq \gamma \Rightarrow \hat{y}_i = 1$$

$$< \hat{y}_i = 0$$

Логистическая
регрессия

Confusion matrix
матрица ошибок

Accuracy (Доля правильных
ответов)

	y	
	0	1
\hat{y}	0	TN FN
	1	FP TP

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP}$$

	y	
	0	1
\hat{y}	0	950 10
	1	0 0
	950 10	

$$\text{Accuracy} = 0.99$$

$$[0, 1]$$

↑
доля самого частотного
класса

	y	
	0	1
\hat{y}	0	TN FN
	1	FP TP

Точность, но
по колонкам

$$\text{Precision} = \frac{TP}{TP + FP}$$

Точность

	y	
	0	1
\hat{y}	0	TN FN
	1	FP TP

$$\text{Recall} = \frac{TP}{TP + FN}$$

Полнота

	y	
	0	1
\hat{y}	0	TN FN
	1	FP TP

ошибка
+ рога
(ур. знан.)

$$\text{FPR} = \frac{FP}{FP + TN}$$

доля неверно клф.
количество

Задача 1

y_i	1	0	0	1	0
\hat{p}_i	0.7	0.2	0.3	0.25	0.1

$$\gamma = 0.6$$

	y	
\hat{y}		
	0	1
0	3	1
1	0	1

$$\text{Precision} = 1$$

$$\text{Recall} = 1/2$$

$$\gamma = 0.26$$

	y	
\hat{y}		
	0	1
0	2	1
1	1	1

$$\text{Precision} = 1/2$$

$$\text{Recall} = 1/2$$

$$\gamma = 0.24$$

	y	
\hat{y}		
	0	1
0	2	0
1	1	2

$$2/3$$

$$1$$

$$\gamma = 0.15$$

	y	
\hat{y}		
	0	1
0	1	0
1	2	2

$$1/2$$

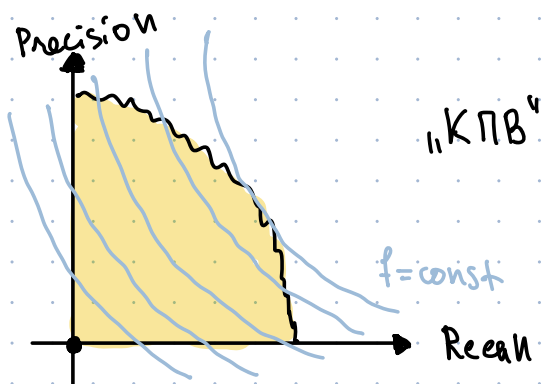
$$1$$

$$\gamma = 0$$

	y	
\hat{y}		
	0	1
0	0	0
1	3	2

$$2/5$$

$$1$$



PR-curve

PR-auc

f_1 -мера

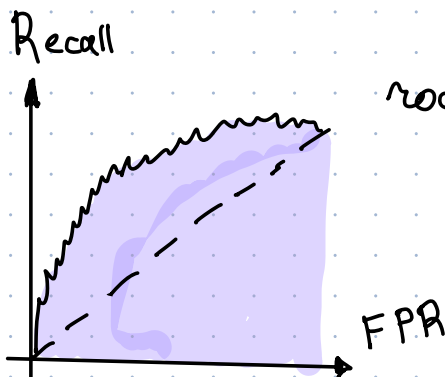
$$f_1 = (1 + \beta)^2 \cdot \frac{p \cdot r \cdot \gamma}{\beta^2 \cdot p + r + \gamma}$$

$$\beta = 1$$

f_1 -мера

Roc-кривая

Roc-AUC



$$\text{roc-auc} \in [1/2; 1]$$

Упражнение

y_i	1	0	0	1	0
\hat{p}_i	0.7	0.2	0.3	0.25	0.1

1) Сортируем по \hat{p}_i

	\hat{p}_i	y_i
①	0.7	1
②	0.3	0
③	0.25	1
④	0.2	0
⑤	0.1	0

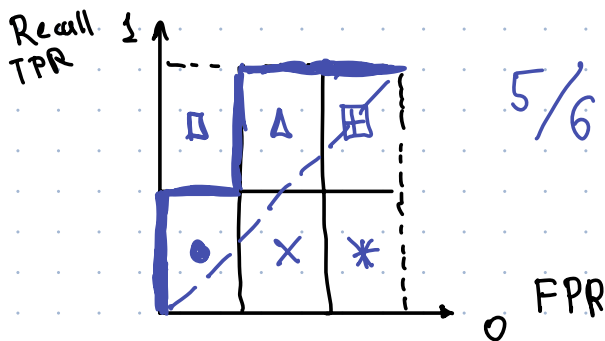
2 eqn.
3 коника } $3 \cdot 2 = 6$ пар

- ①② $0.7 > 0.3$
- × ①④ $0.7 > 0.2$
- * ①⑤ $0.7 > 0.1$
- ③② $0.25 < 0.3$
- △ ③④ $0.25 > 0.2$
- ⊞ ③⑤ $0.25 > 0.1$

Done
хотимых
пар

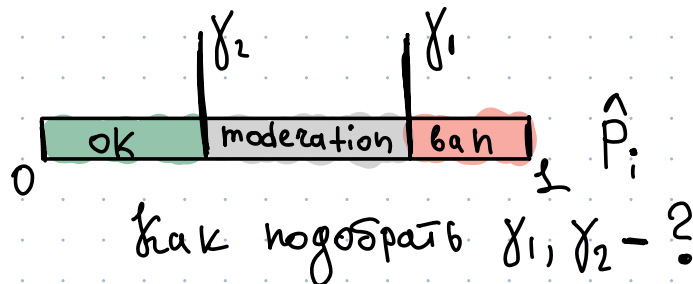
Roc-Auc = $5/6$

2) Строим Roc-curve



Упражнение

youtube
porno



$$\gamma_2: \begin{cases} \text{Recall} \rightarrow \max_{\gamma_2} \\ \text{s.t. video_cnt} \leq 1000 \end{cases}$$

квота
на модераторов

$$\gamma_1: \text{Precision} \geq 0.9$$

или

$$\text{FPR} \leq 0.001$$