

Машинное обучение

Оценки качества

Predicted condition			Sources: [11][12][13][14][15][16][17][18][19] view · talk · edit		
Actual condition	Total population $= P + N$	Positive (PP)	Negative (PN)	Informedness, bookmaker informedness (BM) $= TPR + TNR - 1$	Prevalence threshold (PT) $= \frac{\sqrt{TPR \times FPR} - FPR}{TPR - FPR}$
	Positive (P)	True positive (TP), hit	False negative (FN), type II error, miss, underestimation	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power $= \frac{TP}{P} = 1 - FNR$	False negative rate (FNR), miss rate $= \frac{FN}{P} = 1 - TPR$
Negative (N)	False positive (FP), type I error, false alarm, overestimation	True negative (TN), correct rejection	False positive rate (FPR), probability of false alarm, fall-out $= \frac{FP}{N} = 1 - TNR$	True negative rate (TNR), specificity (SPC), selectivity $= \frac{TN}{N} = 1 - FPR$	
Prevalence $= \frac{P}{P + N}$	Positive predictive value (PPV), precision $= \frac{TP}{PP} = 1 - FDR$	False omission rate (FOR) $= \frac{FN}{PN} = 1 - NPV$	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$	Negative likelihood ratio (LR-) $= \frac{FNR}{TNR}$	
Accuracy (ACC) $= \frac{TP + TN}{P + N}$	False discovery rate (FDR) $= \frac{FP}{PP} = 1 - PPV$	Negative predictive value (NPV) $= \frac{TN}{PN}$ $= 1 - FOR$	Markedness (MK), deltaP (Δp) $= PPV + NPV - 1$	Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$	
Balanced accuracy (BA) $= \frac{TPR + TNR}{2}$	F_1 score $= \frac{2PPV \times TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$	Fowlkes–Mallows index (FM) $= \sqrt{PPV \times TPR}$	Matthews correlation coefficient (MCC) $= \sqrt{TPR \times TNR \times PPV \times NPV}$ $- \sqrt{FNR \times FPR \times FOR \times DFR}$	Threat score (TS), critical success index (CSI), Jaccard index $= \frac{TP}{TP + FN + FP}$	

Содержание лекции

- Оценки качества бинарной классификации
- Кривые ROC и PR
- Цены за ошибку. Plug-in подход. Выбор порогов
- Калибровка вероятностей

Оценки качества бинарной классификации

		Predicted condition		Sources: [11][12][13][14][15][16][17][18][19] view · talk · edit
		Total population $= P + N$		
Actual condition	Total population $= P + N$	Positive (PP)	Negative (PN)	Informedness, bookmaker informedness (BM) $= TPR + TNR - 1$
	Positive (P)	True positive (TP), hit	False negative (FN), type II error, miss, underestimation	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power $= \frac{TP}{P} = 1 - FNR$
	Negative (N)	False positive (FP), type I error, false alarm, overestimation	True negative (TN), correct rejection	False positive rate (FPR), probability of false alarm, fall-out $= \frac{FP}{N} = 1 - TNR$
	Prevalence $= \frac{P}{P + N}$	Positive predictive value (PPV), precision $= \frac{TP}{PP} = 1 - FDR$	False omission rate (FOR) $= \frac{FN}{PN} = 1 - NPV$	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$
	Accuracy (ACC) $= \frac{TP + TN}{P + N}$	False discovery rate (FDR) $= \frac{FP}{PP} = 1 - PPV$	Negative predictive value (NPV) $= \frac{TN}{PN} = 1 - FOR$	Negative likelihood ratio (LR-) $= \frac{FNR}{TNR}$
	Balanced accuracy (BA) $= \frac{TPR + TNR}{2}$	F_1 score $= \frac{2PPV \times TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$	Fowlkes–Mallows index (FM) $= \sqrt{PPV \times TPR}$	Markedness (MK), deltaP (Δp) $= PPV + NPV - 1$
				Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$
				Matthews correlation coefficient (MCC) $= \sqrt{TPR \times TNR \times PPV \times NPV} - \sqrt{FNR \times FPR \times FOR \times DFR}$
				Threat score (TS), critical success index (CSI), Jaccard index $= \frac{TP}{TP + FN + FP}$

Д/з: как в случае несбалансированной выборки сравнить качество классификации задолжников и надежных клиентов в задаче кредитного scoring?

ROC-кривая

- ROC – receiver operating characteristic
- Каждая точка кривой соответствует одному значению порога (цен за ошибки, w_0)
- По оси X:
FPR (false positive rate) – процент объектов с $y=-1$ и $a(x)=+1$ среди всех $y=-1$
- По оси Y:
TPR (true positive rate) – процент объектов с $y=+1$ и $a(x)=+1$ среди всех $y=+1$

FPR и TPR

		Predicted	
		Positive (PP)	Negative (PN)
Actual	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

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

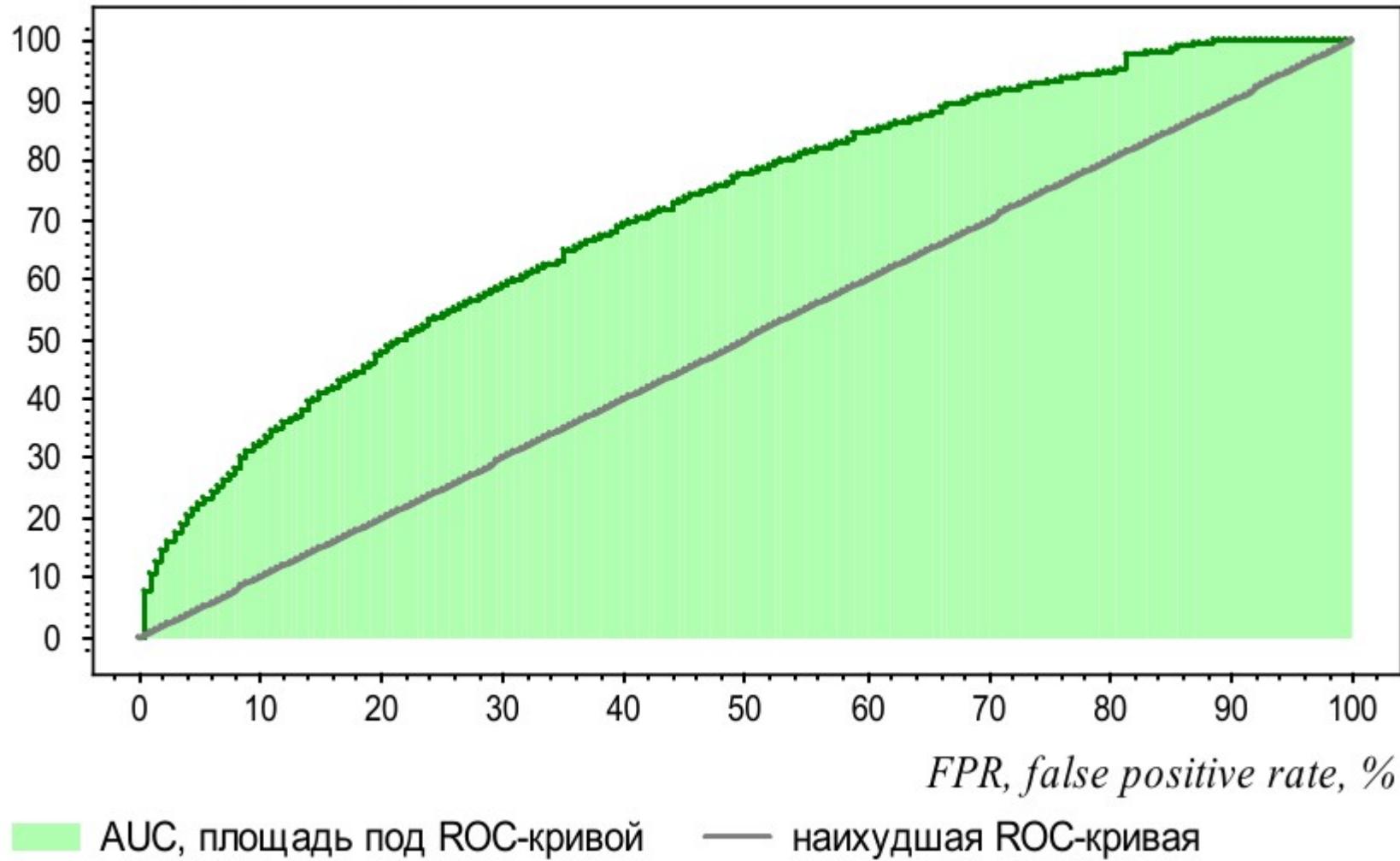
Другие названия:
полнота (recall),
чувствительность (sensitivity)

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

$\text{FPR} = 1 - \text{TNR} =$
 $1 - \text{Recall}$ для
класса "-"

Пример

TPR, true positive rate, %



■ АУС, площадь под ROC-кривой

— наихудшая ROC-кривая

Алгоритм построения ROC-кривой

Вход: выборка X^ℓ ; дискриминантная функция $g(x, w)$;

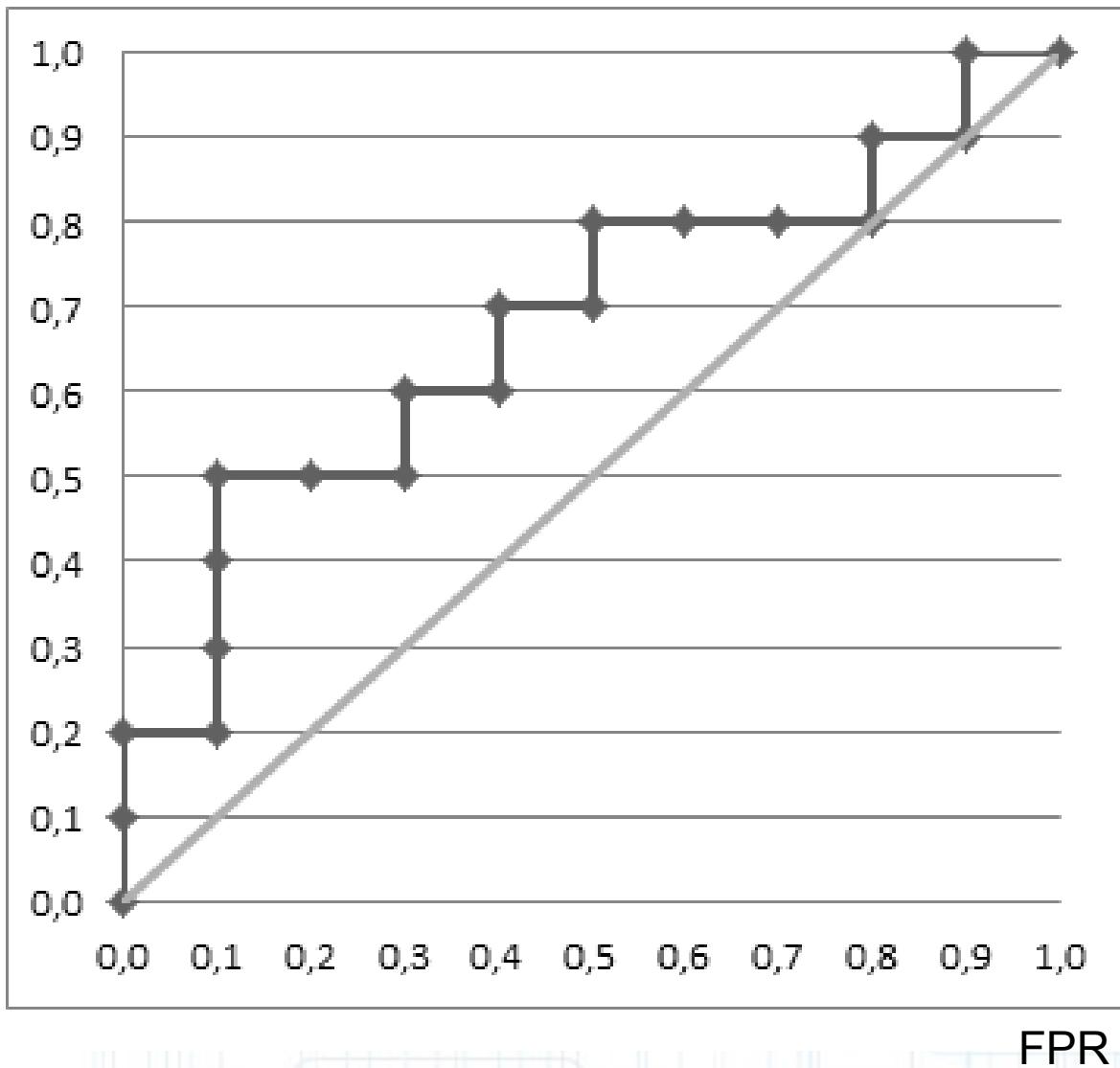
Выход: $\{(FPR_i, TPR_i)\}_{i=0}^\ell$, AUC — площадь под ROC-кривой.

- 1: $\ell_y := \sum_{i=1}^\ell [y_i = y]$, для всех $y \in Y$;
- 2: упорядочить выборку X^ℓ по убыванию значений $g(x_i, w)$;
- 3: поставить первую точку в начало координат:
 $(FPR_0, TPR_0) := (0, 0)$; $AUC := 0$;
- 4: **для** $i := 1, \dots, \ell$
- 5: **если** $y_i = -1$ **то** сместиться на один шаг вправо:
- 6: $FPR_i := FPR_{i-1} + \frac{1}{\ell_-}$; $TPR_i := TPR_{i-1}$;
 $AUC := AUC + \frac{1}{\ell_-} TPR_i$;
- 7: **иначе** сместиться на один шаг вверх:
- 8: $FPR_i := FPR_{i-1}$; $TPR_i := TPR_{i-1} + \frac{1}{\ell_+}$;

Пример

TPR

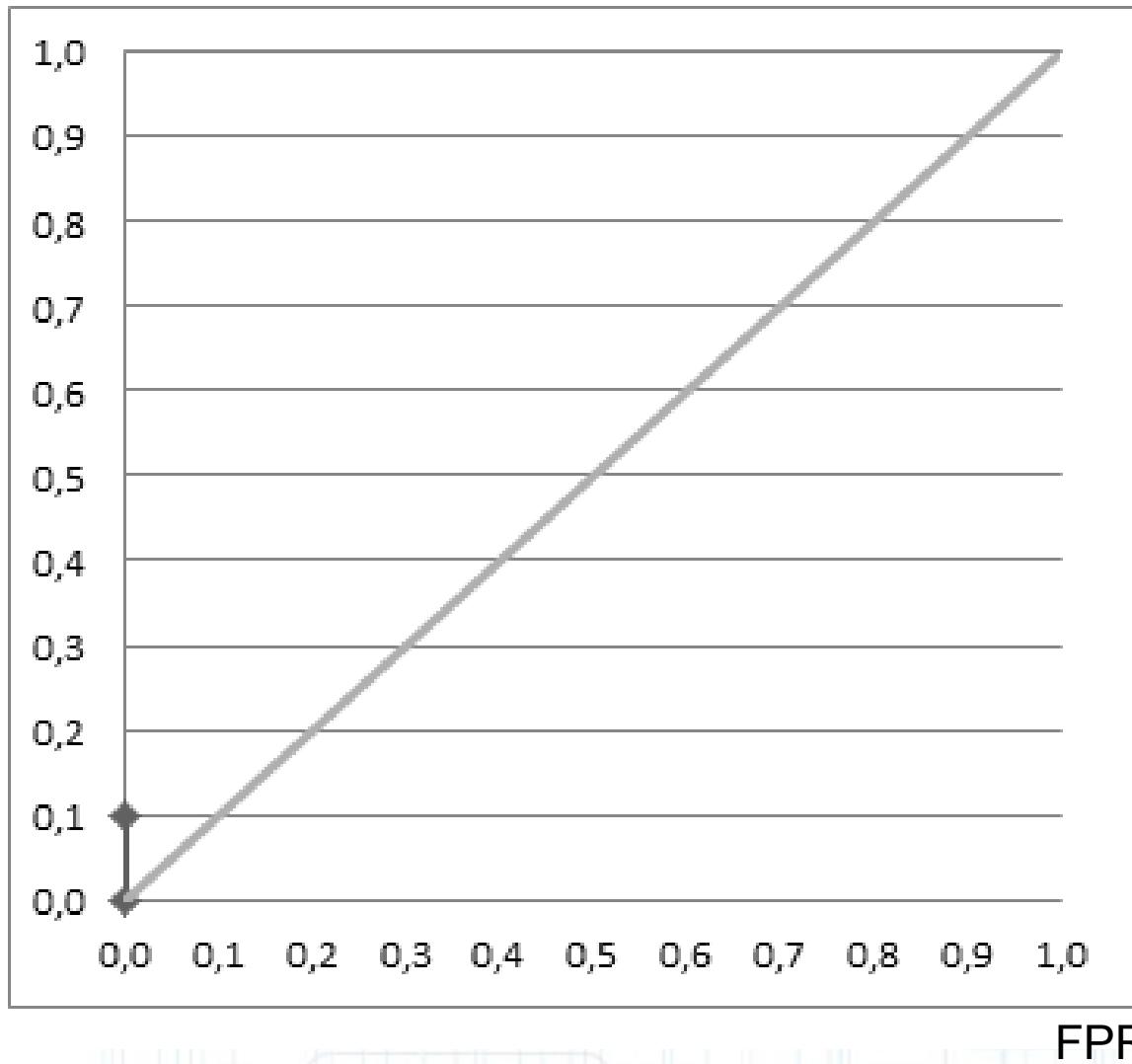
#	C	Score
1	P	0,9
2	P	0,8
3	N	0,7
4	P	0,6
5	P	0,55
6	P	0,54
7	N	0,53
8	N	0,52
9	P	0,51
10	N	0,505
11	P	0,4
12	N	0,39
13	P	0,38
14	N	0,37
15	N	0,36
16	N	0,35
17	P	0,34
18	N	0,33
19	P	0,3
20	N	0,1



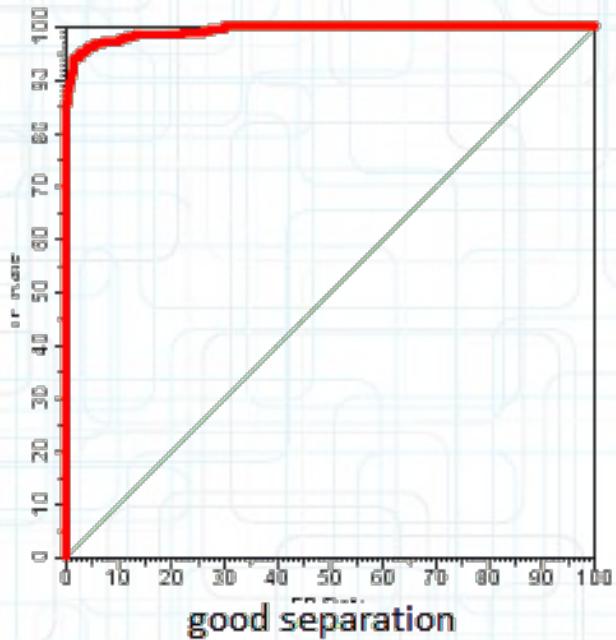
Пример

TPR

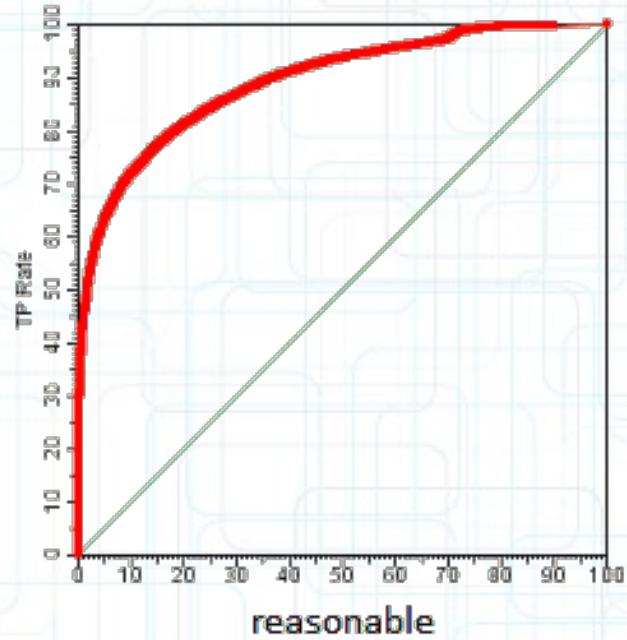
#	C	Score
1	P	0,9
2	P	0,8
3	N	0,7
4	P	0,6
5	P	0,55
6	P	0,54
7	N	0,53
8	N	0,52
9	P	0,51
10	N	0,505
11	P	0,4
12	N	0,39
13	P	0,38
14	N	0,37
15	N	0,36
16	N	0,35
17	P	0,34
18	N	0,33
19	P	0,3
20	N	0,1



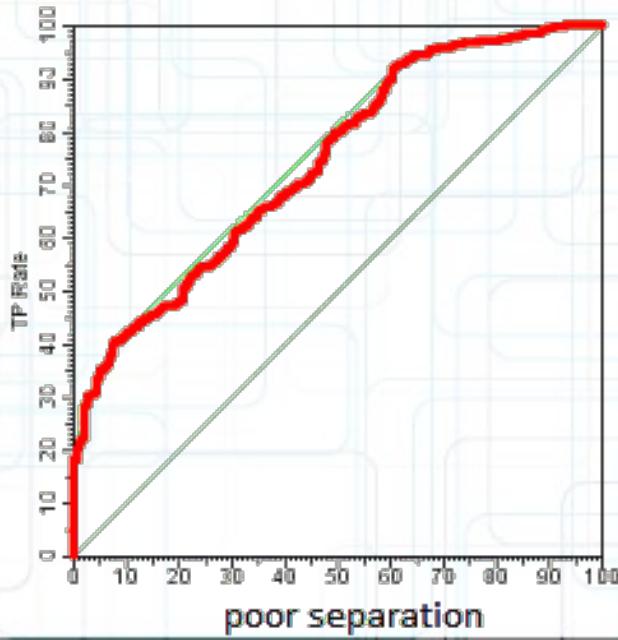
Еще примеры



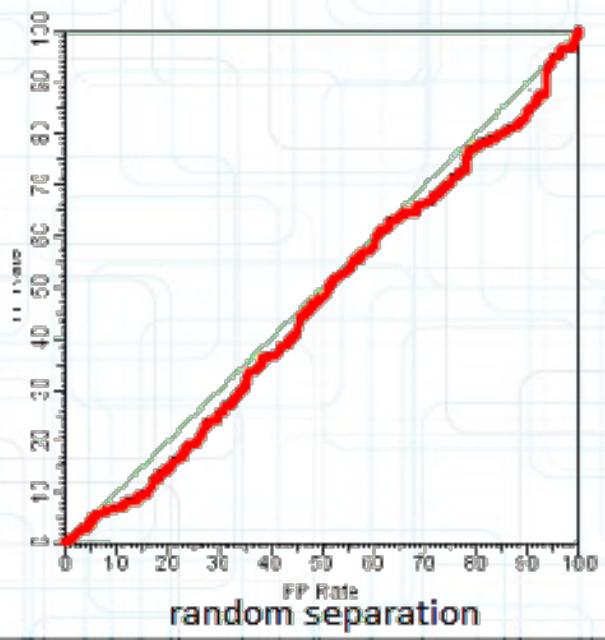
good separation



reasonable



poor separation



random separation

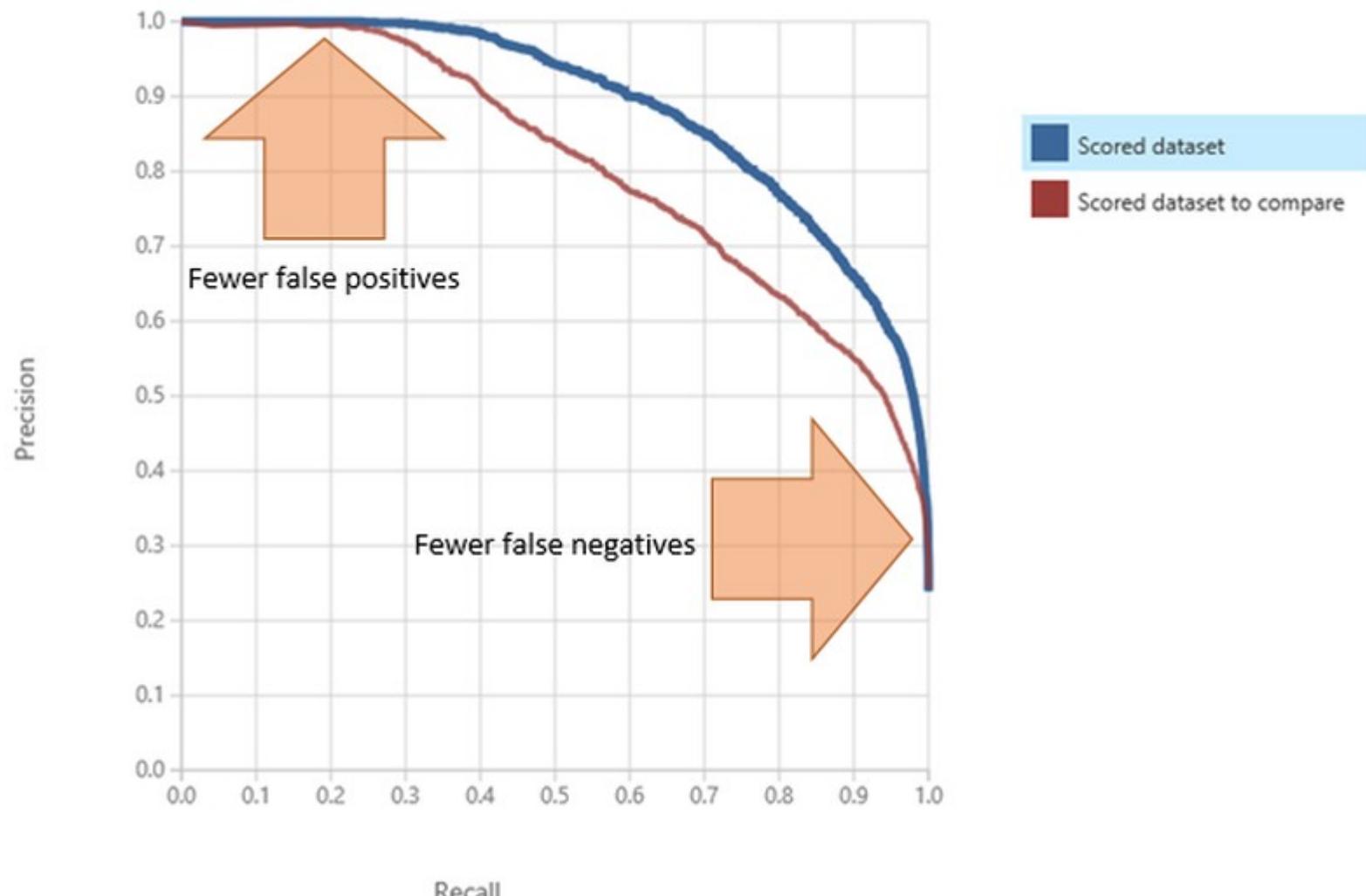
Точность (precision) и полнота (recall)

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

		Predicted	
		Positive (PP)	Negative (PN)
Actual	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

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

PR-кривая



Средняя точность (AveP) – площадь под PR-кривой.

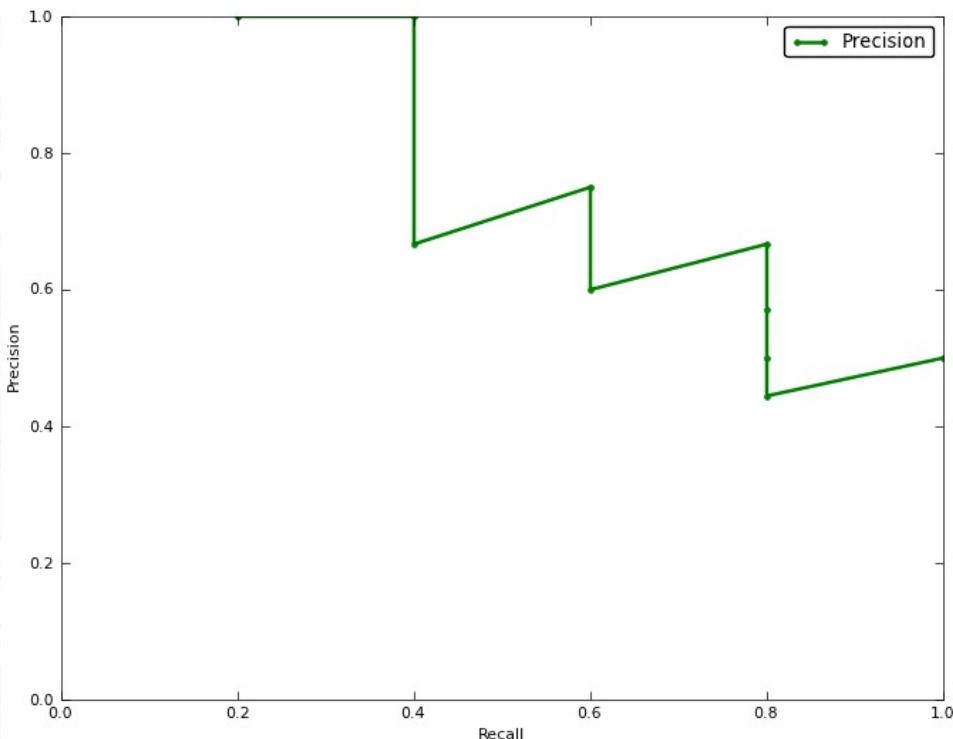
Как выглядит график для случайного порядка?

Пример расчета PR-кривой в задаче определения самолетов

Результат ранжирования



Порог	Precision	Recall
Top 1	100%	20%
Top 2	100%	40%
Top 3	66%	40%
Top 4	75%	60%
Top 5	60%	60%
Top 6	66%	80%
Top 7	57%	80%
Top 8	50%	80%
Top 9	44%	80%
Top 10	50%	100%



Чувствительность (sensitivity)

$$\text{Sensitivity} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- Чувствительность измеряет, насколько восприимчива наша модель (или диагностический тест) к обнаружению искомого, положительного класса.
- Она отвечает на вопрос: "Если пациент на самом деле болен, какова вероятность того, что тест это почувствует и окажется положительным?"
- Чувствительность = Полнота (recall)

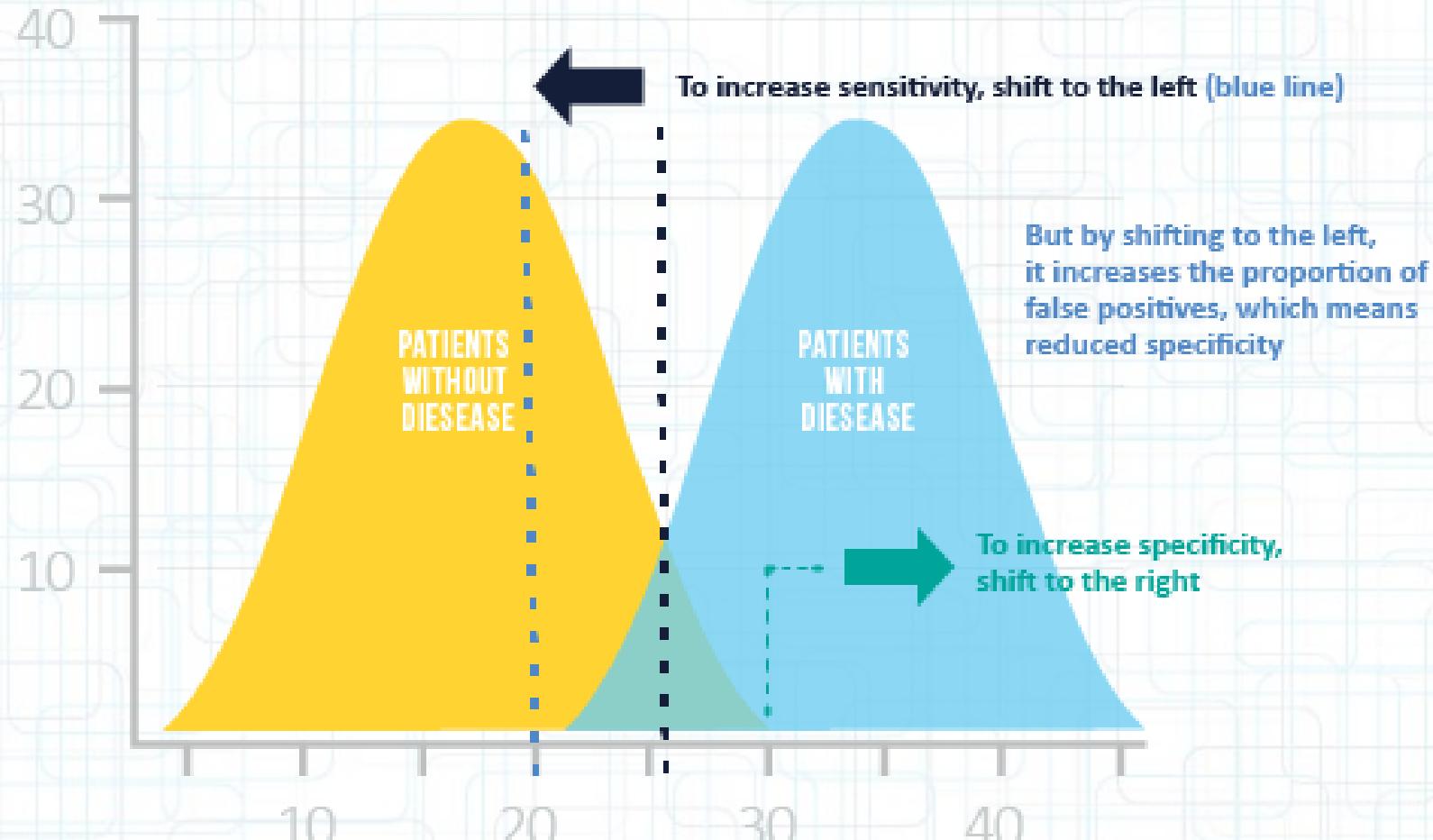
Специфичность/селективность (specificity)

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

- Диагностический тест будет иметь 100% чувствительность (sensitivity) если он положительный у любых людей. Нужна еще одна характеристика, которая показывает, насколько диагностический тест правильно классифицирует здоровых.
- Специфичность/селективность отвечает на вопрос: для какой доли здоровых людей тест отрицательный?
- Специфичность - это та же полнота, но для класса "-"

Порог задает баланс между полнотой для класса 1 и класса 0

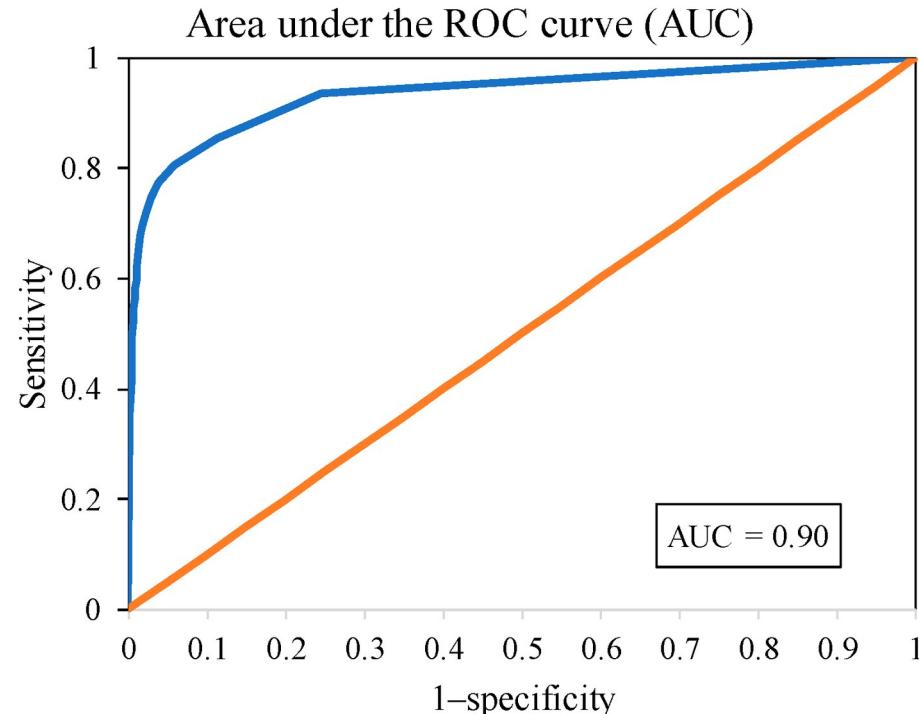
SENSITIVITY AND SPECIFICITY



гистограммы вероятностей класса "+" для объектов, которые на самом деле принадлежат классу "-" (желтая) и классу "+" (синяя)

Чувствительность (sensitivity) и специфичность (specificity)

Так как Специфичность = $1 - FPR$, то ROC-кривая также отражает баланс между полнотой в классе 1 (sensitivity) и полнотой в классе 0 для разных значений порога

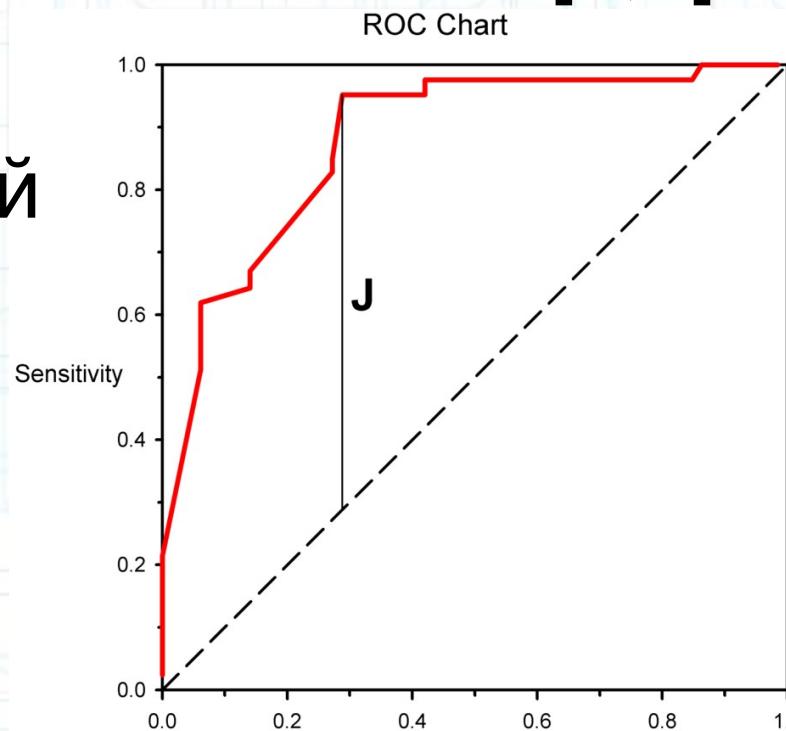


Проблемы метрик качества

- Популярная Accuracy является плохой оценкой для датасетов с несбалансированными классами (пример: 90% класса 1 и 10% класса 0, предсказываем всем 1)
- Точность (Precision) равна 1, если брать маленький порог в отсортированном ряду предсказаний
- Полнота (Recall) равна 1, если брать большой порог
- F1-мера = $2/(P^{-1}+R^{-1})$ – очень популярна, но не зависит от TN (True Negative)

Универсальные метрики

- Balanced accuracy – среднее значение между Sensitivity и Specificity (полнотой в классе 0 и 1). Меняется от 0.5 (random) до 1.
- Индекс Юдена = $Sensitivity + Specificity - 1$.
Почти то же самое, но меняется на $[0, 1]$.
- Равен высоте точки ROC над биссектрисой
- Реализация scikit-learn:
`balanced_accuracy_score`



Вероятностная постановка задачи

- $p(x,y)$ – неизвестная точная плотность распределения на $X \times Y$
- X^ℓ – выборка из случайных, независимых и одинаково распределенных прецендентов
- $p(X^\ell) = p((x_1, y_1), \dots, (x_\ell, y_\ell)) = p(x_1, y_1) \times \dots \times p(x_\ell, y_\ell)$
- $\varphi(x, y, \theta)$ – модель
- Принцип максимума правдоподобия:

$$L(\theta, X^\ell) = \prod_{i=1}^{\ell} \varphi(x_i, y_i, \theta) \rightarrow \max_{\theta}$$

Decision function

- Предположим, что мы нашли вероятность $p(y|x) = p(x,y)/p(x)$. Какое значение y нужно предсказать для заданного x ?
- Минимизация среднего риска:

$$a(x) = \arg \min_s E_y \mathcal{L}(s, y)$$

Несимметричные потери

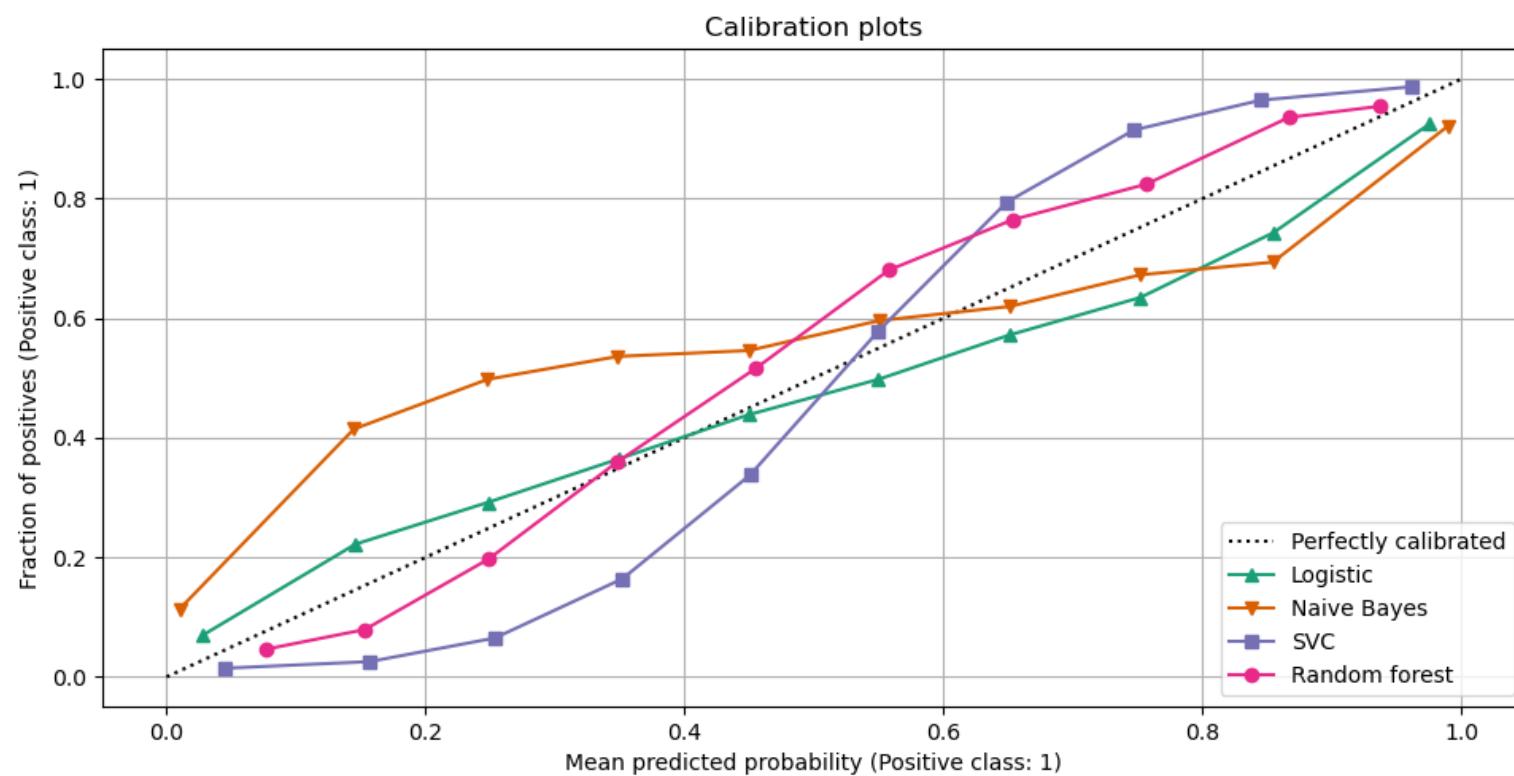
- Пример: нужно предсказать, сколько фирма потратит на рекламу (m - мало 50 т.р, B - много 200 т.р.)
- $L(a=m, y=B) = ???$
- $L(a=B, y=m) = ???$

Несимметричные потери

- Пример: нужно предсказать, сколько фирма потратит на рекламу (m - мало 50 т.р, B - много 200 т.р.)
- $L(a=m, y=B)$ — не сделаем подарочную скидку и от нас могут отказаться и не заказать много рекламы
$$L(m, B) = 200\text{т.р.} * 0.8 \text{ (верть отказа)}$$
- $L(a=B, y=m)$ — сделаем большую скидку, а закажут мало рекламы по низкой цене
$$L(B, m) = 50\text{т.р.} * 0.3 \text{ (скидка)}$$
- Допустим, что метод МО предсказал $p(m|x) = 0.9$ $p(B|x) = 0.1$
Скидку делать?

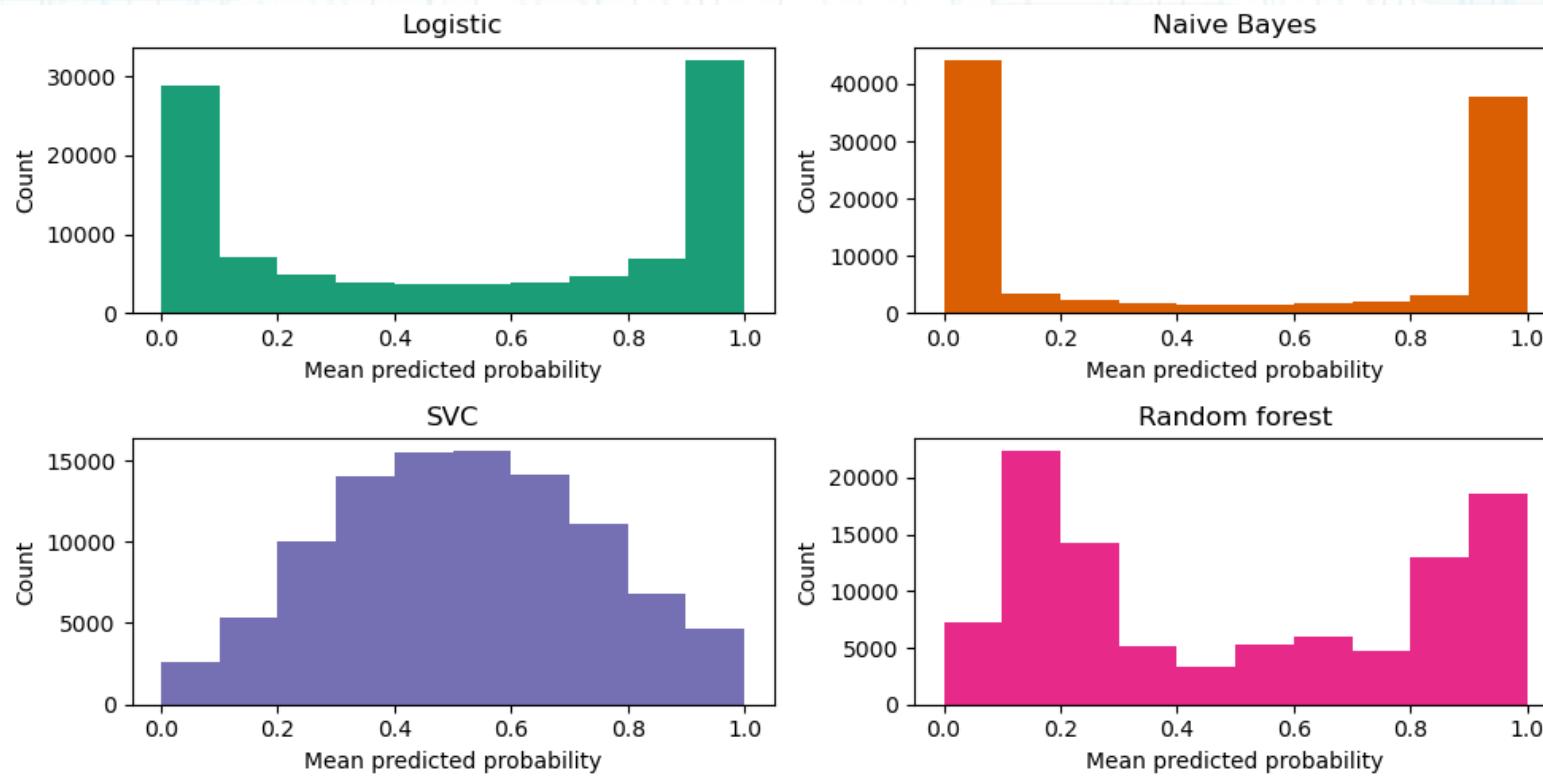
Калибровка вероятностей

- Классификаторы не всегда возвращают точные вероятности



Калибровка вероятностей

- Самые точные получаются у логистической регрессии. NB часто возвращает крайние значения: 0 и 1. RF и SVC ведут себя наоборот



Калибровка в Sklearn

- `model2 = CalibratedClassifierCV(model, cv=2, method="isotonic" или "sigmoid")`
- `sigmoid` менее точная, но не переобучается, предпочтительнее для маленьких выборок

