

A background image showing a pair of hands being examined. One hand is resting on a surface, while the other hand is placed over it, with fingers spread, suggesting a medical or scientific examination. The image is slightly blurred and has a soft, warm tone.

курс «Прикладные задачи анализа данных»

**Функции ошибки /
функционалы качества**

Часть 2: чёткая бинарная классификации

Александр Дьяконов

24 февраля 2021 года

План на эти несколько лекций

задача регрессии

задача бинарной классификации

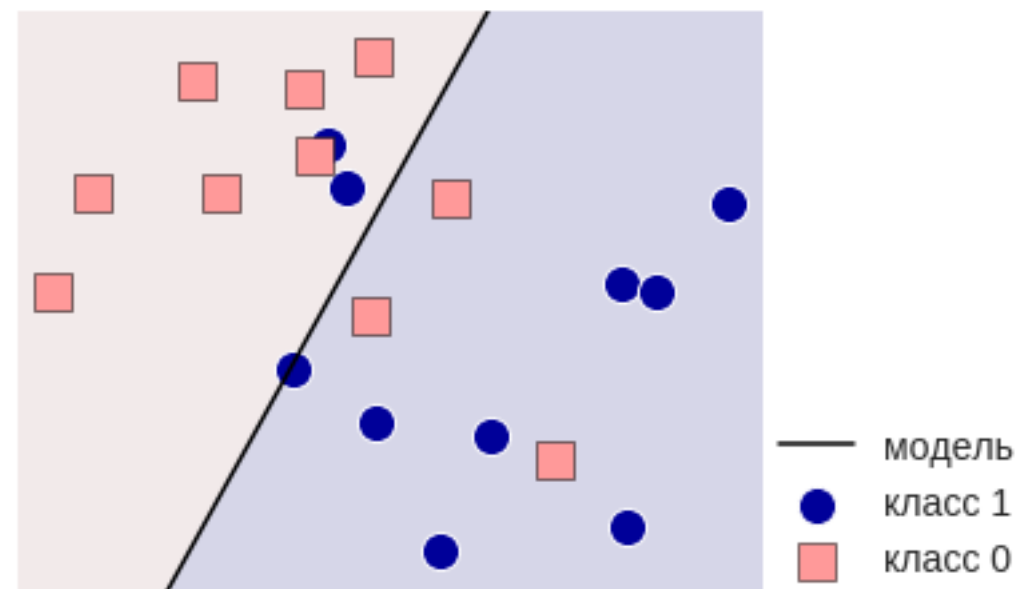
- **чёткая классификация**
- **скоринговые функции**
- кривые в ML**

задача классификации с несколькими классами

задачи ранжирования

задачи кластеризации

Задача классификации



сначала – чёткая классификация

«Confusion Matrix» – матрица ошибок / несоответствий

ОТВЕТЫ

	у	а
0	1	1
1	1	1
2	1	2
3	2	1
4	2	3
5	3	2
6	3	3
7	3	3
8	1	2
9	2	2

матрица
ошибок

	а	1	2	3
у				
1	2	2	0	
2	1	1	1	
3	0	1	2	

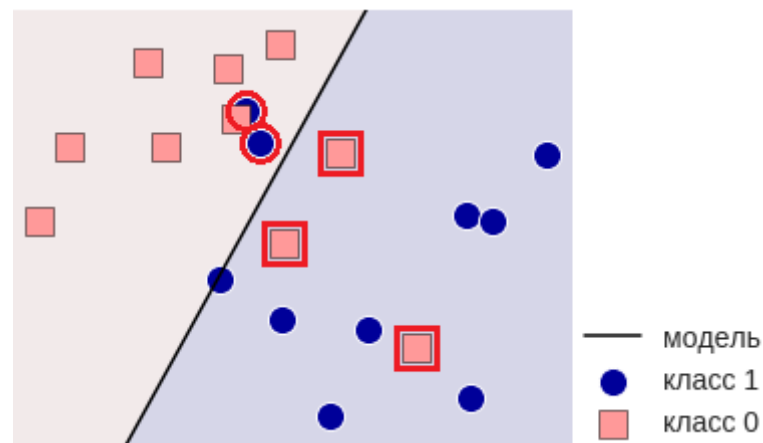
Для классов $\{1, 2, \dots, l\}$

$$N = \| m_{ij} \|_{l \times l}$$

$$m_{ij} = \sum_{t=1}^m I[a_t = i] I[y_t = j]$$

```
from sklearn.metrics import confusion_matrix
n = confusion_matrix(df.y, df.a) # 1й способ
n = pd.crosstab(df.y, df.a) # 2й способ
```

Обычная точность – Accuracy, Mean Consequential Error



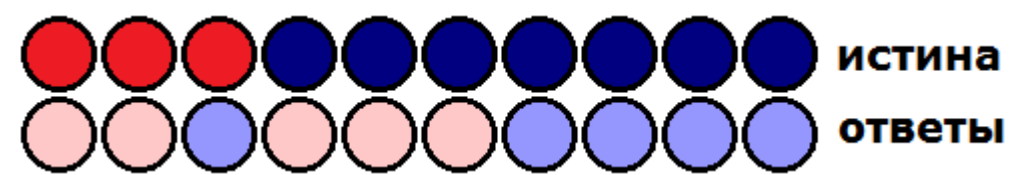
$$\text{MCE} = \frac{1}{m} \sum_{i=1}^m I[a_i \neq y_i] = \frac{\sum_{t=1}^l m_{tt}}{\sum_{t=1}^l \sum_{s=1}^l m_{ts}}$$

- первое, что приходит в голову
- не учитывает разную мощность классов

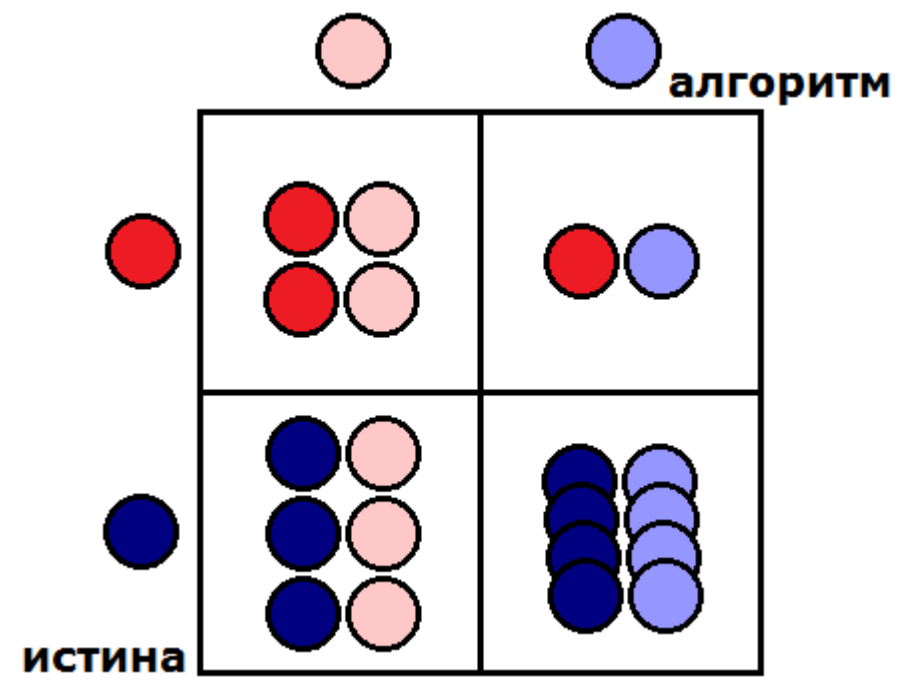
$y = [0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0]$

Выгодно выдавать решение – константу 0!

«Confusion Matrix» в задаче классификации с двумя классами



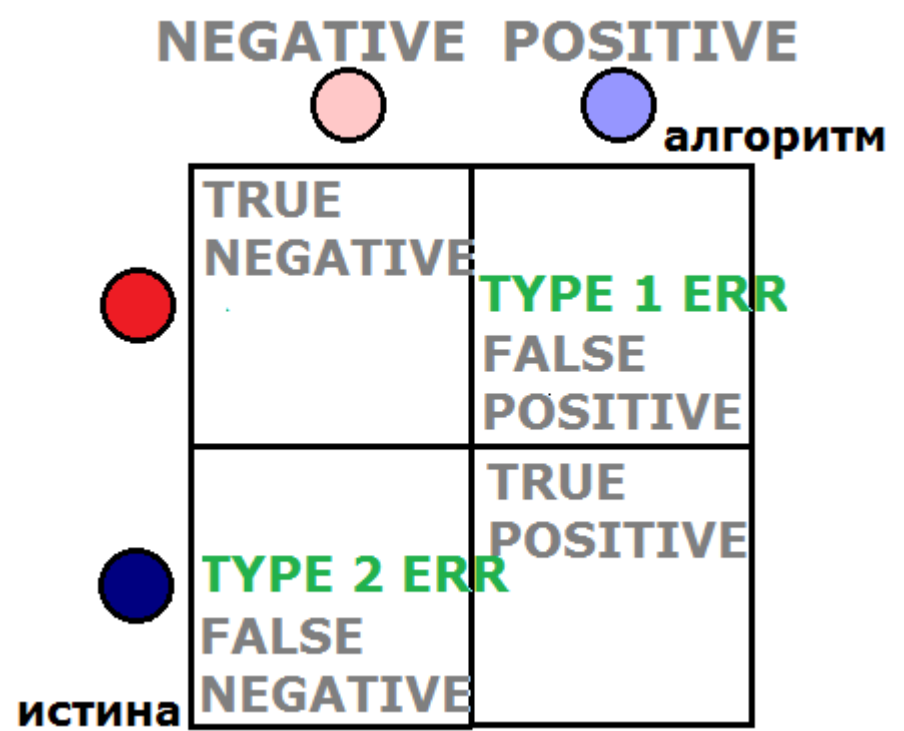
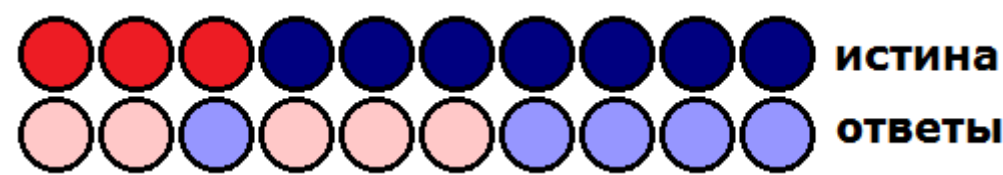
	$a = 0$	$a = 1$
$y = 0$	13599	2600
$y = 1$	898	903



в scikit-learn-е такая ориентация!
Иногда: наоборот!

```
from sklearn.metrics import confusion_matrix
confusion_matrix(y_test, a_test)
```

Задача классификации с двумя классами



```
tn, fp, fn, tp = confusion_matrix(y, a).ravel() # вычисление tn, ...
```

Как запомнить названия ошибок

1 рода – **не учил**, но **сдал** (= знает по мнению экзаменатора)

2 рода – **учил**, но **не сдал** (= не знает по мнению экзаменатора)



Ошибка 1 рода

FP / m

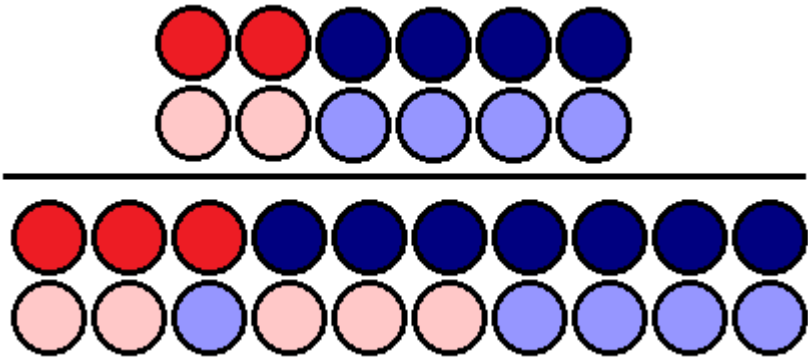


Ошибка 2 рода

FN / m

Точность Accuracy

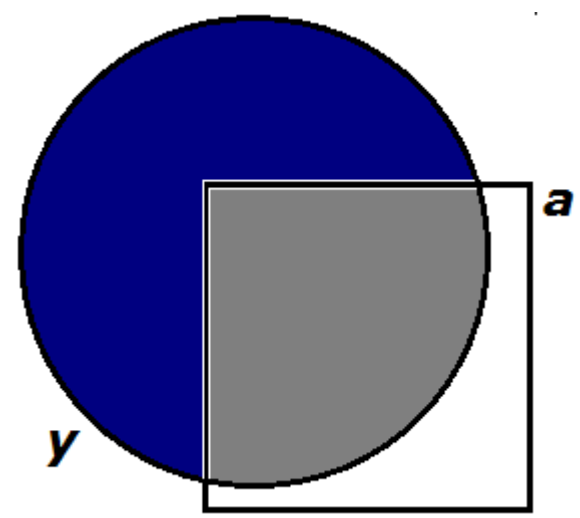
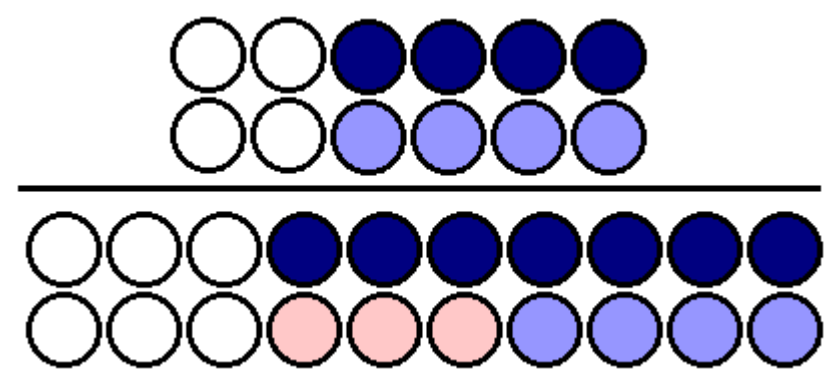
	<i>a</i> = 0	<i>a</i> = 1
<i>y</i> = 0	13599	2600
<i>y</i> = 1	898	903



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

Полнота (Sensitivity, True Positive Rate, Recall, Hit Rate)

	$a = 0$	$a = 1$
$y = 0$	13599	2600
$y = 1$	898	903

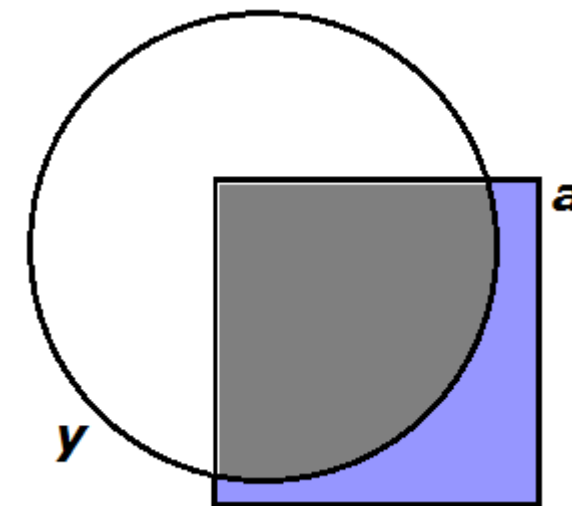
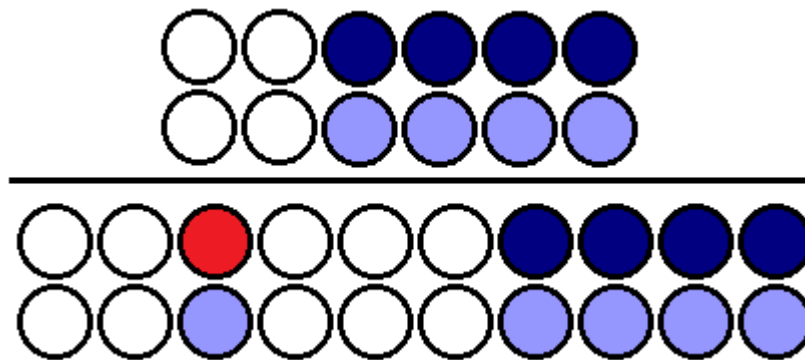


$$TPR = R = \frac{TP}{TP + FN}$$

какой процент объектов положительного класса мы правильно классифицировали

Точность (Precision, Positive Predictive Value)

	$a = 0$	$a = 1$
$y = 0$	13599	2600
$y = 1$	898	903

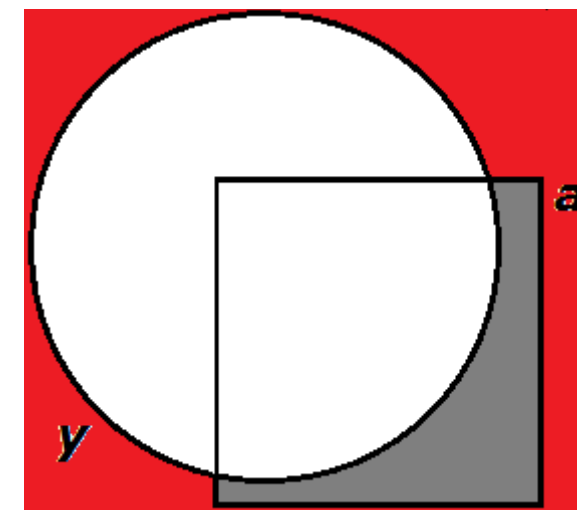
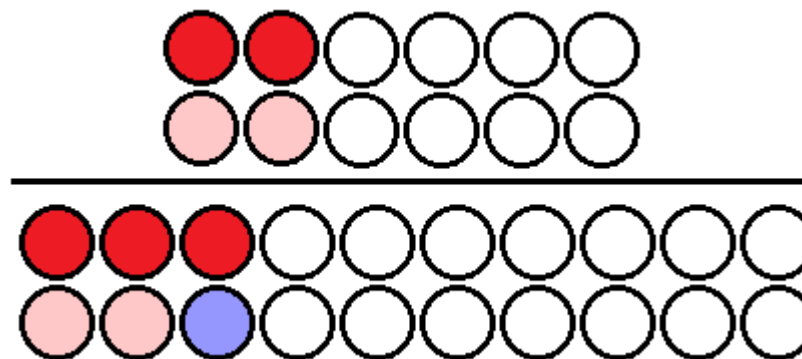


$$PPV = P = \frac{TP}{TP + FP}$$

какой процент положительных объектов
(т.е. тех, что мы считаем положительными)
правильно классифицирован

Специфичность (Specificity, True Negative Rate)

	$a = 0$	$a = 1$
$y = 0$	13599	2600
$y = 1$	898	903

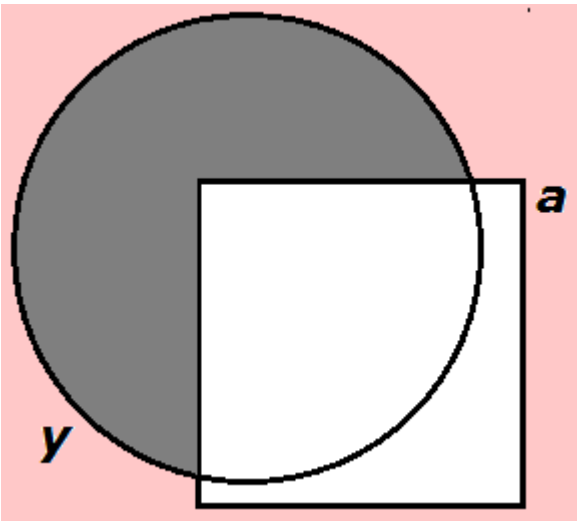
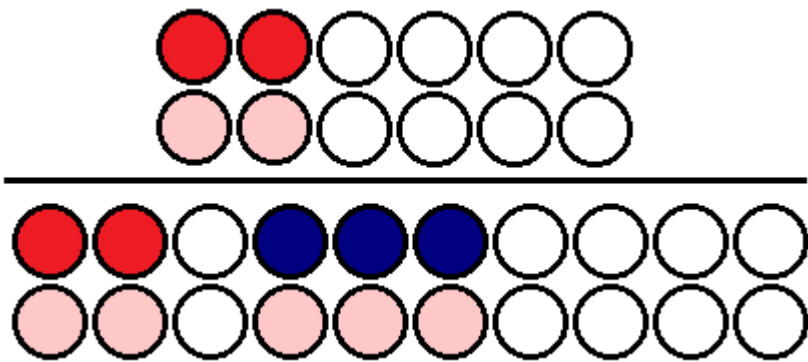


$$\text{TNR} = \text{Specificity} = R_0 = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

процент правильно классифицированных объектов
негативного класса
«полнота для негативного класса»!

Negative Predictive Value (Inverse Precision)

	$a = 0$	$a = 1$
$y = 0$	13599	2600
$y = 1$	898	903



$$NPV = P_0 = \frac{TN}{TN + FN}$$

точность для нулевого класса

False Positive Rate (FPR, fall-out, false alarm rate)

	$a = 0$	$a = 1$
$y = 0$	13599	2600
$y = 1$	898	903

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

**доля объектов негативного класса,
которых мы ошибочно отнесли к положительному**

F₁ score

$$\frac{2}{\frac{1}{P} + \frac{1}{R}} = \frac{2}{\frac{1}{TP/(TP+FP)} + \frac{1}{TP/(TP+FN)}} = \frac{2TP}{2TP+FP+FN}$$

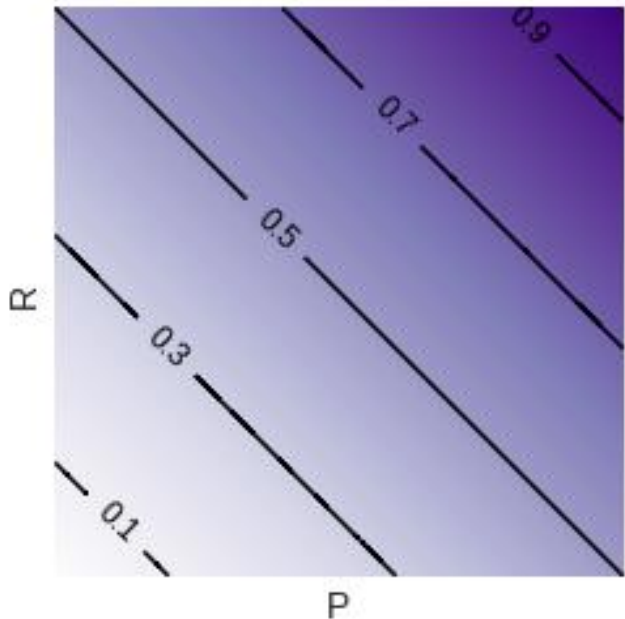
F_β score

$$F_{\beta} = \frac{1}{\frac{\alpha}{P} + \frac{1-\alpha}{R}} = \frac{1}{\alpha} \frac{P \cdot R}{R + \left(\frac{1}{\alpha} - 1\right)P} = (1 + \beta^2) \frac{P \cdot R}{R + \beta^2 P}$$

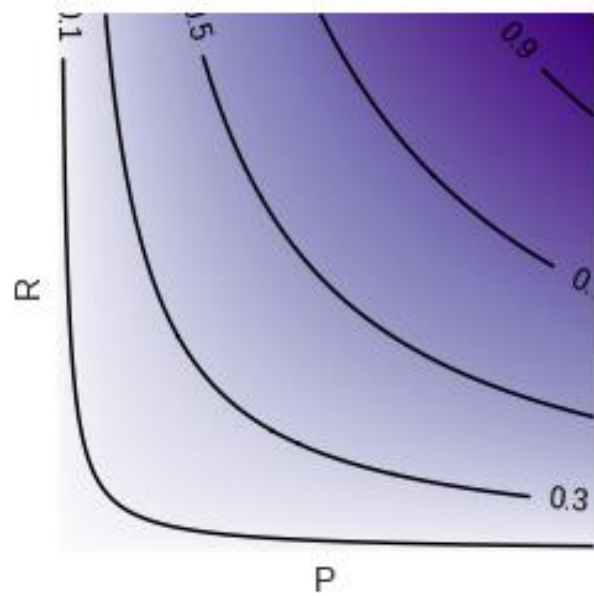
$$\beta^2 = \left(\frac{1}{\alpha} - 1\right)$$

Почему используется F-мера

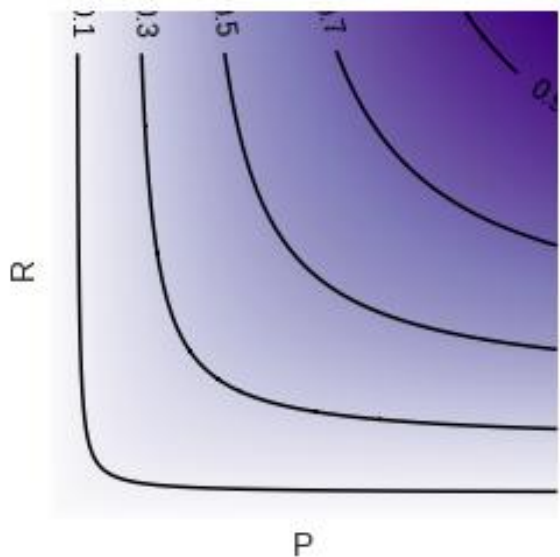
$(P + R) / 2$



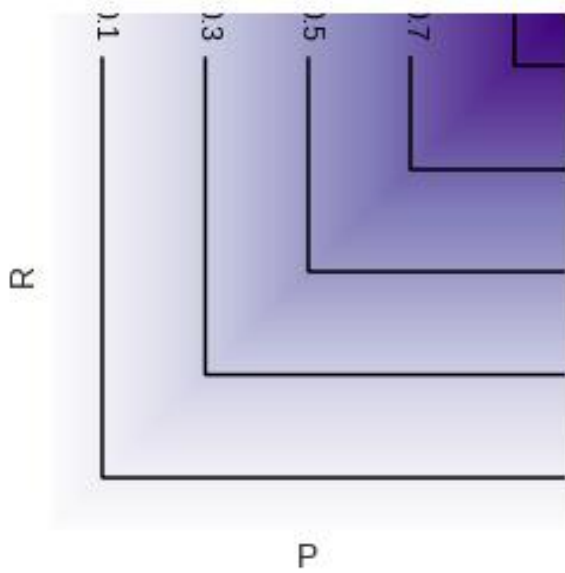
$\sqrt{P \cdot R}$



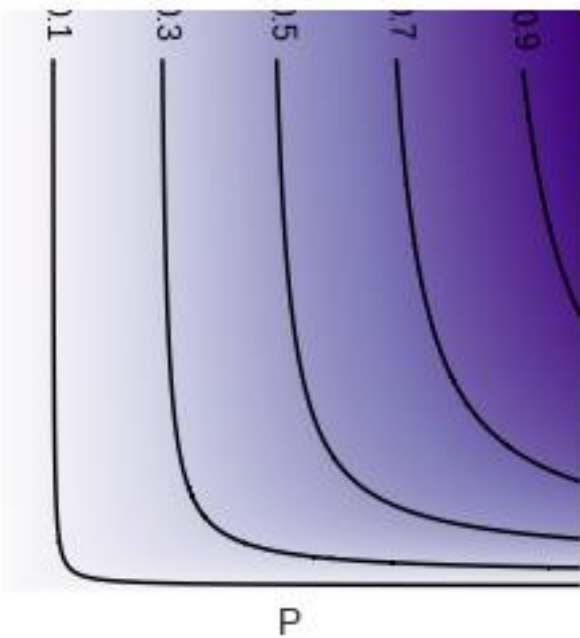
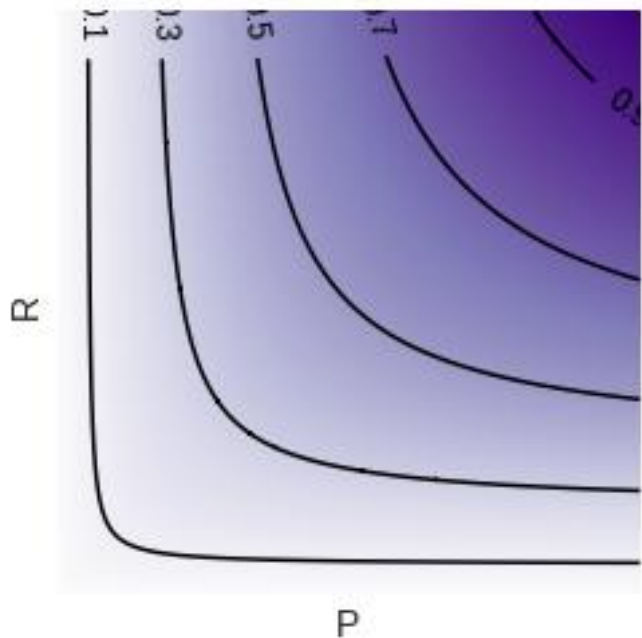
$2 / (1 / P + 1 / R)$



$\min(P, R)$

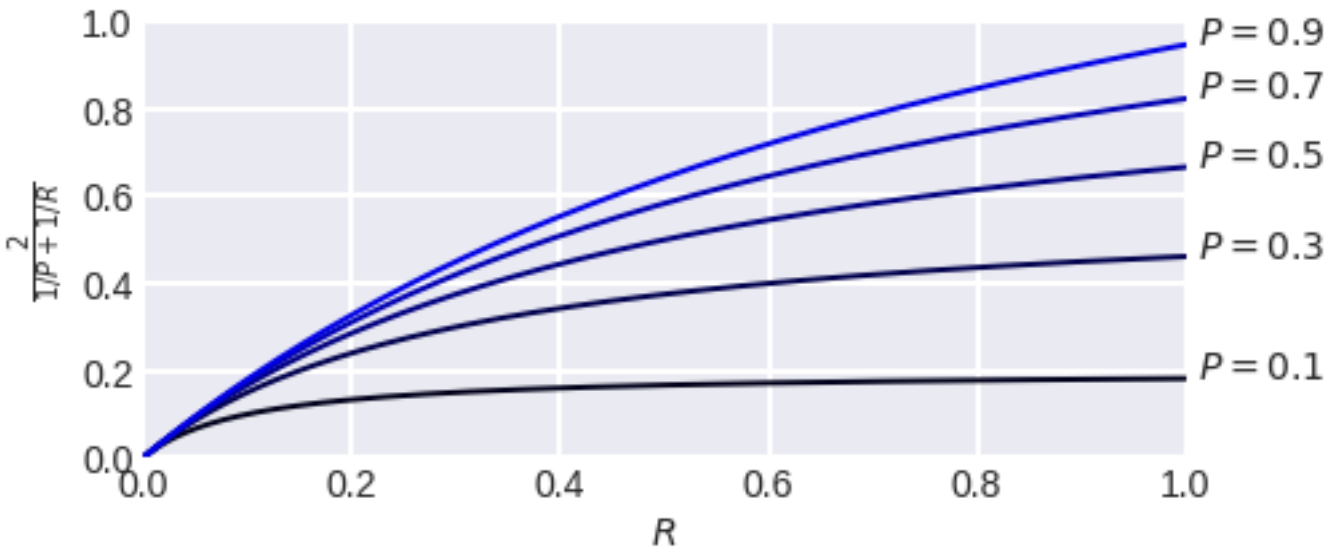


Почему используется F-мера



$2 / (1 / P + 1 / R)$

$1 / (0.9 / P + 0.1 / R)$



Можно сколь угодно улучшать один из показателей (R), если второй не увеличивается (P), то качество ограничено

Каппа Коэна (Cohen's Kappa) в задачах классификации

Идея: поскольку использование точности вызывает сомнение в задачах с сильным дисбалансом классов, надо её значения немного перенормировать

$$K = \frac{\text{Accuracy} - \text{Accuracy}_{\text{chance}}}{1 - \text{Accuracy}_{\text{chance}}}$$

Chance adjusted index – статистика для измерения согласованности между ответами (Accuracy) с нормировкой на согласованность по случайности ($\text{Accuracy}_{\text{chance}}$)

Каппа Коэна (Cohen's Kappa) в задачах классификации

$$K = \frac{\text{Accuracy} - \text{Accuracy}_{\text{chance}}}{1 - \text{Accuracy}_{\text{chance}}}$$

	$a = 0$	$a = 1$
$y = 0$	m_{00}	m_{01}
$y = 1$	m_{10}	m_{11}

$$\text{Accuracy} = \frac{m_{00} + m_{11}}{m}$$

$$\text{Accuracy}_{\text{chance}} = \frac{m_{00} + m_{01}}{m} \frac{m_{00} + m_{10}}{m} + \frac{m_{10} + m_{11}}{m} \frac{m_{01} + m_{11}}{m}$$

- – вероятность угадать класс 0
- – вероятность угадать класс 1

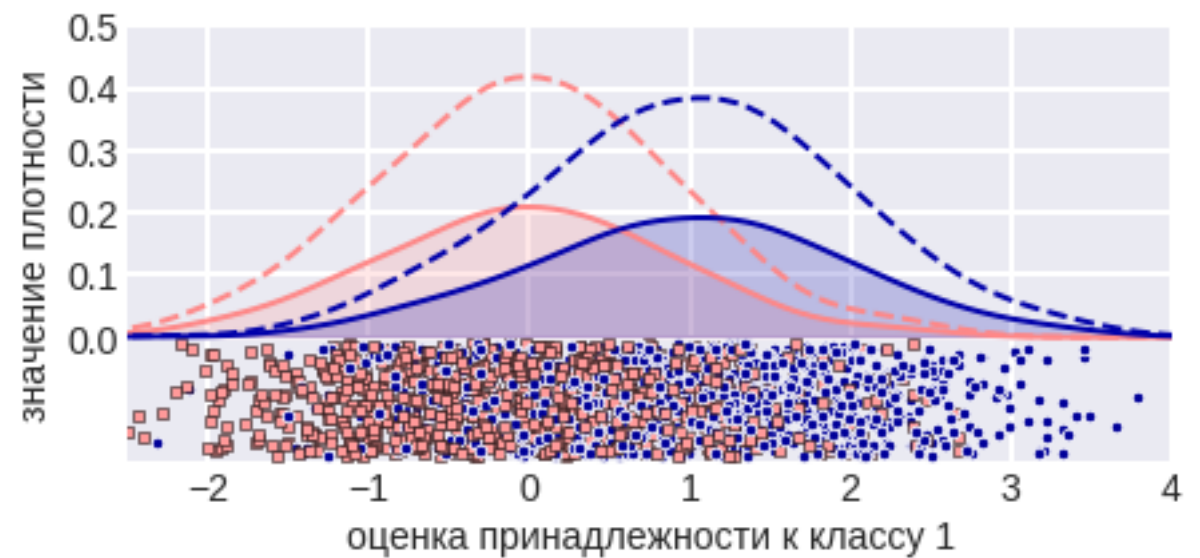
Каппа Коэна (Cohen's Kappa)

смысл: поправка значения точности.

Как раз для решения проблемы дисбаланса классов.

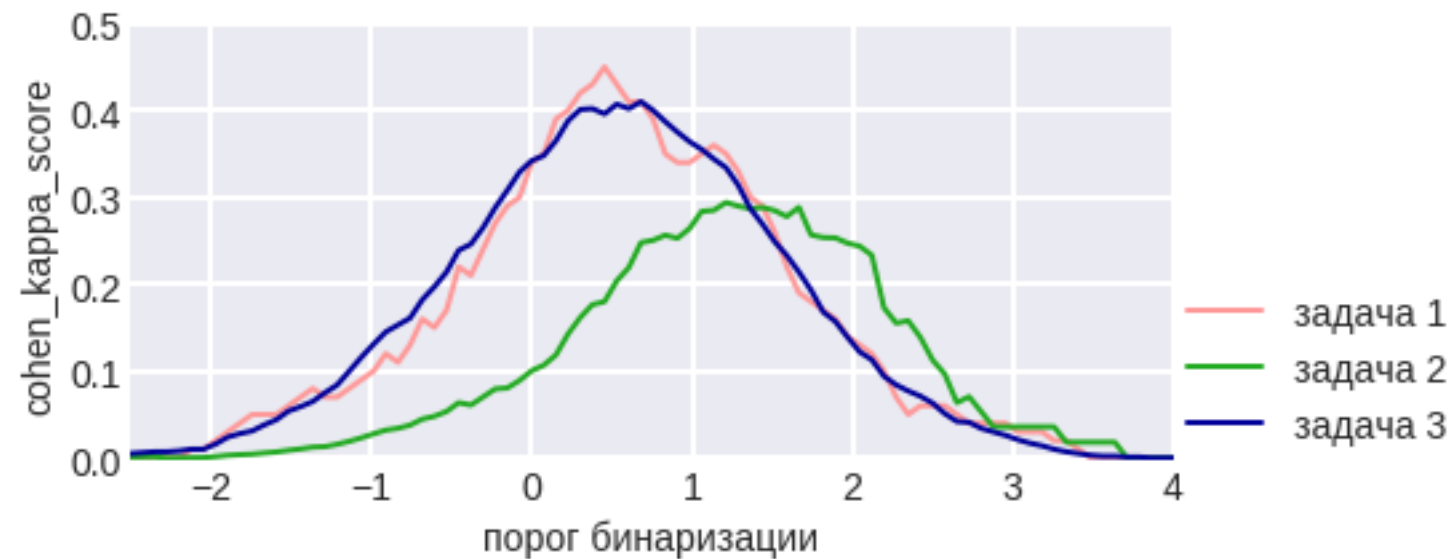
```
from sklearn.metrics import cohen_kappa_score  
cohen_kappa_score(a, y)
```

Каппа Коэна (Cohen's Kappa) три модельные задачи



Каппа Коэна (Cohen's Kappa)

график СК от порога бинаризации



ROC AUC: 0.77 во всех задачах!

Коэффициент Мэттьюса
Matthews correlation coefficient (MCC)

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \in [-1, +1]$$

Подходит для несбалансированных выборок

Коэффициент Мэттьюса

Для понимания смысла...

Рассмотрим среднее геометрическое точности и полноты:

$$\sqrt{P \cdot R} = \sqrt{\frac{TP}{TP+FP} \cdot \frac{TP}{TP+FN}} = \frac{TP}{\sqrt{(TP+FP)(TP+FN)}}$$

это точность и полнота для класса 1,

теперь умножим это на точность и полноту для класса 0:

$$\sqrt{P_1 R_1 P_0 R_0} = \frac{TP \cdot TN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \in [0, 1]$$

теперь посмотрим на формулу MCC...

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \in [-1, +1]$$

Сбалансированная точность (Balanced Accuracy)

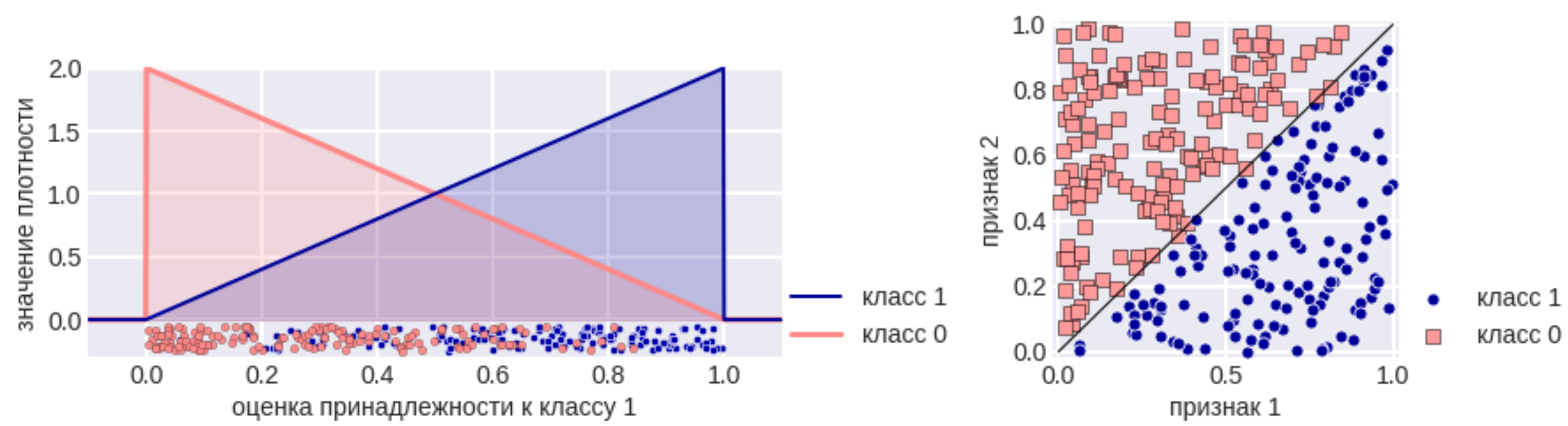
$$BA = \frac{R_1 + R_0}{2} = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right)$$

среднее арифметическое чувствительности и специфичности

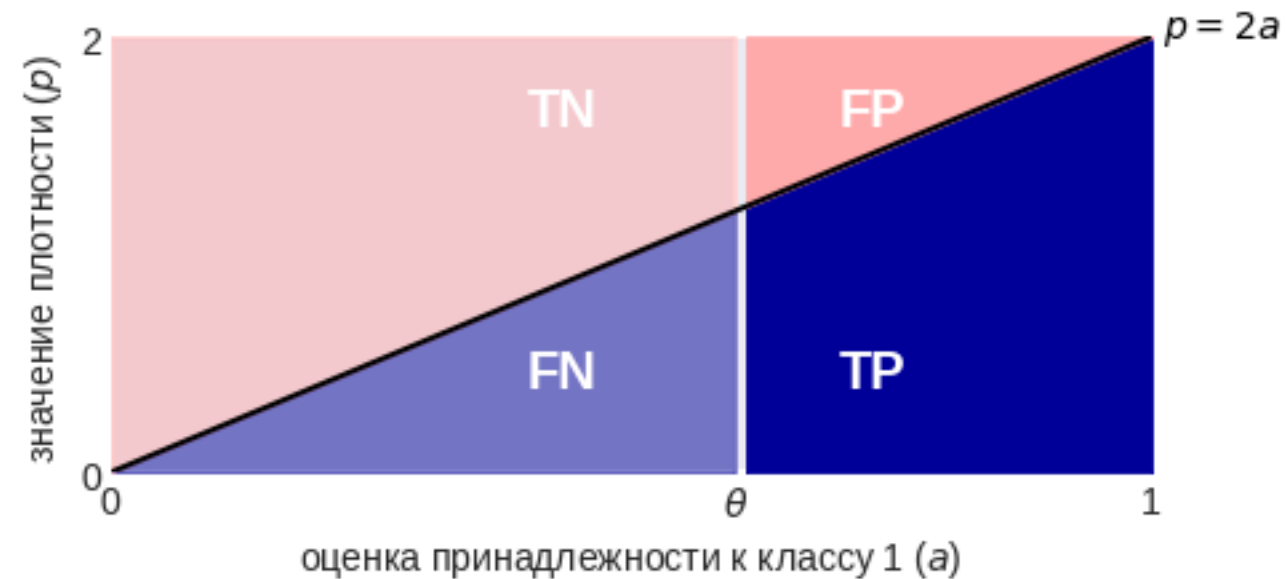
Если классы примерно равноможны...

$$TP + FN \approx TN + FP \approx m / 2$$

Сравнения функционалов качества – модельная задача



Сравнения функционалов качества – модельная задача



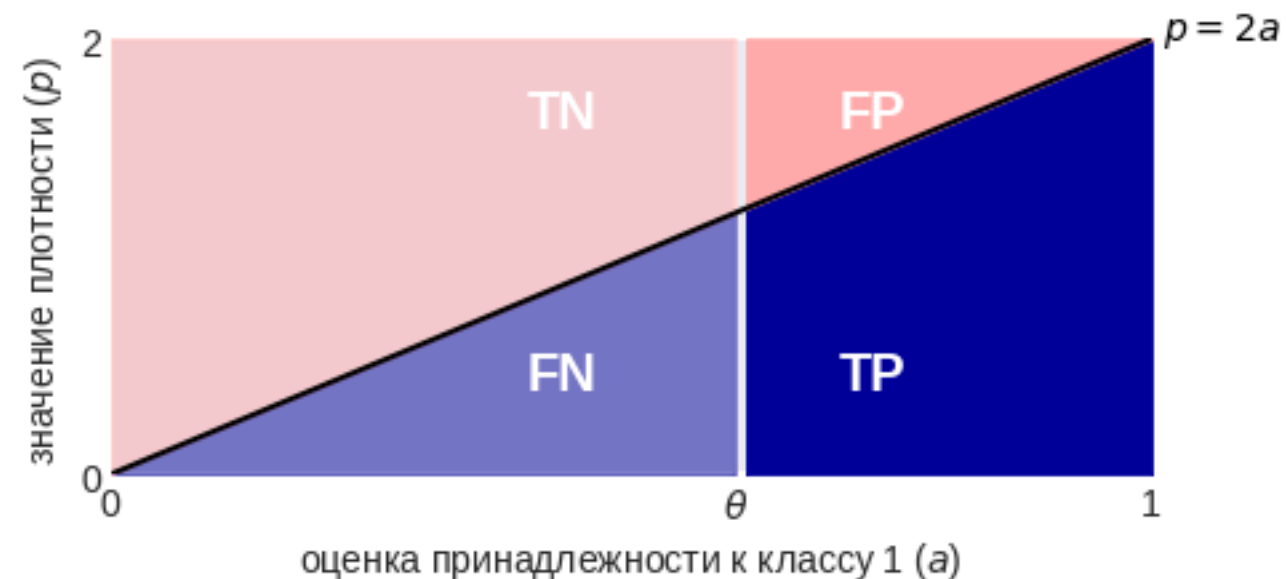
$$R = 1 - \theta^2$$

$$P = (1 + \theta) / 2$$

$$F_1 = \frac{1 - \theta^2}{1.5 - \theta}$$

$$\text{MCC} = \sqrt{\theta(1 - \theta)}$$

$$\kappa = \frac{\frac{1 + 2\theta - 2\theta^2}{2} - \frac{1}{2}}{1 - \frac{1}{2}} = 2\theta(1 - \theta)$$

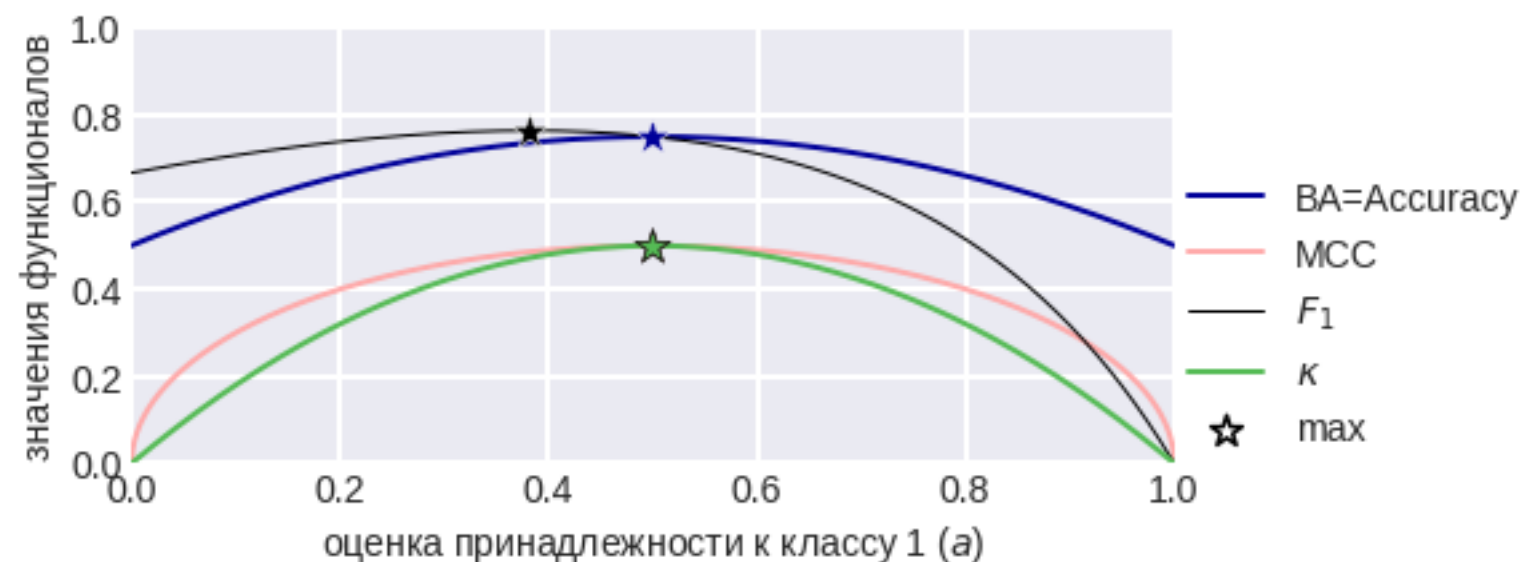
Сравнения функционалов качества – модельная задача

$$BA=Accuracy = \frac{(1-\theta^2) + (1-(1-\theta)^2)}{2} = \frac{1+2\theta-2\theta^2}{2}$$

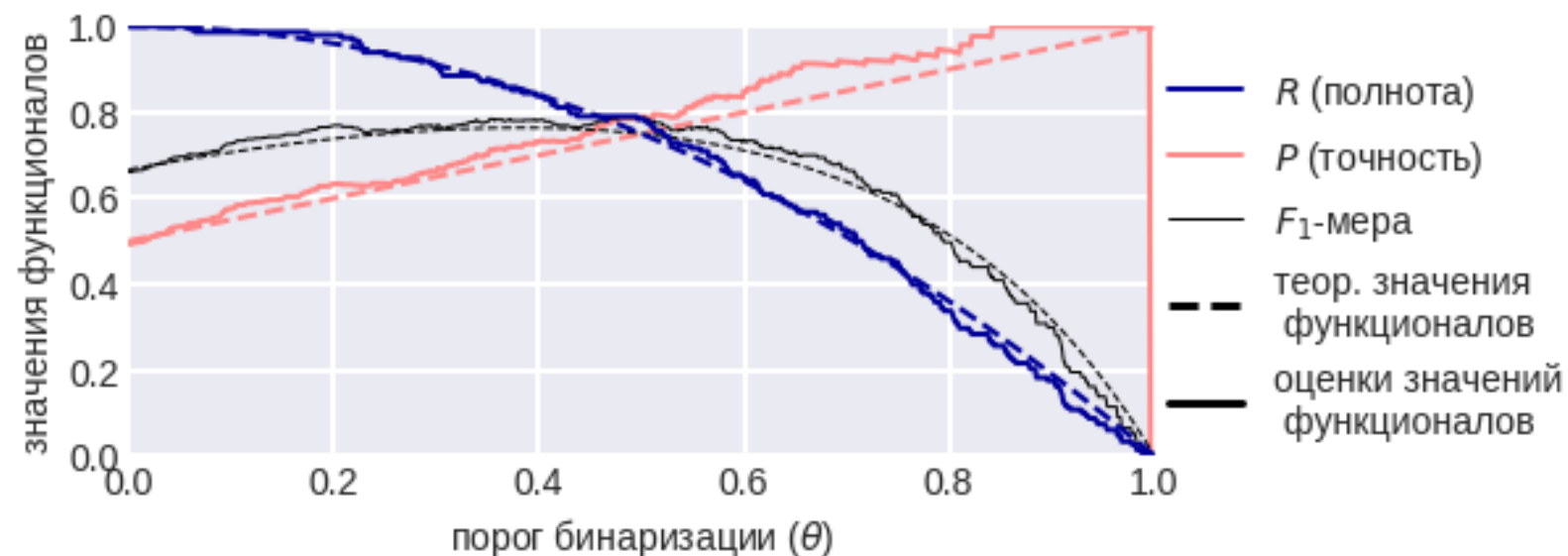
оптимальный порог для F-меры

$$\theta_{opt} = \frac{3-\sqrt{5}}{2} \approx 0.38$$

Сравнения функционалов качества – модельная задача

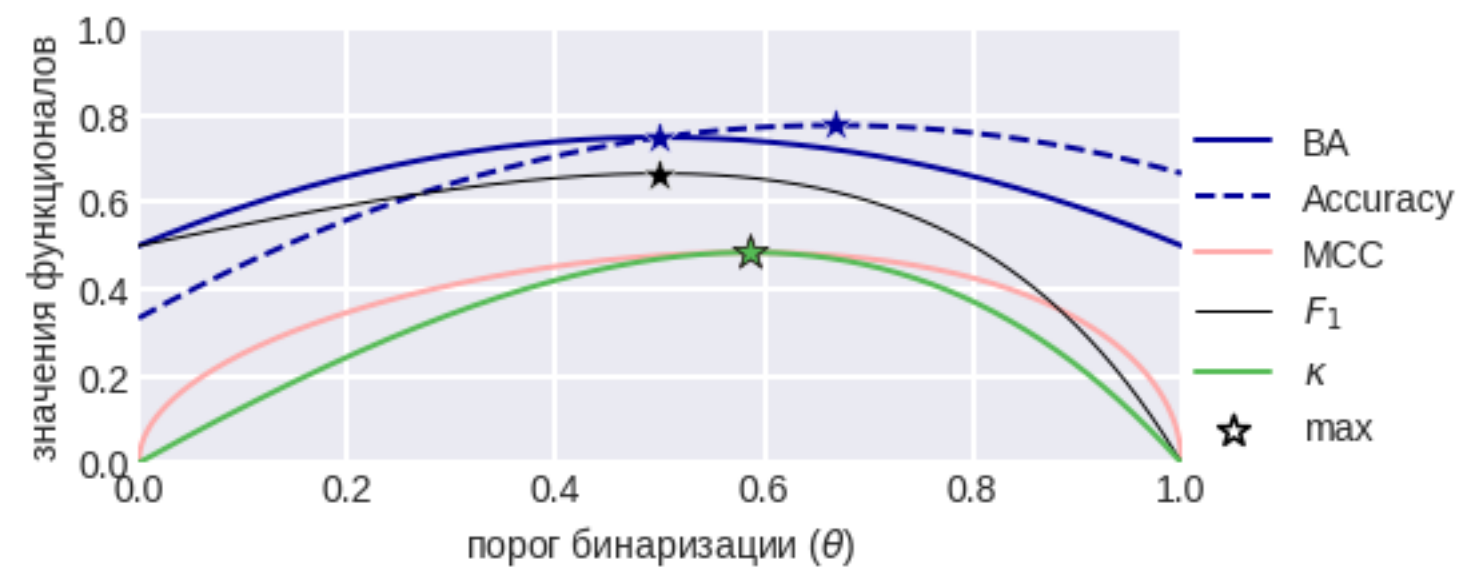


в теории и оценённые по выборке 300 объектов...

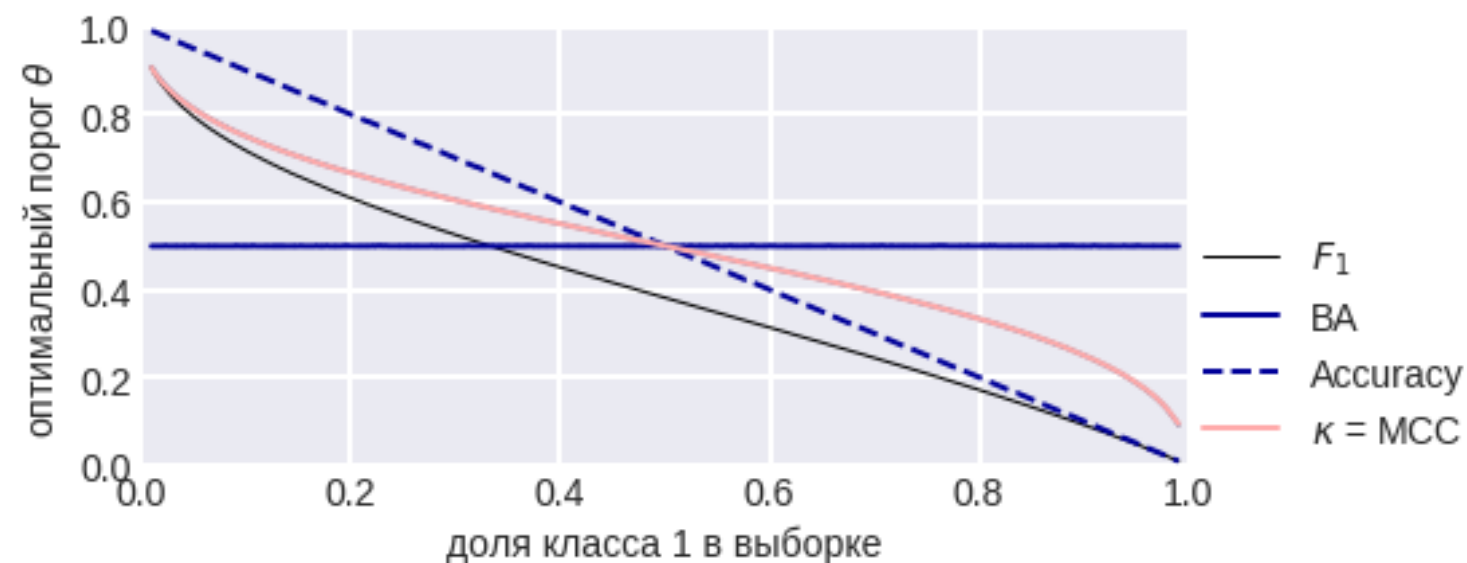


Дисбаланс классов

если сделать дисбаланс классов – **какой?**



Дисбаланс классов



Оптимальный порог бинаризации для разных функционалов при изменении доли класса 1 в выборке

Вопросы

- **у какого функционала качества самый маленький оптимальный порог бинаризации (в общем случае), почему?**
- **какой функционал качества действительно имеет смысл использовать в задачах с сильным дисбалансом классов (заметим, что стандартные советы: ВА, МСС, карра, F1 обладают совершенно разными свойствами)?**
- **какой «самый неустойчивый» из перечисленных функционалов (его значения на небольших выборках сильнее отличаются от вычисленных на достаточно больших)?**
 - **что изменится в примерах выше, если от линейных плотностей перейти к нормальным? Как это сделать корректно (и в чём некорректность описанной модельной задачи)?**

Вопросы

- **верно ли, что максимальное значение точности (т.е. значение точности при оптимальном выборе порога) всегда не меньше максимального значения сбалансированной точности?**
- **выведите явные формулы функционалов качества в задаче с дисбалансом классов (придётся повозиться с формулами).**

Минутка кода

```
from sklearn.metrics import classification_report
print (classification_report(y_test, a_test)) # нужен print
```

	precision	recall	f1-score	support
0.0	0.94	0.84	0.89	16199
1.0	0.26	0.50	0.34	1801
micro avg	0.81	0.81	0.81	18000
macro avg	0.60	0.67	0.61	18000
weighted avg	0.87	0.81	0.83	18000

```
from sklearn.metrics import cohen_kappa_score
from sklearn.metrics import accuracy_score
from sklearn.metrics import matthews_corrcoef
from sklearn.metrics import f1_score
from sklearn.metrics import roc_auc_score
from sklearn.metrics import balanced_accuracy_score
```

	score
cohen_kappa_score	0.24
accuracy_score	0.81
matthews_corrcoef	0.26
f1_score	0.34
roc_auc_score	0.67
balanced_accuracy_score	0.67

Итог

В задаче чёткой бинарной классификации вся информация об ошибках в 2×2 -матрице несоответствий

Много разных функционалов качества

- естественные
- из информационного поиска
- для учёта дисбаланса

Accuracy	P / R	
BA	F₁	κ / MCC

Литература

Jeffrey M Girard «Inter-observer reliability» //

<https://github.com/jmgirard/mReliability/wiki>

Функционалы качества бинарной классификации

<https://dyakonov.org/2019/05/31/функционалы-качества-в-задаче-бинарн/>