

Метрическая классификация

Матрица ошибок precision и recall

Skillbox

образовательная платформа

Правильные метки	Предсказанные метки
1	1
1	1
0	0
1	0
0	1

Правильные метки		Предсказанные метки	
1		1	
1		1	
0		0	
1		0	
0		1	

Матрица сопряжённости (Confusion matrix)

	Фактические значения	
Модель	+	–
+	TP	FP
–	FN	TN

Точность (Accuracy)

Точность (Accuracy) — число верно классифицированных объектов по модели:

$$Ac = \frac{TP + TN}{n}$$

Модель

Без модели

	Фактические значения	
Модель	+	—
+	5	10
—	5	90

$$A_c = \frac{90 + 5}{110} = 86,3$$

$$A_c = \frac{100}{110} = 90,9$$

Precision & Recall

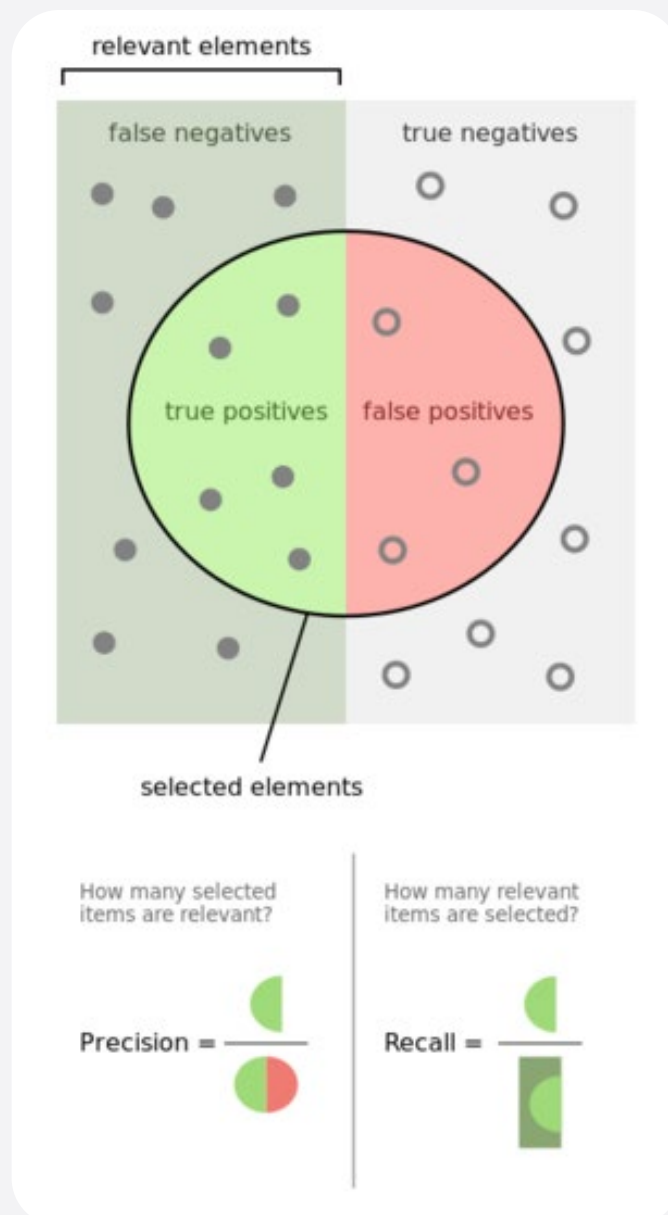
Точность (Precision) — доля положительных объектов среди тех, кого назвали положительными:

$$\text{Precision} = \frac{TP}{TP + FP} \times 100 \%$$

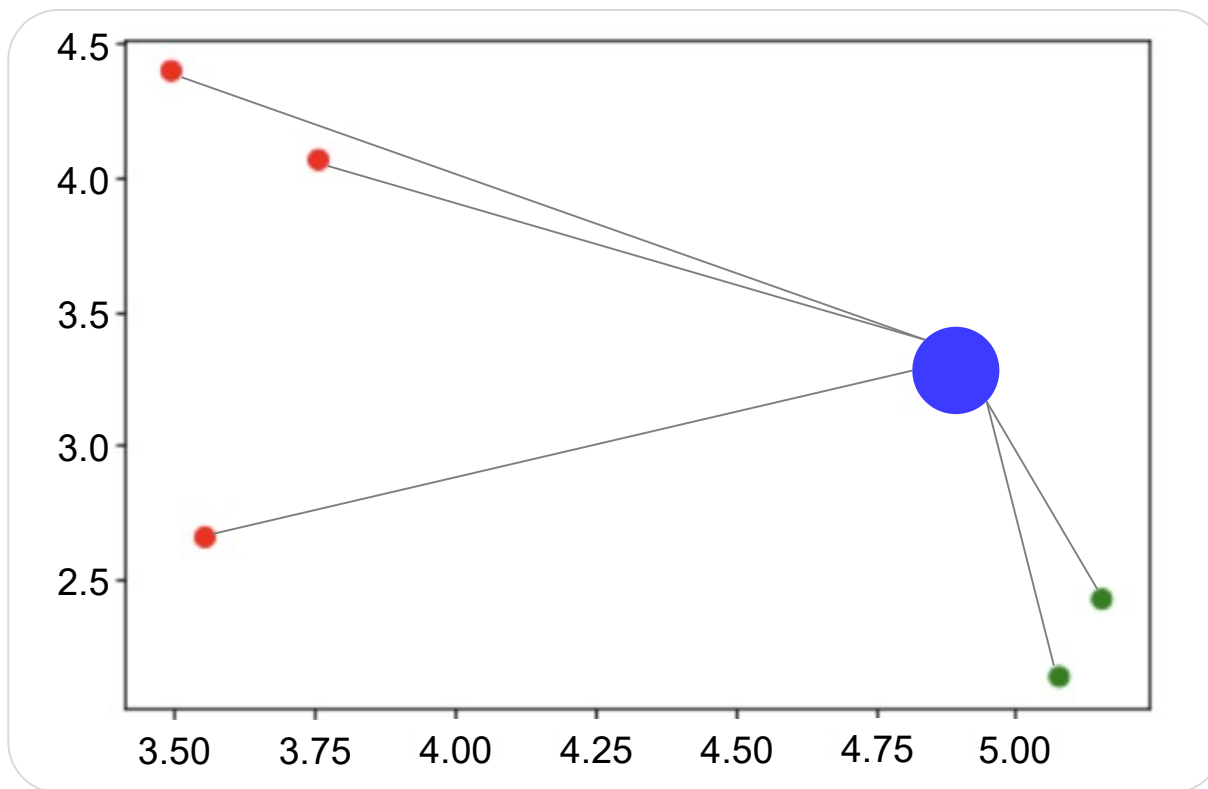
Полнота (Recall) — доля истинно положительных объектов:

$$\text{Recall} = \frac{TP}{TP + FN} \times 100 \%$$

Матрица сопряжённости (Confusion matrix)



Уровень отсечения (cut-off)



если 3 из 5 объектов в методе ближайших соседей
как бы проголосовали за 1-й класс

вероятность = $\frac{3}{5} = 0.6$

порог cut-off = 0.5, так как $0.6 > 0.5$, то объекту присваивается класс 1.

Уровень отсечения (cut-off)

Y	P	Y'
1	0.2	
1	0.4	
0	0.1	
0	0.3	
1	0.5	

Уровень отсечения 0.1

Y	P	Y'
1	0.2	1
1	0.4	1
0	0.1	0
0	0.3	1
1	0.5	1

Уровень отсечения 0.2

Y	P	Y'
1	0.2	0
1	0.4	1
0	0.1	0
0	0.3	1
1	0.5	1

Уровень отсечения 0.9

Y	P	Y'
1	0.2	0
1	0.4	0
0	0.1	0
0	0.3	0
1	0.5	0