

```
In [2]: import IPython.display
        from IPython.display import HTML
        IPython.display.display_latex(IPython.display.Latex(filename="macros.tex"))
```

Метрики качества в задачах классификации

Бинарная классификация:

$$\alpha(x) = \text{sign}(f(x, \theta) - tr)$$

- f - уровень уверенности алгоритма
- tr - порог уверенности

accuracy

Пусть у нас задача бинарной классификации:

$$Y = \{0, 1\}$$

- P — количество правильных предсказаний
- N — количество не правильных предсказаний

$$accuracy = \frac{P}{P + N}$$

Проблема с accuracy: Несбалансированная выборка:

Предположим, что у нас есть некоторая обучающая выборка из 100 объектов, у которой 98 объектов принадлежат классу 0 и всего лишь 2 классу 1. Пусть наш алгоритм всегда предсказывает только класс объектов 0. Если посчитать accuracy такого классификатора, то $acc=0.98$. Можно ли его назвать хорошим?

ТОЧНОСТЬ И ПОЛНОТА precision and recall

		Real answers	
		1	0
Prediction of algorithm	1	TP	FP
	0	FN	TN

- TP - true positive answers
- FP - false positive answers (Type I error)
- TN - true negative answers
- FN - false negative answers (Type II error)

Итого имеем 2 метрики:

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

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

Тут нет симметрии по классам. Вы выбираете позитивный класс (H_0 гипотезу). Либо можно говорить о точности и полноте для класса "А" и для класса "В".

Полнота меряет, действительно ли мы все объекты класса 1 отнесли к этому классу. А точность меряет, насколько хорошо мы понимаем, что делаем и не относим все подряд объекты к классу 1. Понятно, что максимальные значения для точности и полноты это 1.

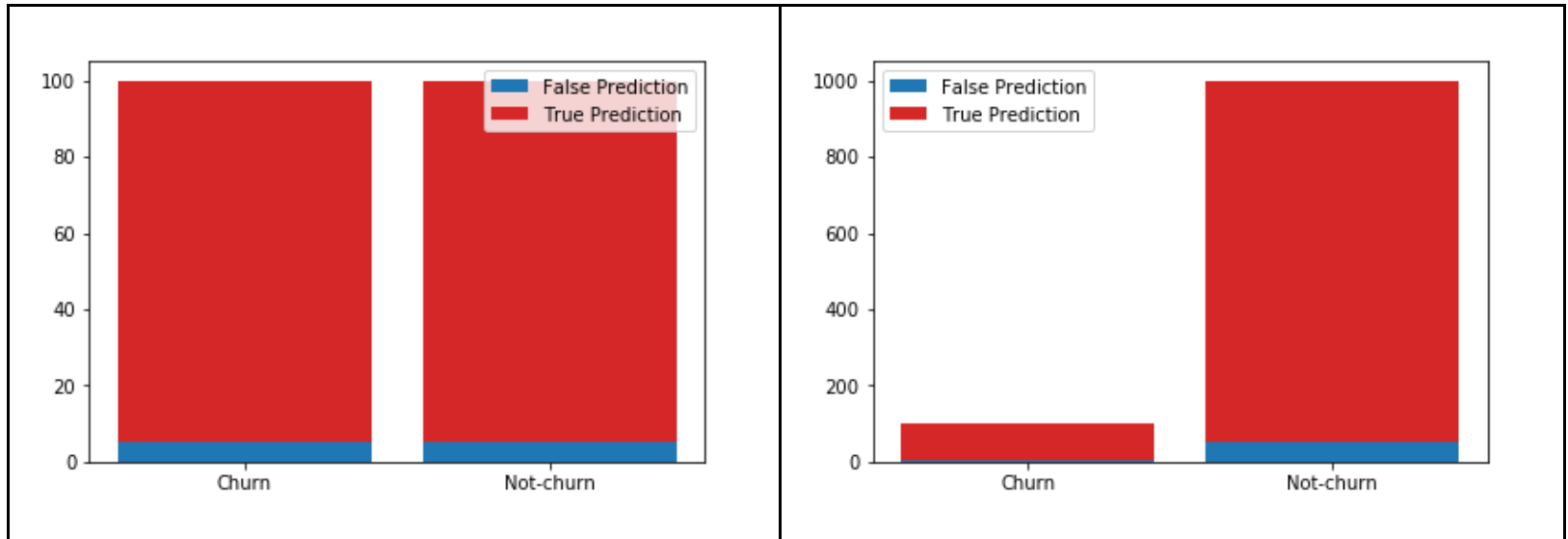
Precision и Recall это не 2 разных метрики, они всегда используются в паре.

$$\forall x \alpha(x) = 1 \Rightarrow \textit{Recall} = 100\%$$

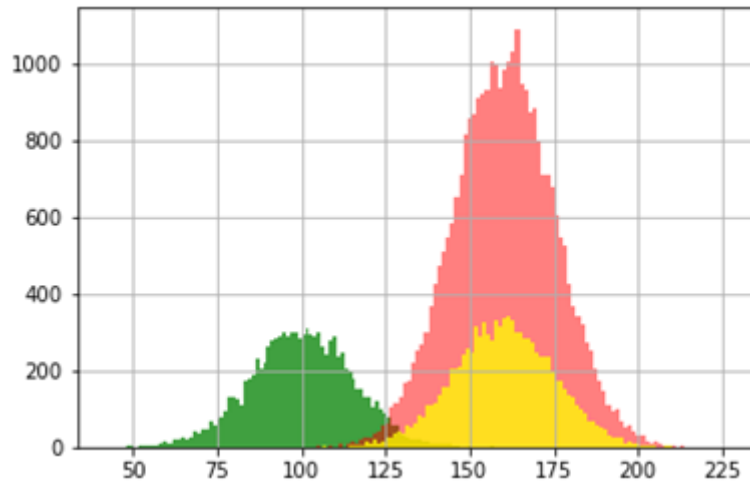
$$\forall x \alpha(x) = 0 \Rightarrow \textit{Type I error} = 0$$

Но в зависимости от задачи вы можете выбирать, что приоритетнее.

баланс классов



Посмотрим на графики. Пусть мы ошибаемся на в 5 % случаях в обе стороны. То есть 5% не оточных наша модель помечает как оточных, и 5% оточных как не оточных. На верхнем графике у нас сбалансированная выборка по 100 клиентов обоих классов. И все выглядит нормально. На нижней уже не оточных 1000 по сравнению со 100 отточными. Мы видим что теперь 5% неправильно распознанных не отточных это FP - это половина от всех отточных. Понятно что точность у нас сильно уменьшится с 95 % до примерно 62 %, и это при показателе 10 % оттока, а если отток и того меньше, точность упадет еще сильнее.



Придставим что мы научились разделять наши выборки на 2 класса. На самом деле нашли функцию которая переводит все характеристики в одну цифру, и класс "0" распределен по нормальному распределению с матожиданием, например 160 как на картинке, а "1" с матожиданием 100. Тут эти распределения с дисперсией в 15. Допустим у нас сбалансированная выборка и "0" это зеленые а "1" это желтые. Мы видим что есть пересечение, и там мы не можем однозначно определить класс, соответственно это будут наши ошибки. Пересечение зеленого распределения с желтым не такое большое, а теперь мы увеличим количество «желтых» классов всего в 3 раза – красное распределение. Видно, что теперь пересечение стало больше, но что более важно теперь больше «зеленых» клиентов попали в неопределенную зону. Доля "1", которых мы можем отделить стала гораздо меньше, а если потом мы увеличим выборку не в 3 а в большее число раз, то там останется совсем мало чисто зеленой области.

Пусть вы обучили классификатор, пусть он предсказывает в среднем правильно (правильно, это когда объект класса 0 относится к 0, а 1 к 1) с вероятностью $P = 0.8$. Возьмем выборку, где 50 объектов относятся к классу 0, 50 к классу 1. Посчитаем PR.

$$E(TP) = P \times N_1 = 0.8 \times 50 = 40$$

$$E(FP) = (1 - P) \times N_0 = 0.2 \times 50 = 10$$

$$E(FN) = (1 - P) \times N_1 = 0.2 \times 50 = 10$$

$$E(Precision) = E(TP / (TP + FP)) = 40 / 50 = 0.8$$

$$E(Recall) = E(TP / (TP + FN)) = 40 / 50 = 0.8$$

А теперь возьмем тот же классификатор, но будем мерить на другой выборке. Там объектов нулевого класса 50, а объектов первого класса 10. И посчитаем все то же самое

$$E(TP) = P \times N_1 = 0.8 \times 10 = 8$$

$$E(FP) = (1 - P) \times N_0 = 0.2 \times 50 = 10$$

$$E(FN) = (1 - P) \times N_1 = 0.2 \times 10 = 2$$

$$E(Precision) = E(TP / (TP + FP)) = 8 / 18 = 0.44444$$

$$E(Recall) = E(TP / (TP + FN)) = 8 / 10 = 0.8$$

Упр: Классификатор "монетка". На разбалансированной выборке. Несимметричная монетка на сбалансированной/разбалансированной выборке.

F мера (F1-score). гармоническое среднее между точностью и полнотой:

$$F = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Если нужен приоритет:

$$F_{\beta} = (\beta^2 + 1) * \frac{Precision * Recall}{\beta^2 * Precision + Recall}$$

- $0 < \beta < 1$ приоритет у точности
- $\beta > 1$ приоритет у полноты

Имеем:

P, N, TP, FP, TN, FN is function of the threshold

Пусть:

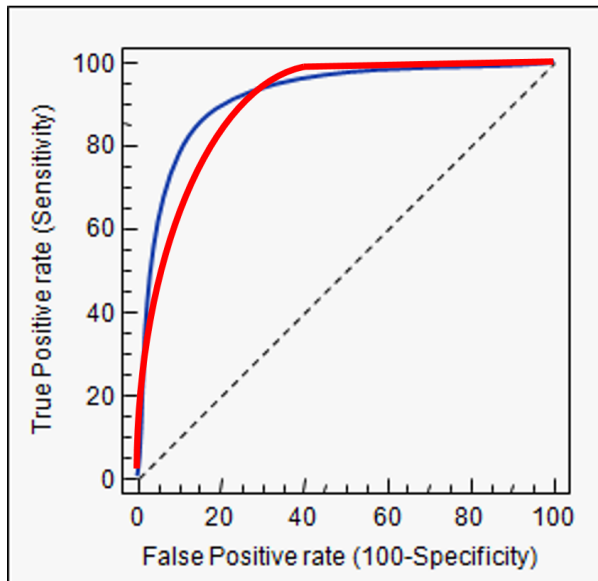
$$TPR = \frac{TP}{P} \quad \text{Recall}$$

$$FPR = \frac{FP}{N}$$

ROC(Receiver operating characteristic) curve:

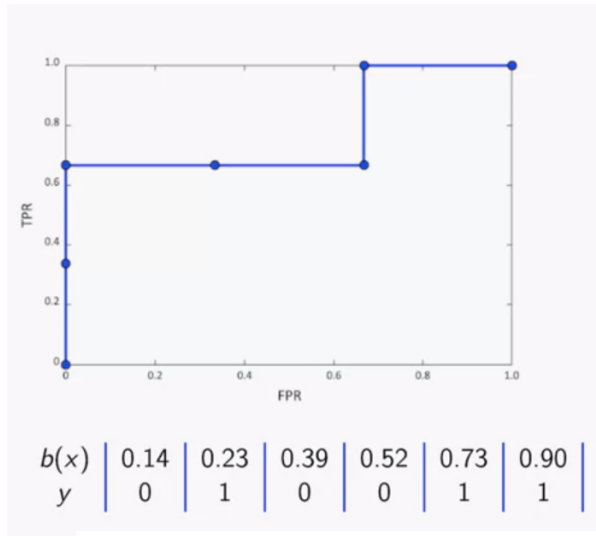
$Ox - FPR(tr)$

$Oy - TPR(tr)$



Монотонно возрастает Идеальная кривая проходит через точки $(0, 0)$, $(0, 1)$, $(1, 1)$.

ROCAUC(area under the curve) Площадь под кривой еще одна метрика

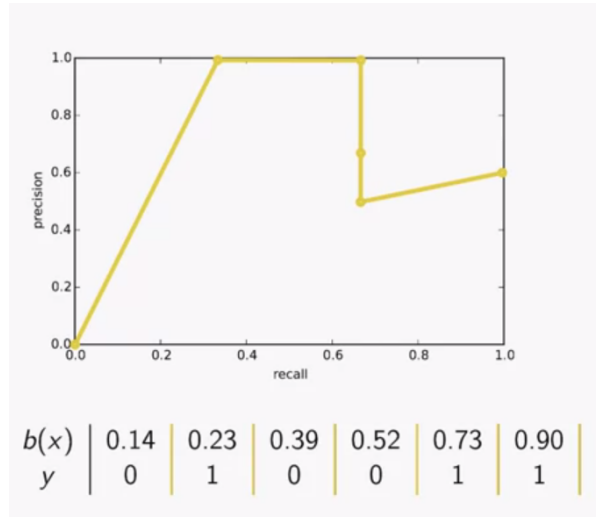


Данная кривая обладает следующими свойствами:

1. Левая точка всегда (0, 0)
2. Правая точка всегда (1, 1)
3. Выборка идеально разделима - пройдем через точку (0, 1)
4. Площадь от 1/2 до 1.
5. Чем больше площадь, тем лучше.

Precision-Recall curve

$$\text{Precision} = f_1(tr) \text{ and } \text{Recall} = f_2(tr)$$



Данная кривая обладает следующими свойствами:

1. Левая точка всегда (0, 0)
2. Правая точка зависит от баланса классов ($1, N_1 / (N_1 + N_0)$)
3. Выборка идеально разделима - пройдем через точку (1, 1)
4. ? Чем больше AUC (area under curve) - тем лучше?

не бинарная классификация

Всегда можно использовать один против всех

МИКРО И МАКРО УСРЕДНЕНИЕ

Пусть выборка состоит из 3 классов. Рассмотрим 3 двухклассовых задач, каждая из которых заключается в отделении своего класса от остальных (угадал свой класс, молодец, назвал любой другой - неправильно). Для каждой из них можно вычислить различные характеристики (TP, FP, и т.д.). При микро-усреднении сначала эти характеристики усредняются по всем классам, а затем вычисляется итоговая двухклассовая метрика — например, точность, полнота или F-мера. При макро-усреднении сначала вычисляется итоговая метрика для каждого класса, а затем результаты усредняются по всем классам.

	TP	FP	FN	TN
$y = 1$	900	120	100	930
$y = 2$	850	70	150	980
$y = 3$	10	100	40	1900

Посчитаем точность.

Микро усреднение.

- $\overline{TP} = 586.7$
- $\overline{FP} = 96.7$
- $\overline{FN} = 96.7$
- $\overline{TN} = 1270$

Точность 86%

Макро усреднение.

- класс 1 точность 88%
- класс 1 точность 92%
- класс 1 точность 9%

Точность 63%

Confusion Matrix

	Real answers			
		1	2	3
Predict of algorithm	1			
	2			
	3			

$$Precision_c = \frac{A_{cc}}{\sum_{i=1}^n A_{ci}}$$

$$Recall_c = \frac{A_{cc}}{\sum_{i=1}^n A_{ic}}$$