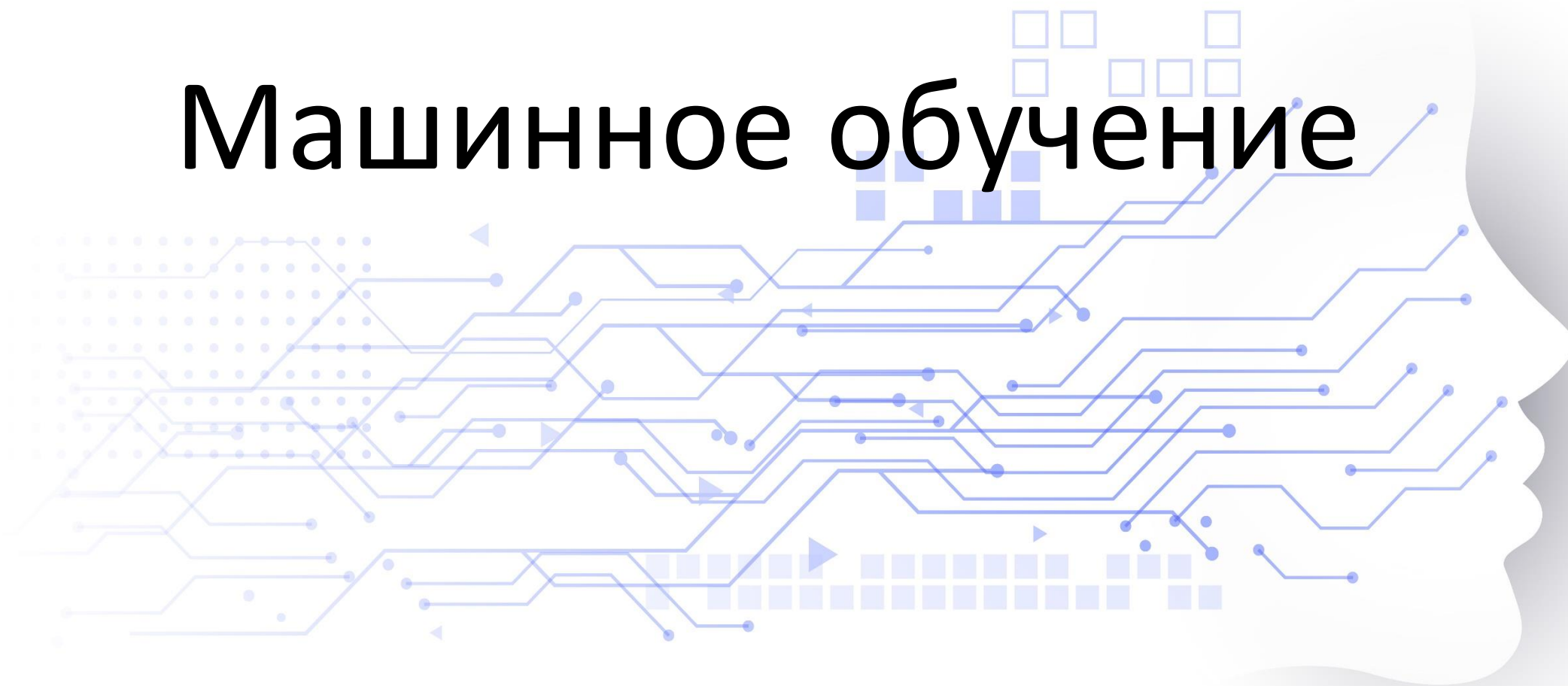
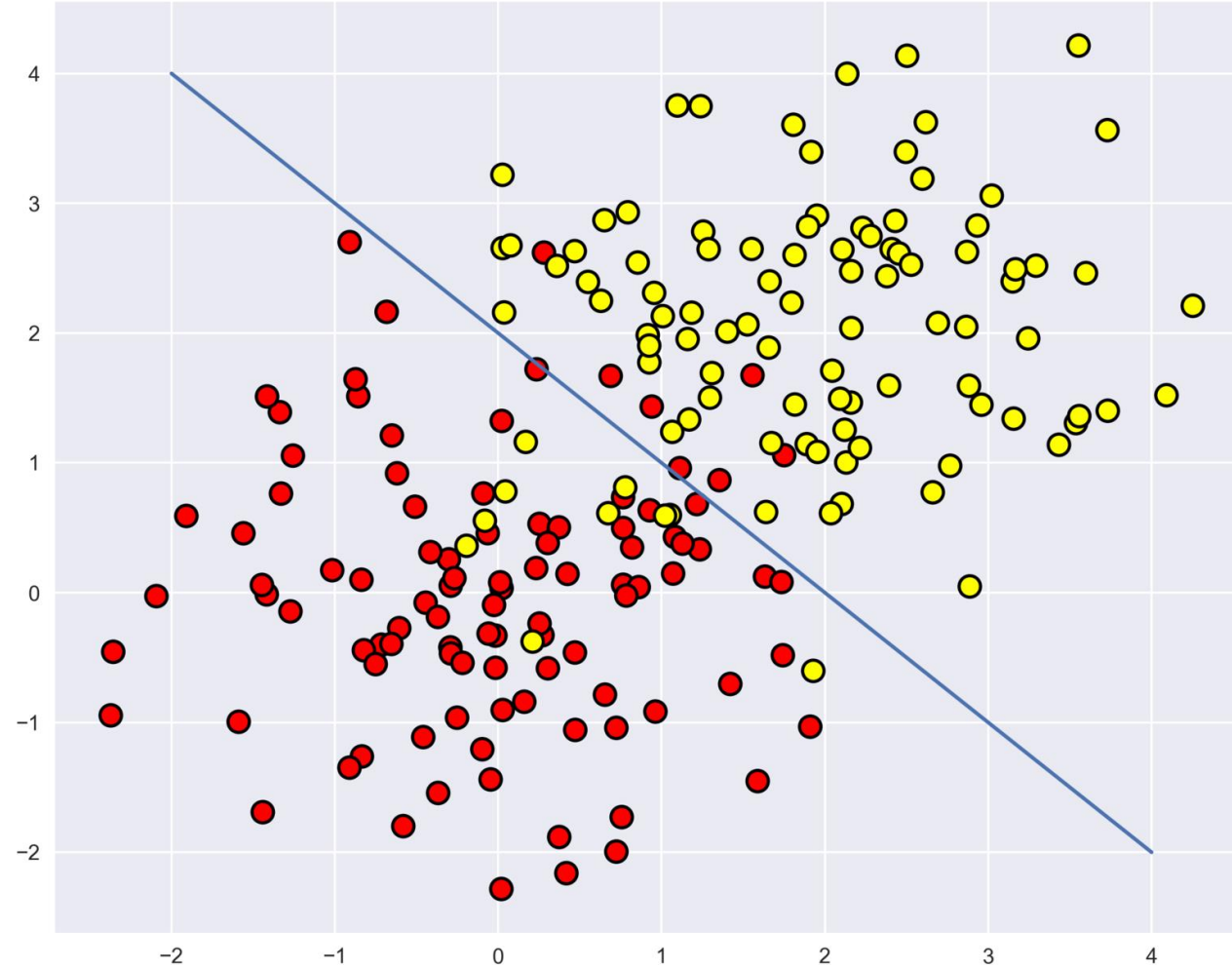


Машинное обучение



Классификация (Classification)



Бинарная классификация

Это метод машинного обучения, который позволяет разделить данные на две категории

$$y \in \{-1, 1\}$$

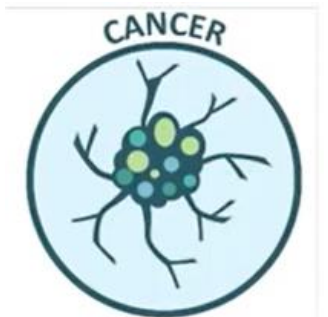
Идентификация спама (Да / Нет)



Обнаружение мошеннических операций (Да / Нет)



Классификация опухолей (Доброкачественные / Злокачественные)



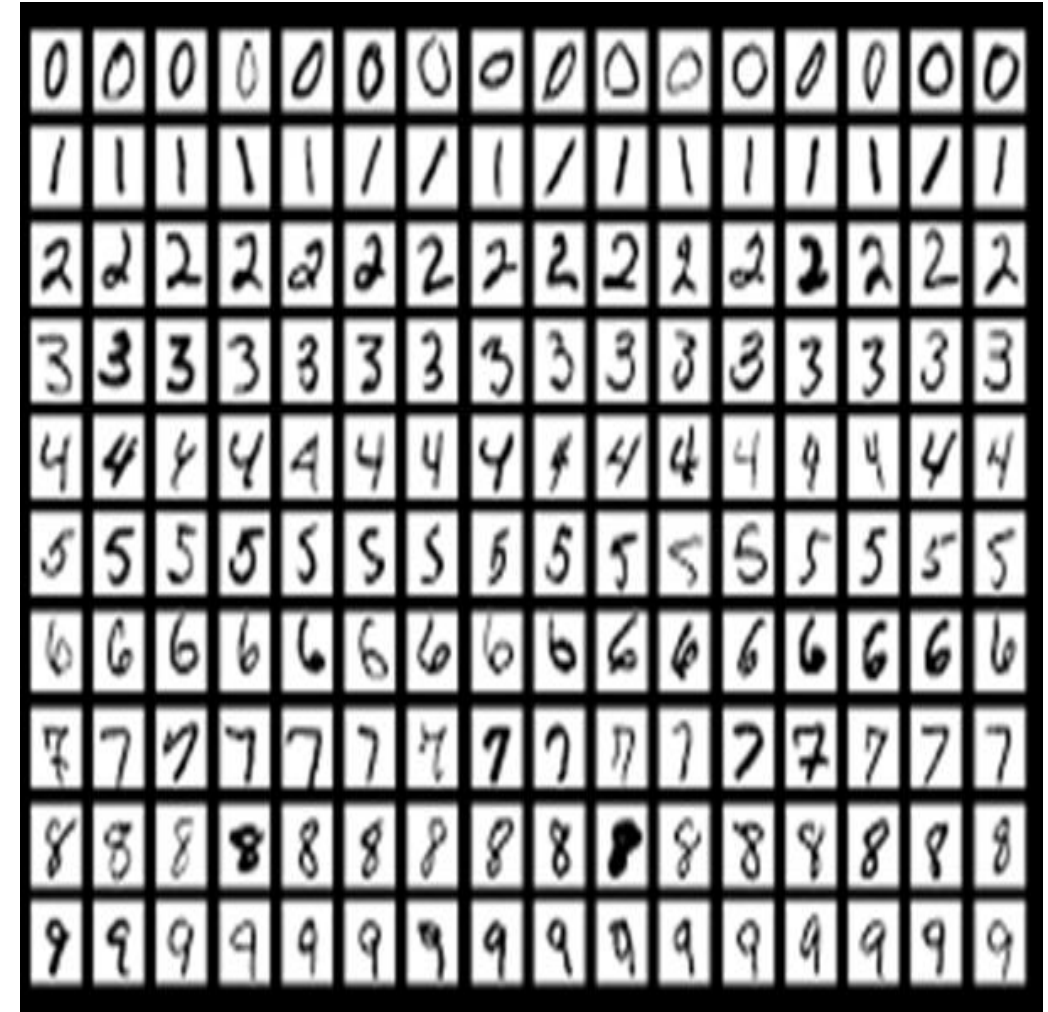
Мультиклассовая классификация

Это метод машинного обучения, который позволяет разделить данные на несколько категорий

$y \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

Вход: Изображение рукописного номера

Выход: Число



Бинарная классификация

$$Y \in \{-1, 1\}$$



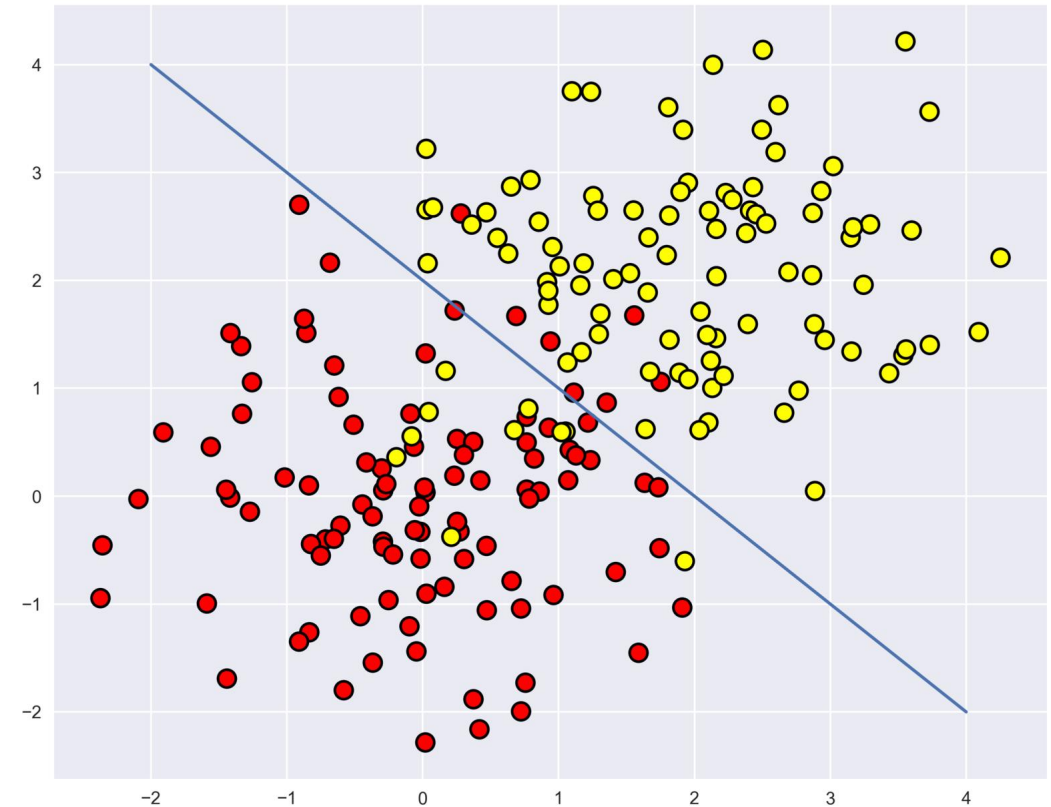
-1 Отрицательный класс



1 Положительный класс



$A(x)$ Должен возвращать одно из двух чисел



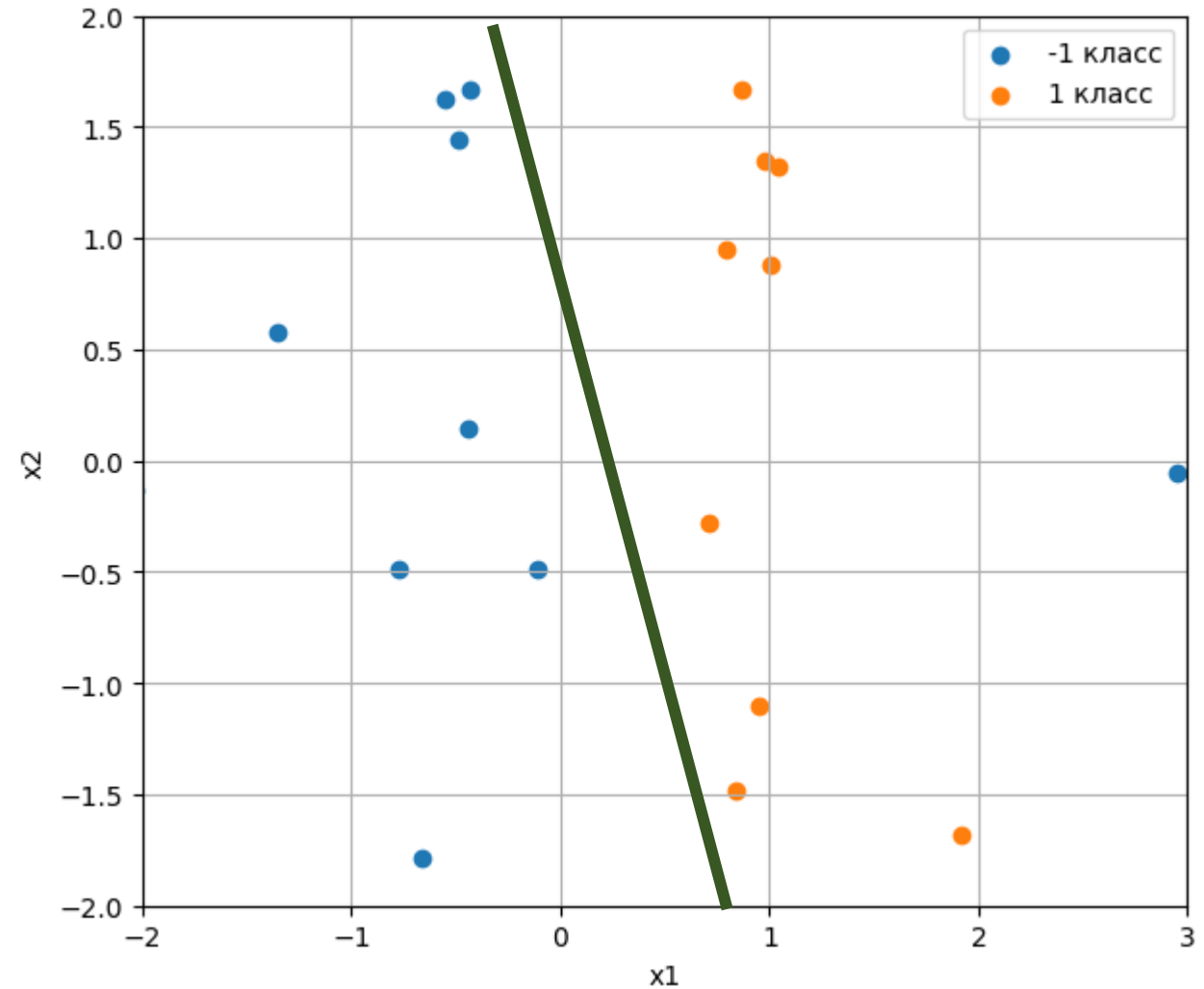
Классификация (Classification)

⚡ Линейная регрессия

$$a(x) = \theta_0 + \sum_{i=1}^d \theta_i x_i$$

⚡ Вещественное число

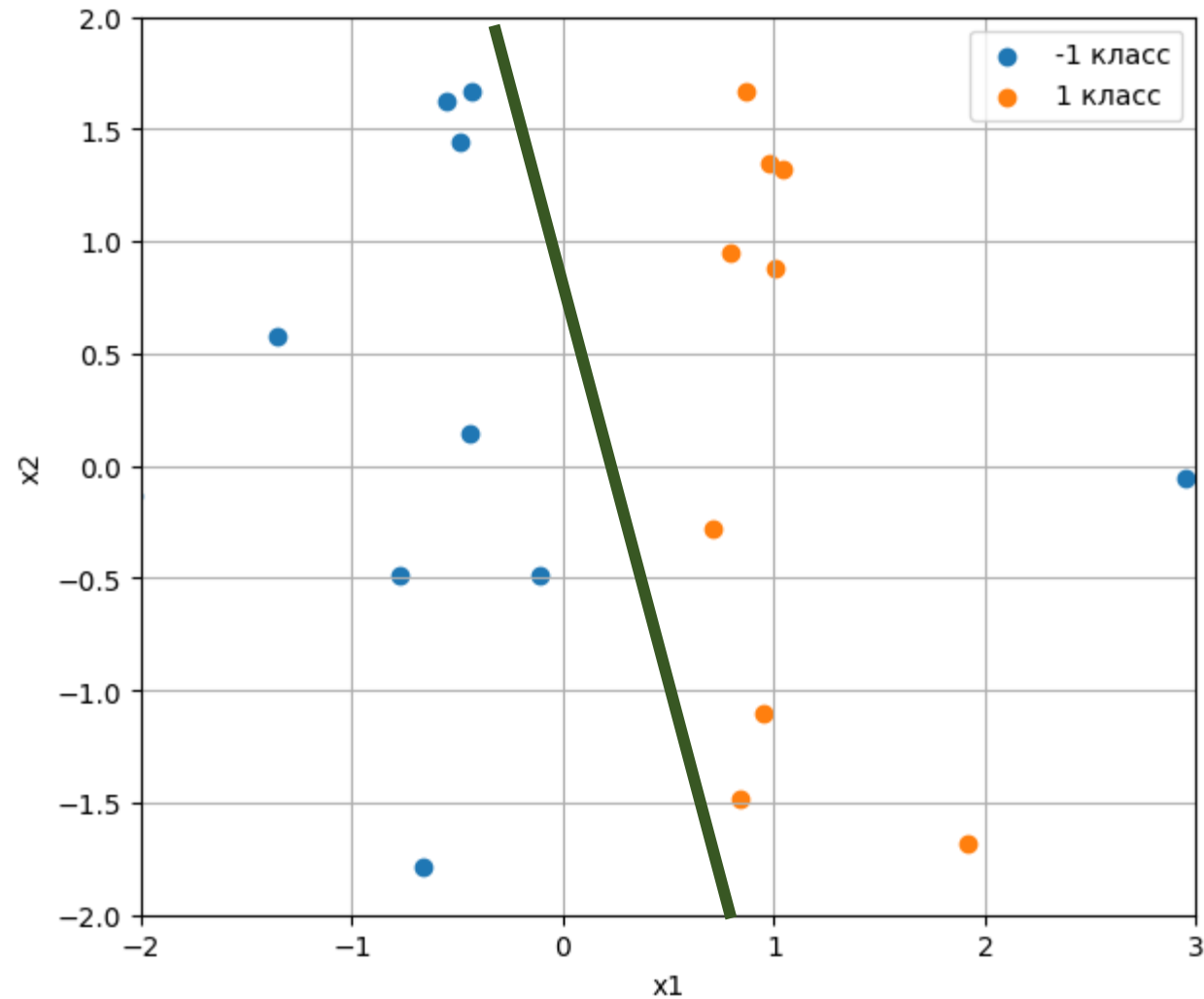
$$-\infty \leq a(x) \leq +\infty$$



Классификация (Classification)

$$a(x) = \theta_0 + \sum_{j=1}^d \theta_j x_j$$

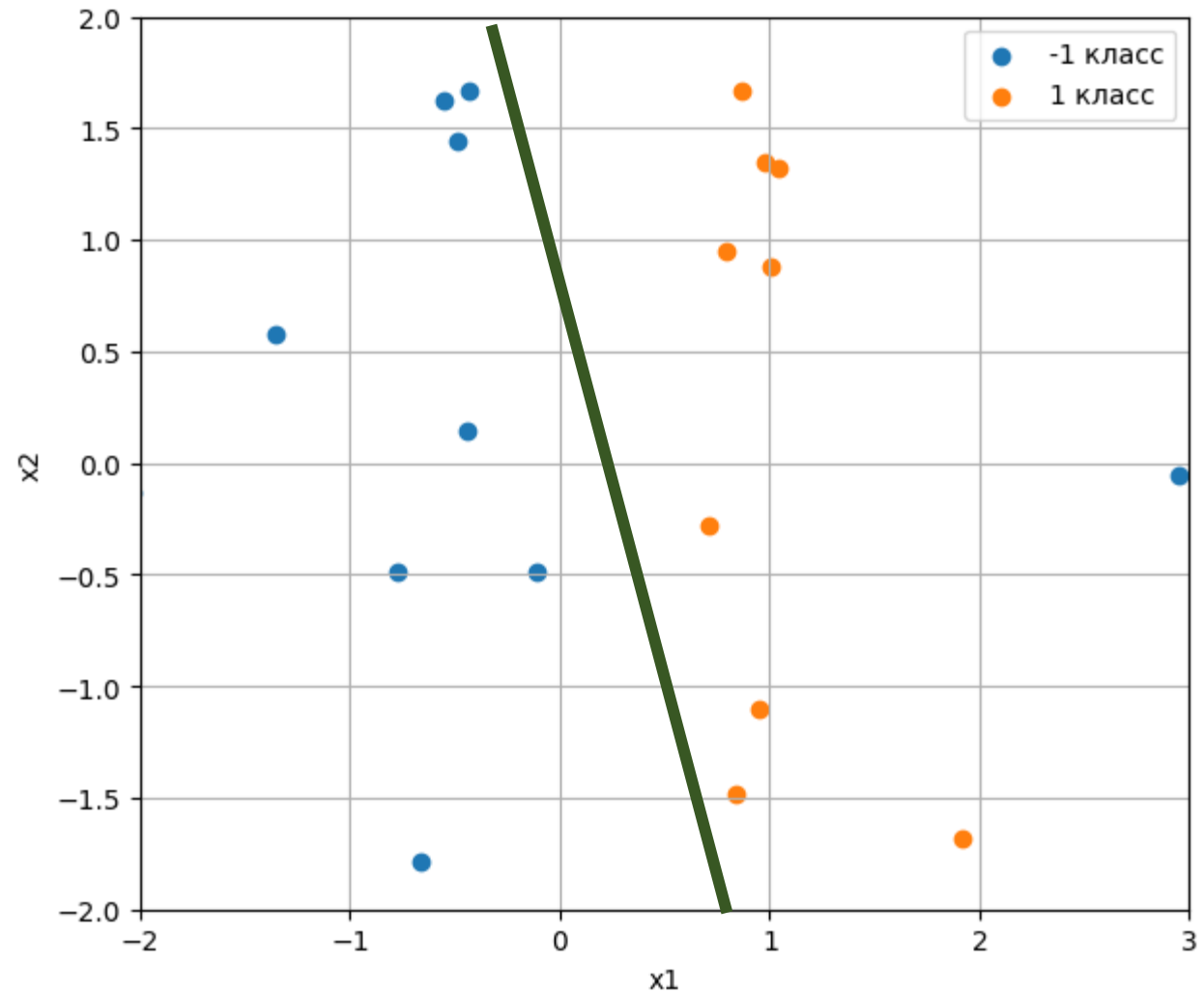
	Class -1	Class 1
0	-0.874904	1.308753
1	-0.413640	1.199750
2	-1.240367	0.653478
3	2.944008	1.246134
4	-0.228938	0.641666
5	-0.198273	0.540919
6	-2.065783	1.182851
7	-0.095081	1.580854
8	-0.211104	0.729224
9	-1.019285	0.979414



Классификация (Classification)

$$a(x) = \theta_0 + \sum_{j=1}^d \theta_j x_j$$

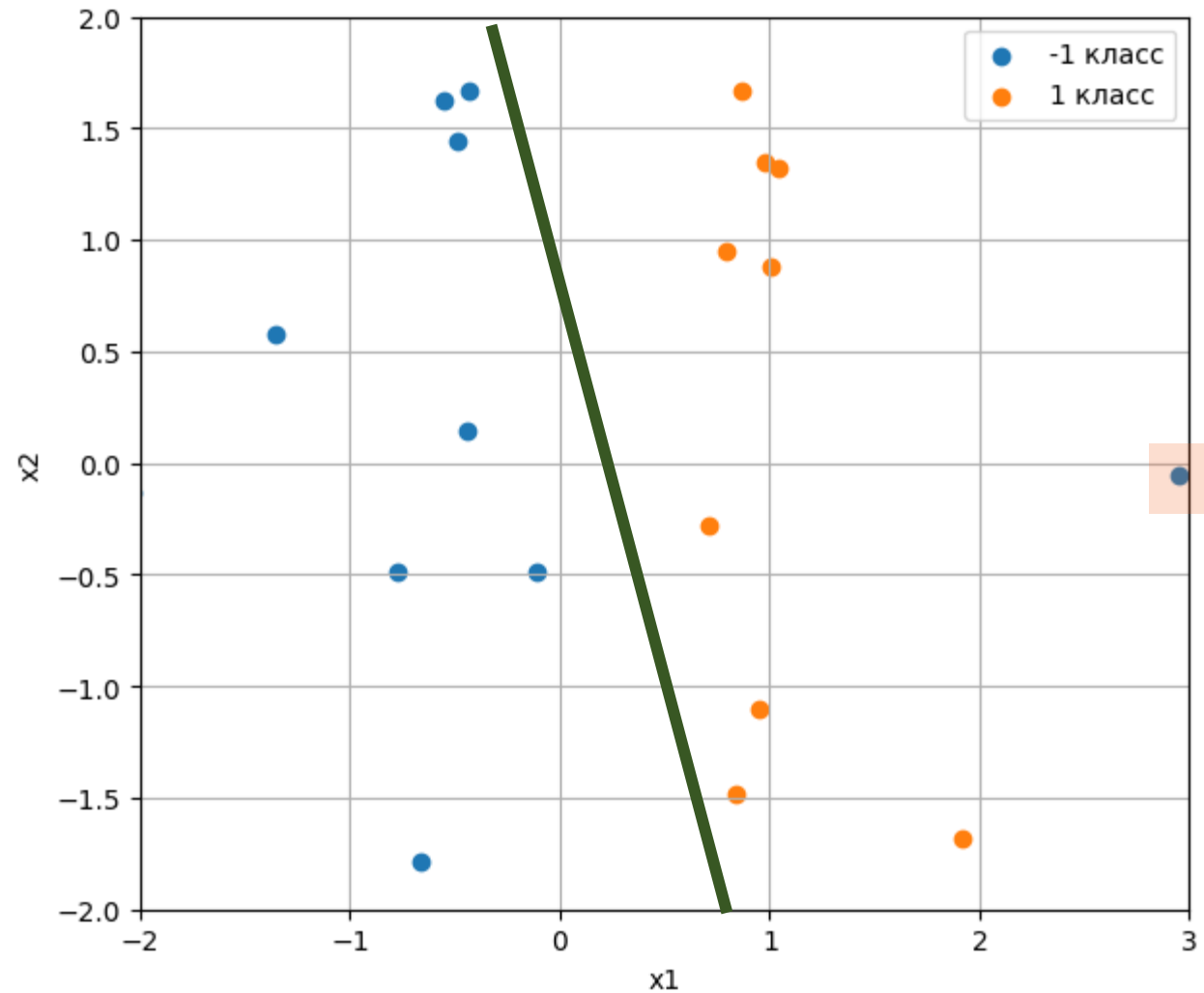
	Class -1	Class 1
0	-0.874904	1.308753
1	-0.413640	1.199750
2	-1.240367	0.653478
3	2.944008	1.246134
4	-0.228938	0.641666
5	-0.198273	0.540919
6	-2.065783	1.182851
7	-0.095081	1.580854
8	-0.211104	0.729224
9	-1.019285	0.979414



Классификация (Classification)

$$a(x) = \theta_0 + \sum_{j=1}^d \theta_j x_j$$

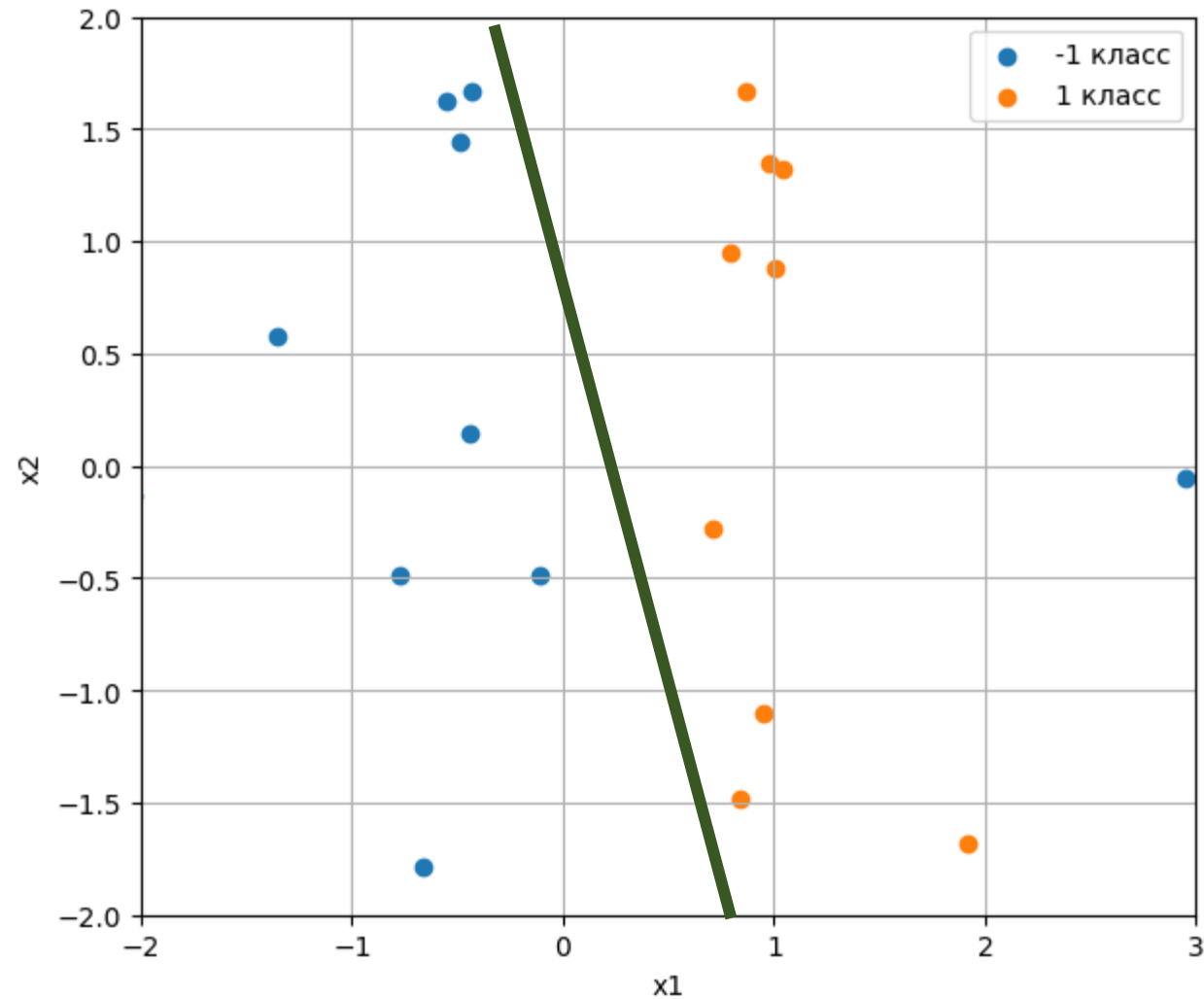
	Class -1	Class 1
0	-0.874904	1.308753
1	-0.413640	1.199750
2	-1.240367	0.653478
3	2.944008	1.246134
4	-0.228938	0.641666
5	-0.198273	0.540919
6	-2.065783	1.182851
7	-0.095081	1.580854
8	-0.211104	0.729224
9	-1.019285	0.979414



Классификация (Classification)

$$a(x) = \theta_0 + \sum_{j=1}^d \theta_j x_j$$

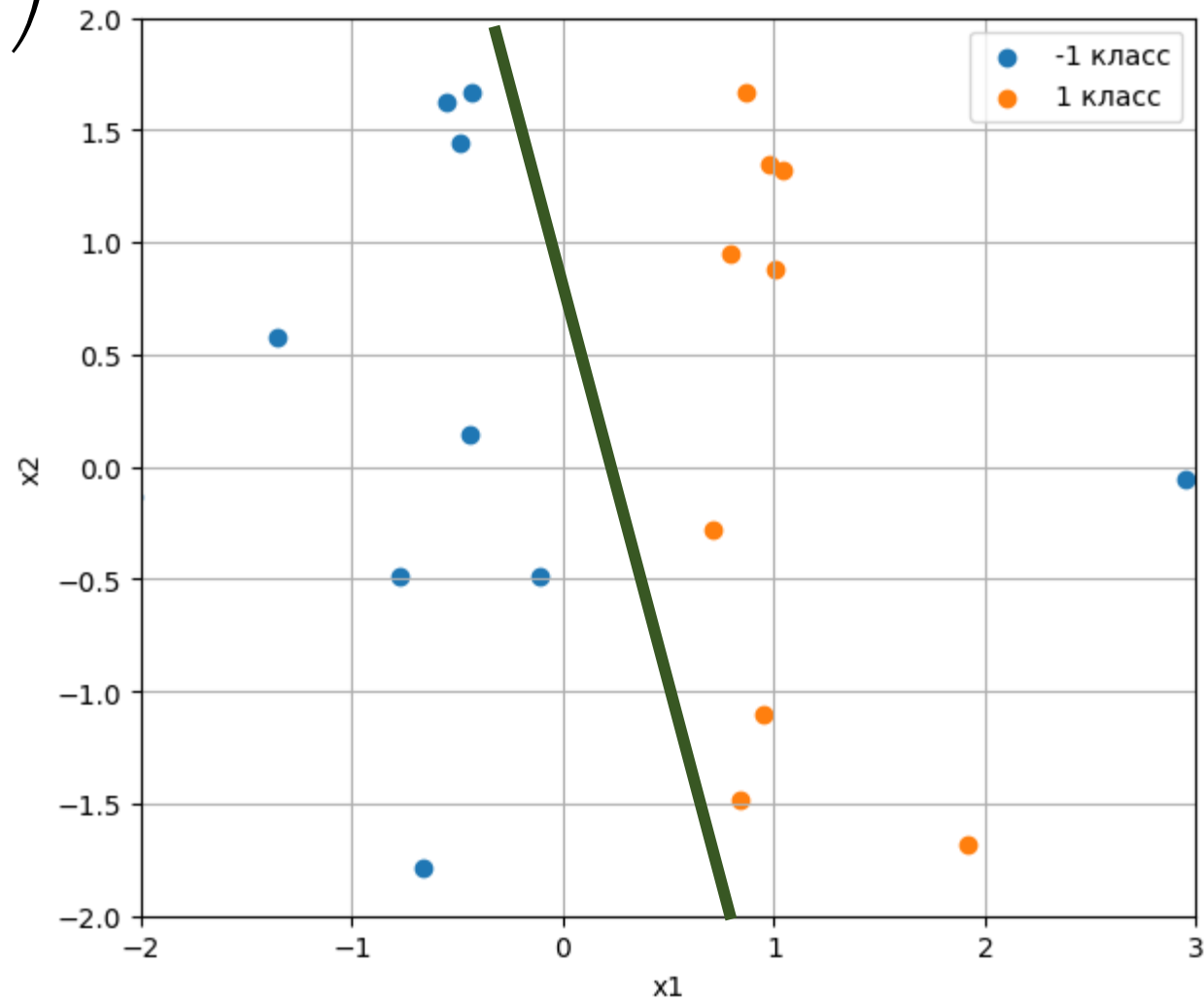
	Class -1	Class 1
0	-0.874904	1.308753
1	-0.413640	1.199750
2	-1.240367	0.653478
3	2.944008	1.246134
4	-0.228938	0.641666
5	-0.198273	0.540919
6	-2.065783	1.182851
7	-0.095081	1.580854
8	-0.211104	0.729224
9	-1.019285	0.979414



Классификация (Classification)

$$a(x) = \text{sign} \left(\theta_0 + \sum_{j=1}^d \theta_j x_j \right)$$

	Class -1	Class 1
0	-0.874904	1.308753
1	-0.413640	1.199750
2	-1.240367	0.653478
3	2.944008	1.246134
4	-0.228938	0.641666
5	-0.198273	0.540919
6	-2.065783	1.182851
7	-0.095081	1.580854
8	-0.211104	0.729224
9	-1.019285	0.979414



Классификация (Classification)

⚡ Линейная регрессия

$$a(x) = \theta_0 + \sum_{j=1}^d \theta_j x_j$$

$$-\infty \leq a(x) \leq +\infty$$

$$a(x) = \langle \theta, x \rangle$$

⚡ Бинарная классификация

$$a(x) = \textit{sign}(\langle \theta, x \rangle)$$

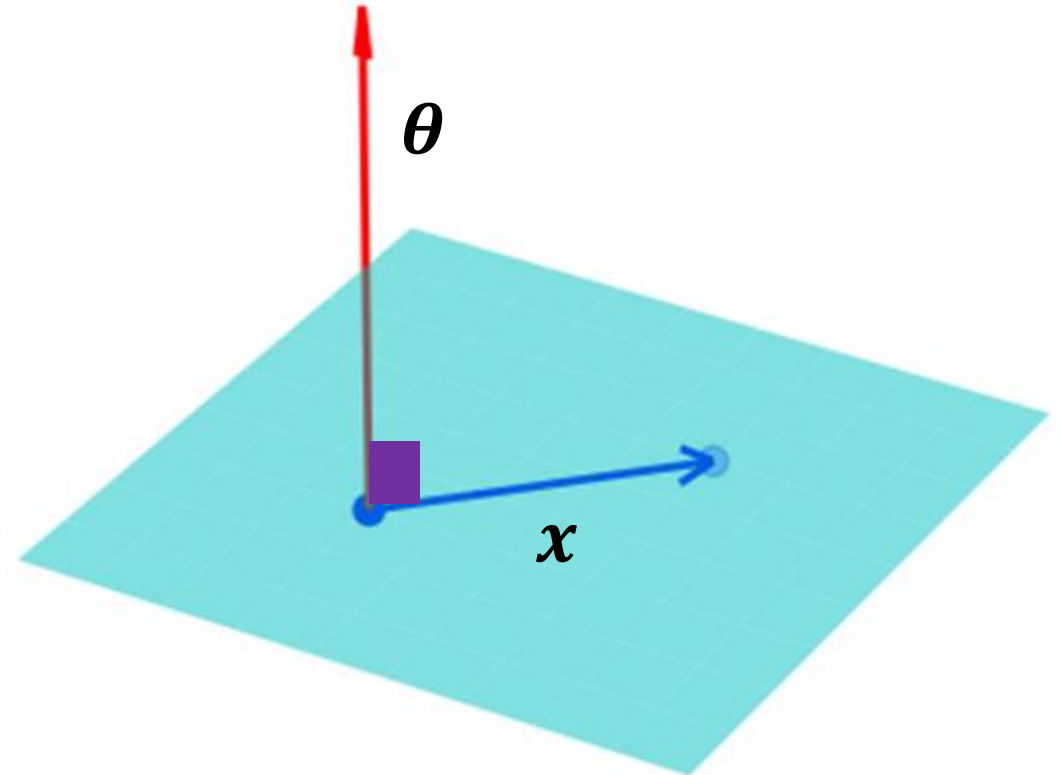
$$a(x) = \begin{cases} -1 \\ +1 \end{cases}$$

Классификация (Classification)

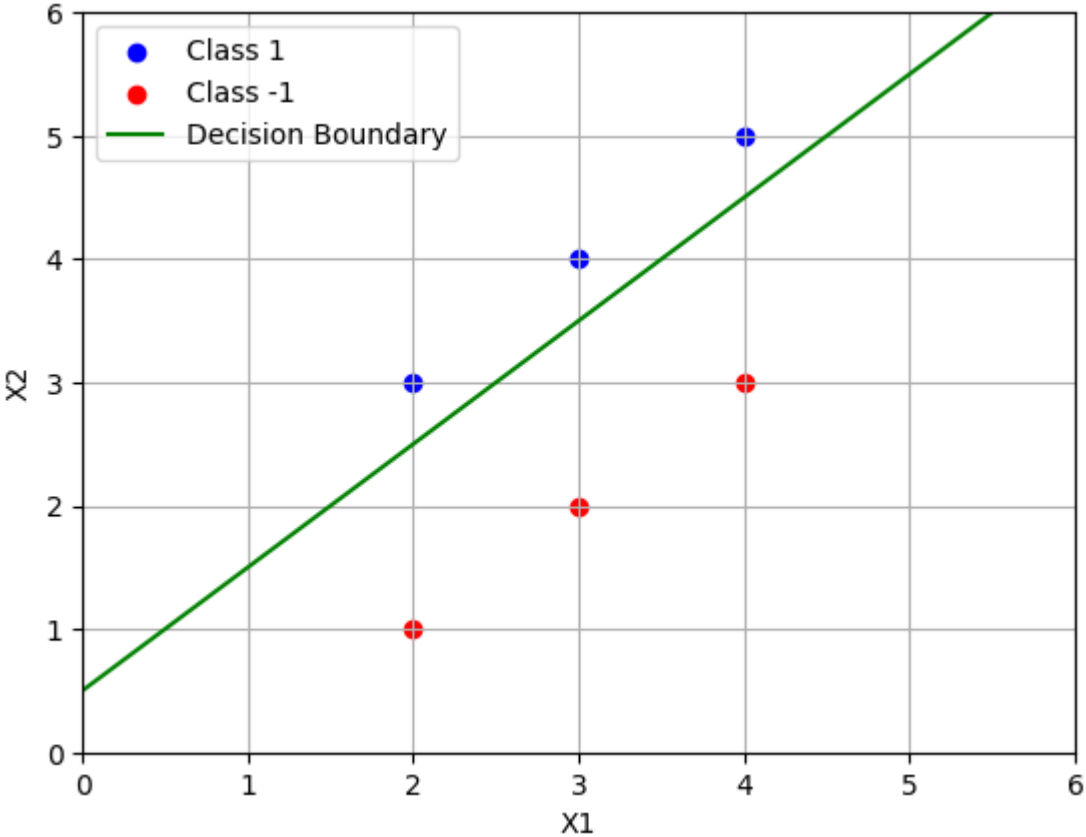
⚡ Уравнение гиперплоскости

$$\langle \theta, x \rangle = 0$$

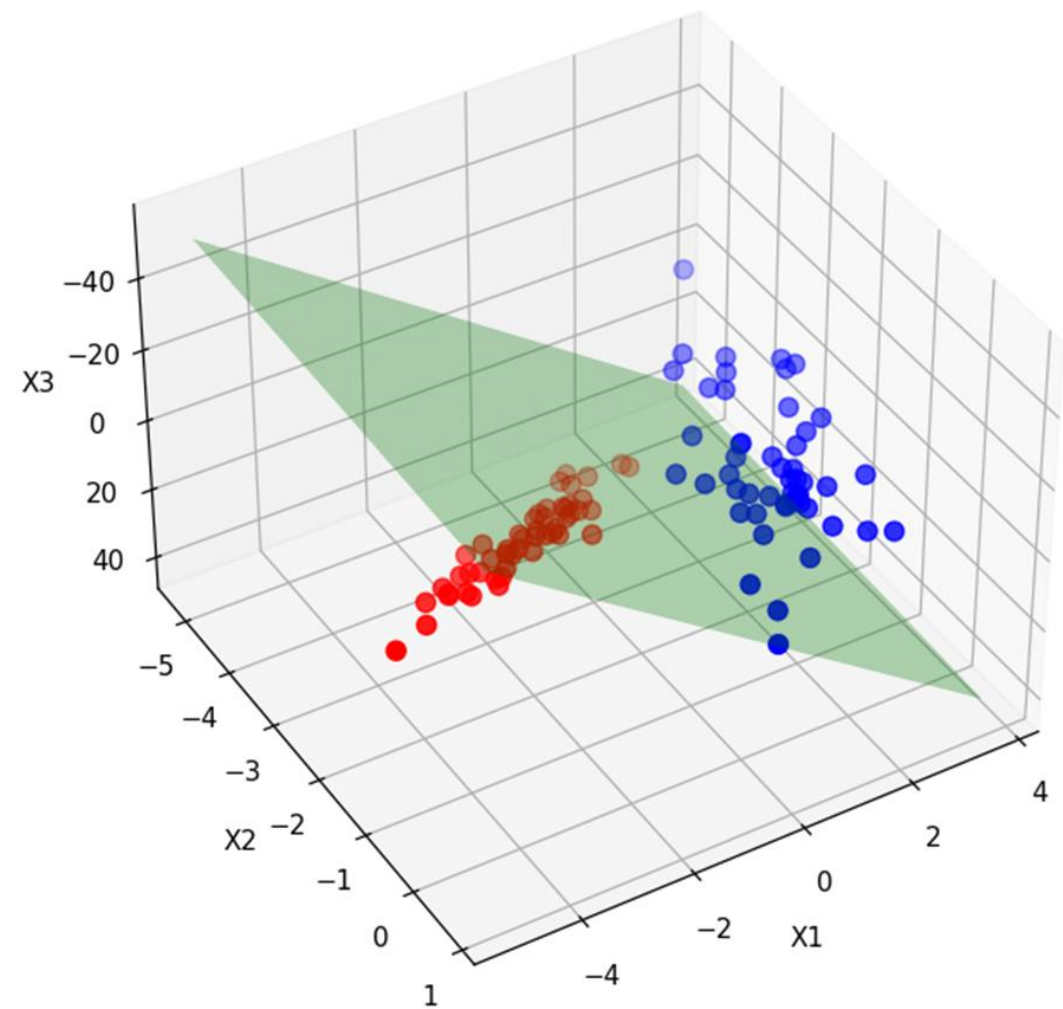
$$\langle \theta, x \rangle = |\theta| \cdot |x| \cdot \cos \alpha = 0$$



Классификация (Classification)



Классификация (Classification)



Классификация (Classification)

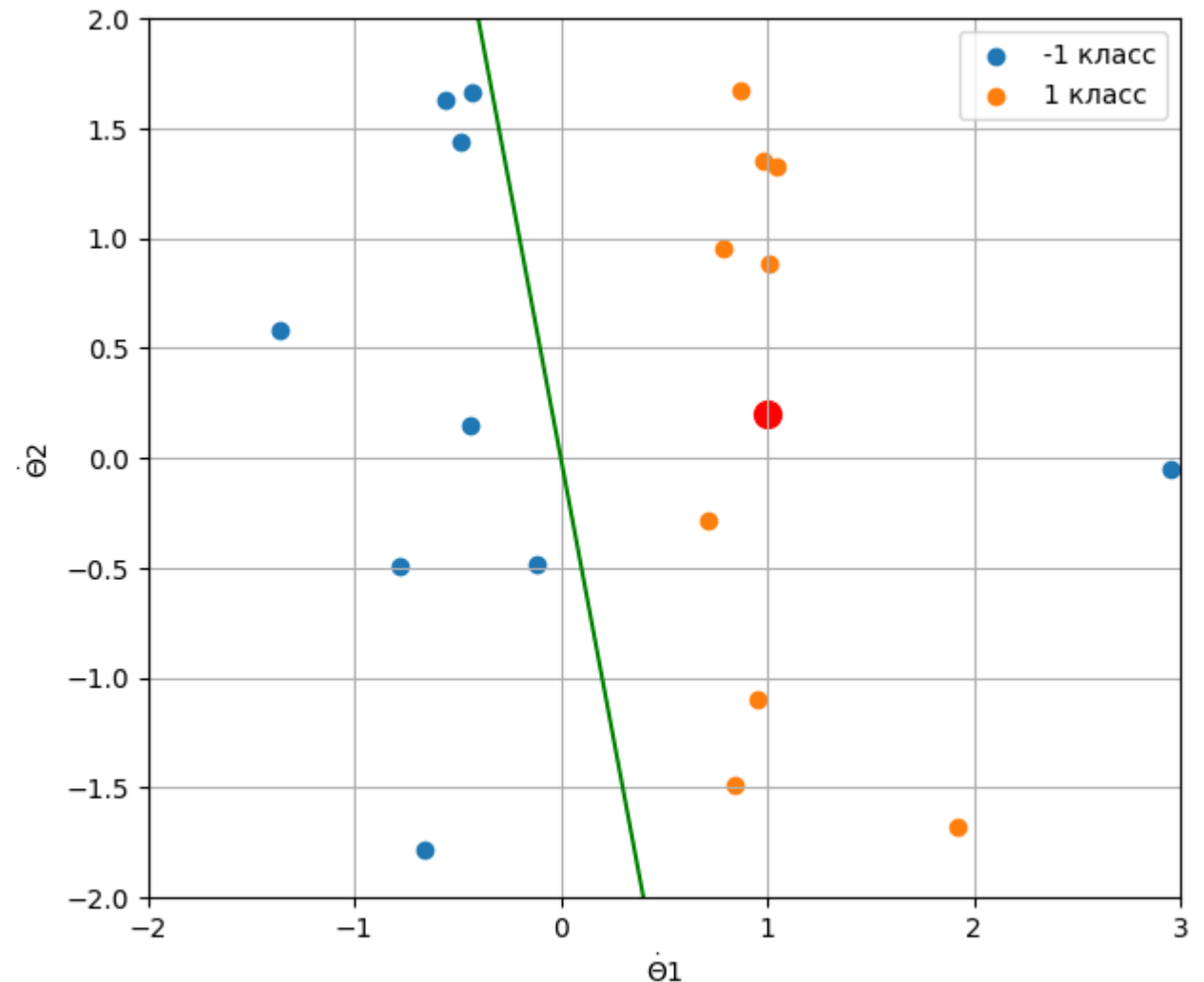
🔗 Уравнение гиперплоскости

Если x **лежит** на гиперплоскости, то
 $\langle \theta, x \rangle = 0$

$\langle \theta, x \rangle < 0$ объект **слева** от неё

$\langle \theta, x \rangle > 0$ объект **справа** от неё

$$\langle \theta, x \rangle = 0$$



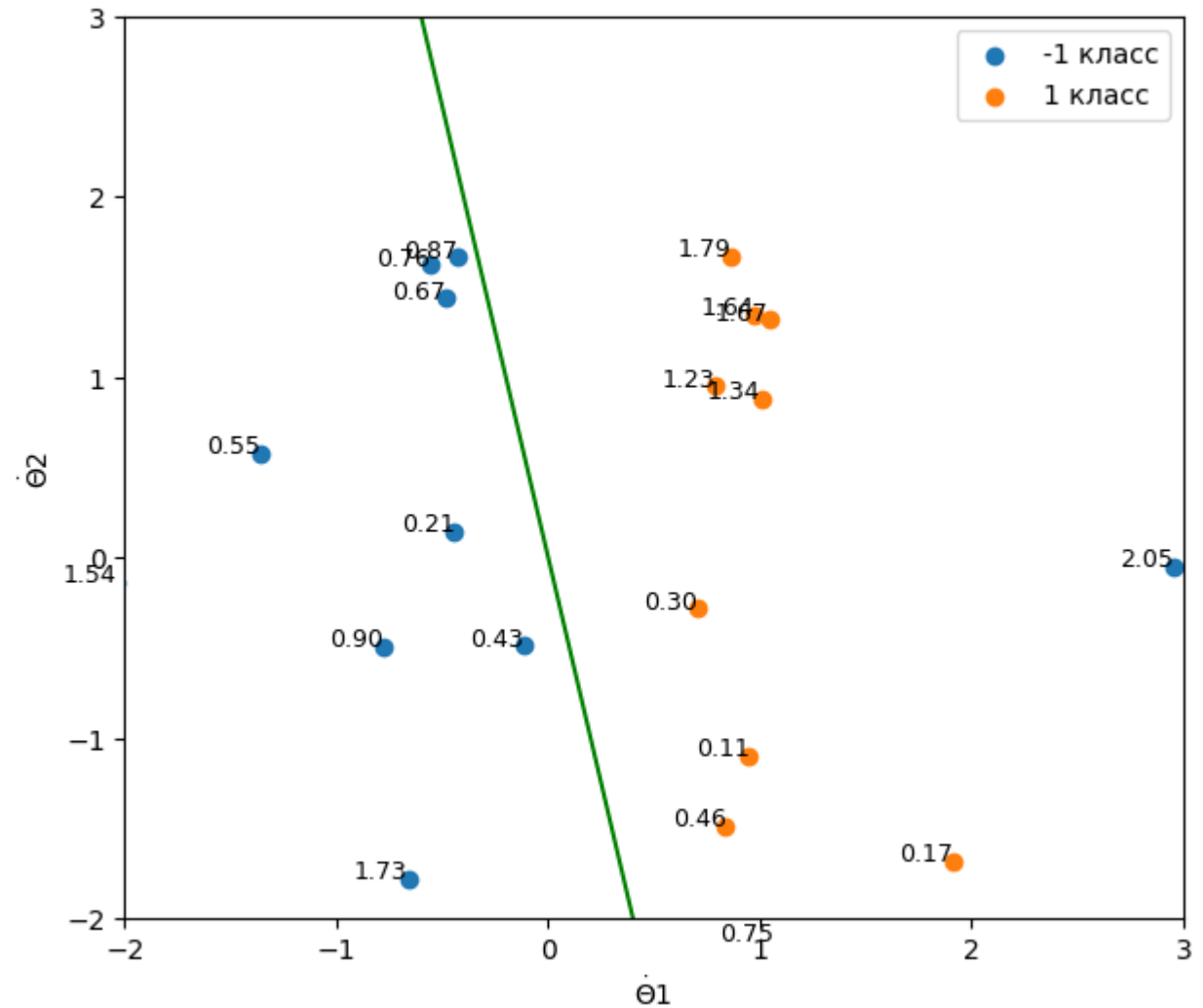
Обучение модели

Классификация (Classification)

Отступы (Margin)

Расстояние от точки до гиперплоскости $\langle \theta, x \rangle = 0$

$$\frac{|\langle \theta, x \rangle|}{\|\theta\|}$$



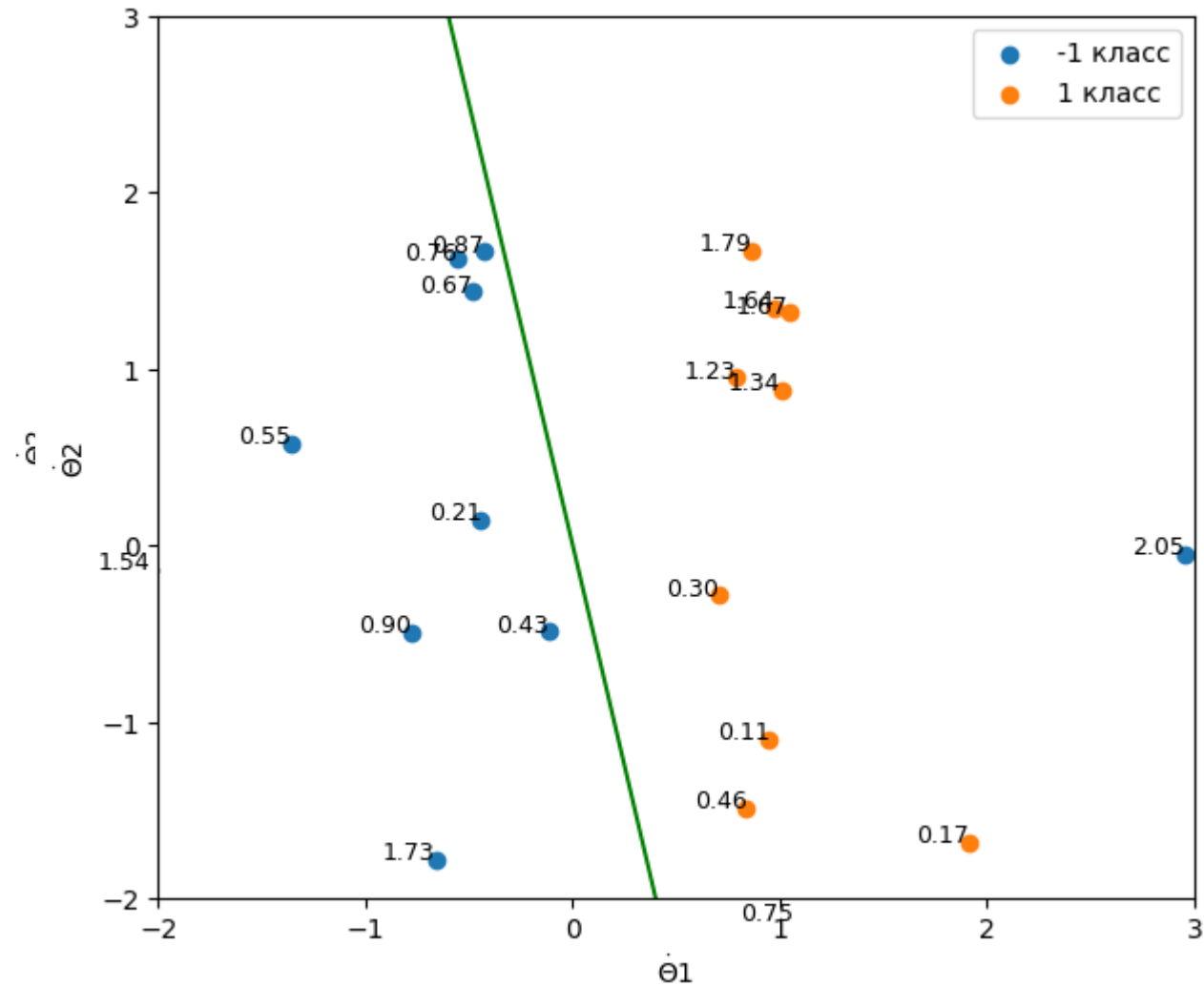
Классификация (Classification)

Отступы (Margin)

$$M_i = y_i \langle \theta, x_i \rangle$$

$M_i > 0$ классификатор дает верный ответ

$M_i < 0$ классификатор ошибается



Функция потерь в классификации

Бинарная функция потерь

$$L(y, a) = [a \neq y]$$

Функционал ошибки - доля ошибок

$$Q(a, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} [a(x_i) \neq y_i]$$

Обучение линейных классификаторов

⚡ Доля ошибок для линейного классификатора

$$a(x) = \textit{sign}(\langle \boldsymbol{\theta}, x \rangle)$$

Функционал ошибки:

$$Q(a, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} [a(x_i) \neq y_i] = \frac{1}{\ell} \sum_{i=1}^{\ell} [\textit{sign}(\langle \boldsymbol{\theta}, x_i \rangle) \neq y_i]$$

Недифференцируемая функция

Отступы для линейного классификатора

Функционал ошибки:

$$Q(\theta, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} [\text{sign}(\langle \theta, x_i \rangle) \neq y_i]$$

Альтернативная запись:

$$\mathbf{M}_i = y_i \langle \theta, x_i \rangle$$

$$Q(\theta, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} [y_i \langle \theta, x_i \rangle < 0]$$

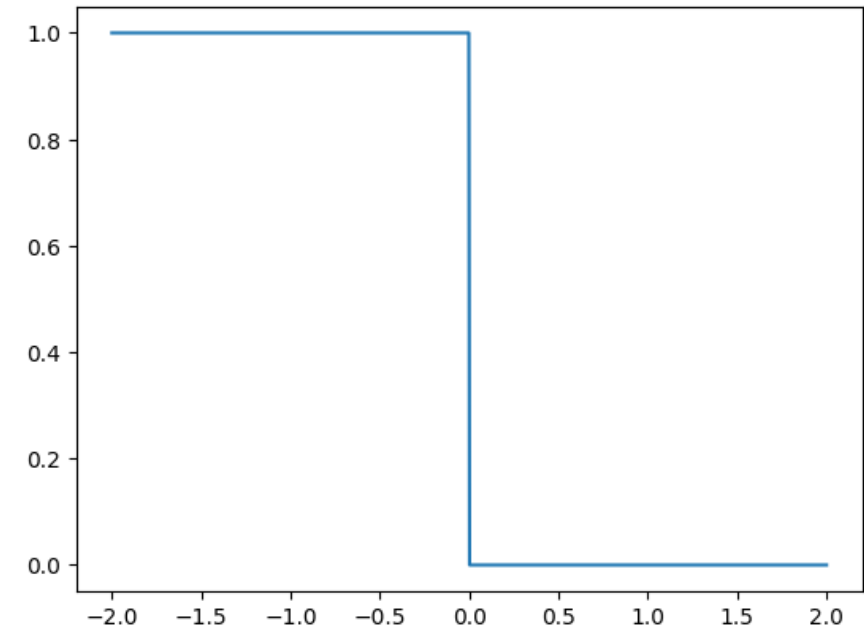
Обучение линейных классификаторов

⚡ Отступы для линейного классификатора

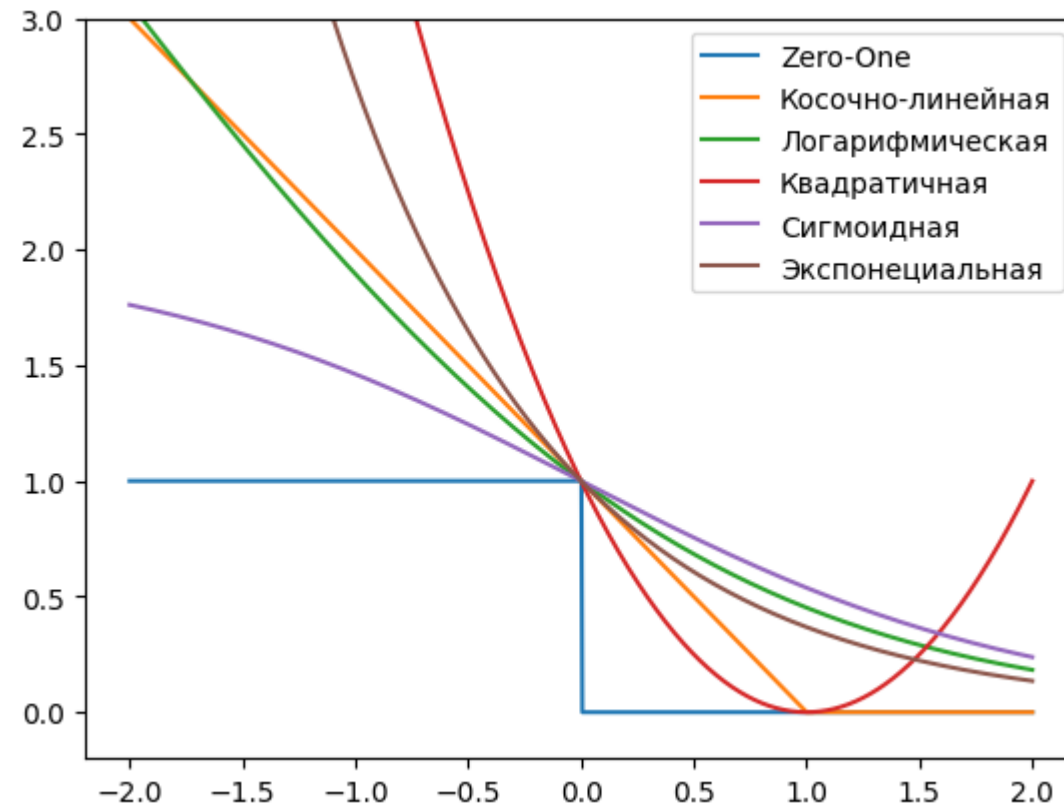
Функция потерь:

$$L(M) = [y_i \langle \theta, x_i \rangle < 0]$$

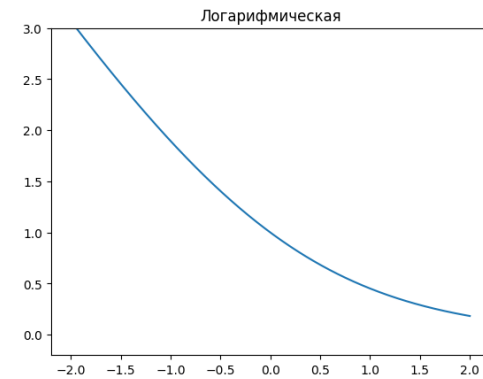
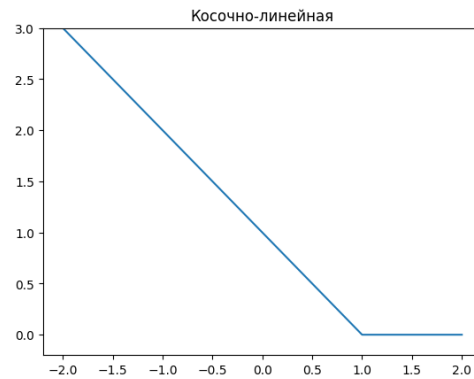
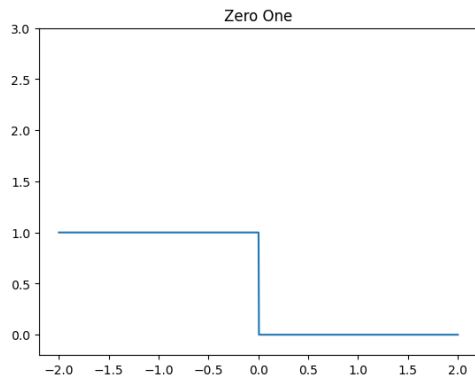
$$L(y, a(x)) = \begin{cases} 0 & \text{if } y = a(x) \\ 1 & \text{if } y \neq a(x) \end{cases}$$



Обучение линейных классификаторов

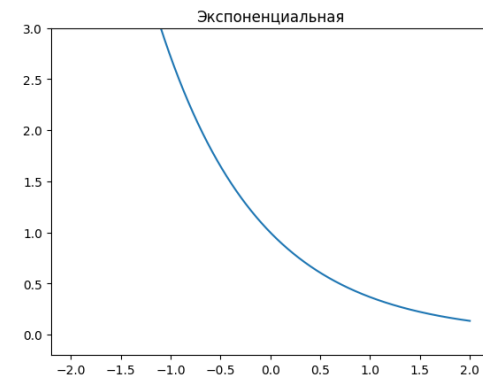
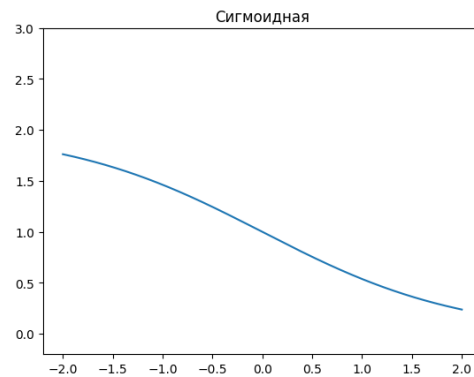
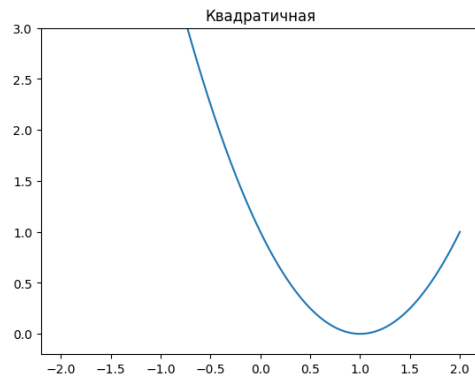


Обучение линейных классификаторов



$$V(M) = \max(1 - M)$$

$$L(M) = \log_2(1 + e^{-M})$$



$$Q(M) = (1 - M)^2$$

$$S(M) = \frac{1}{(1 + e^M)}$$

$$E(M) = e^{-M}$$

Обучение линейных классификаторов

Квадратичная функция потерь

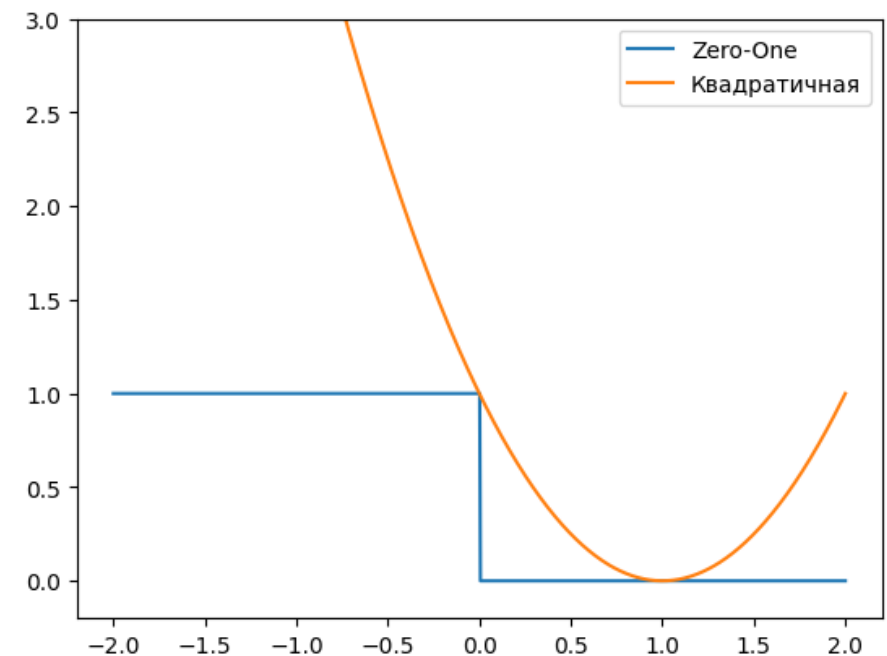
$$L_i(\theta) = (1 - M_i)^2 = (1 - \theta^T \cdot x_i \cdot y_i)^2$$

$$Q(\theta) = \sum_{i=1}^{\ell} (1 - \theta^T \cdot x_i \cdot y_i)^2 \rightarrow \min_{\theta}$$

$$\frac{\partial Q(\theta)}{\partial \theta} = -2 \cdot \sum_{i=1}^{\ell} (1 - \theta^T \cdot x_i \cdot y_i) \cdot x_i^T \cdot y_i = 0$$

$$\sum_{i=1}^{\ell} x_i^T \cdot y_i - \theta^T \cdot \sum_{i=1}^{\ell} x_i \cdot x_i^T \cdot y_i^2 = 0$$

$$\theta^T = \sum_{i=1}^{\ell} x_i^T \cdot y_i \cdot \left(\sum_{i=1}^{\ell} x_i \cdot x_i^T \right)^{-1}$$



Обучение линейных классификаторов

Логистическая функция потерь

$$Q(\theta, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} \log(1 + e^{y_i \langle \theta, x_i \rangle}) \rightarrow \min_{\theta}$$

Вычисляем градиент:

$$\nabla Q(\theta, X) = -\frac{1}{\ell} \sum_{i=1}^{\ell} \frac{y_i x_i}{1 + e^{y_i \langle \theta, x_i \rangle}}$$

Градиентный спуск

Выбираем логистическую функцию потерь:

$$Q(\theta, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} \log(1 + e^{-y_i \langle \theta, x_i \rangle}) \rightarrow \min_{\theta}$$

Вычисляем градиент:

$$\nabla_{\theta} Q(\theta, X) = -\frac{1}{\ell} \sum_{i=1}^{\ell} \frac{x_i y_i}{1 + e^{(y_i \langle \theta, x_i \rangle)}}$$

Делаем градиентный спуск:

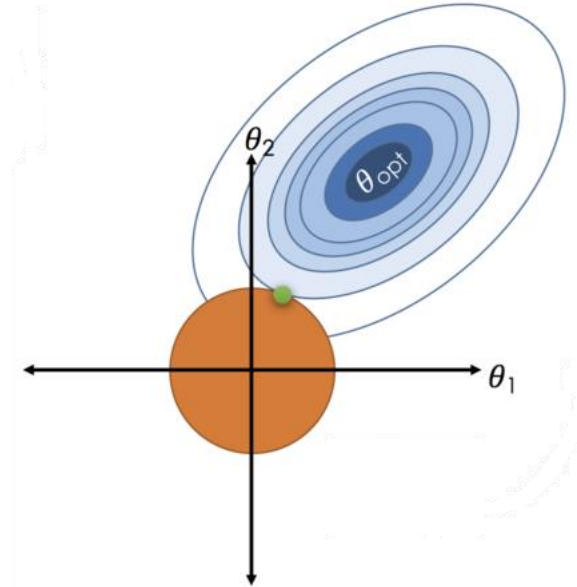
$$\theta^{(t)} = \theta^{(t-1)} + \alpha \frac{1}{\ell} \sum_{i=1}^{\ell} \frac{x_i y_i}{1 + e^{(y_i \langle \theta, x_i \rangle)}}$$

Обучение линейных классификаторов

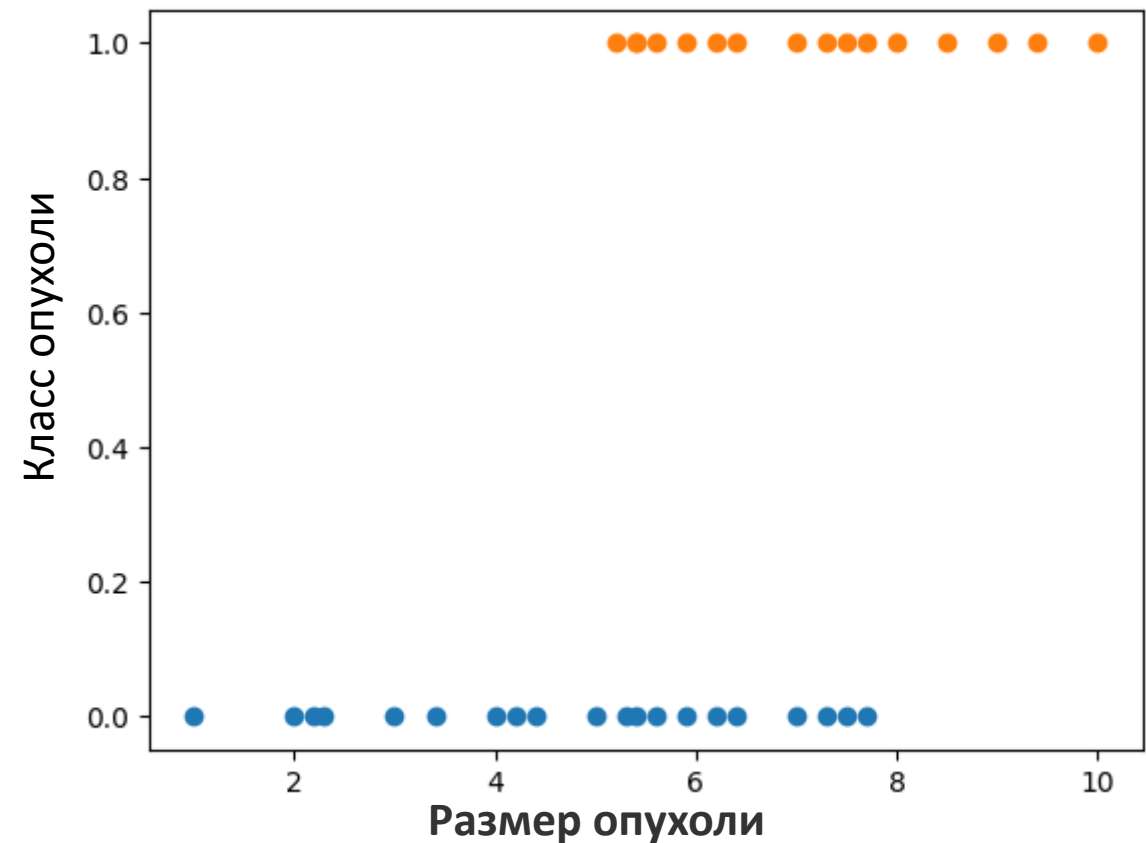
⚡ Пример регуляризации

$$Q(\theta, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} \log(1 + e^{-y_i \langle \theta, x_i \rangle}) \rightarrow \min_{\theta}$$

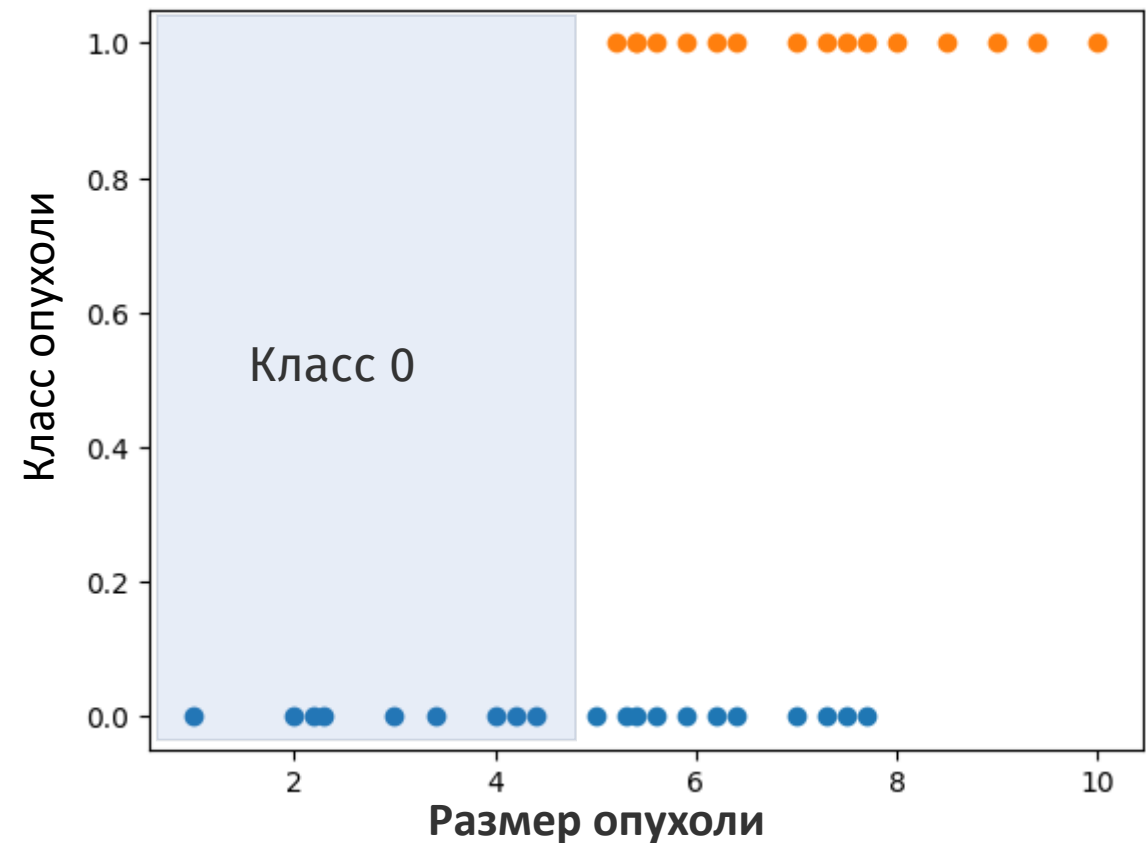
$$\frac{1}{\ell} \sum_{i=1}^{\ell} \log(1 + e^{-y_i \langle \theta, x_i \rangle}) + \lambda \|\theta\|_2 \rightarrow \min_{\theta}$$



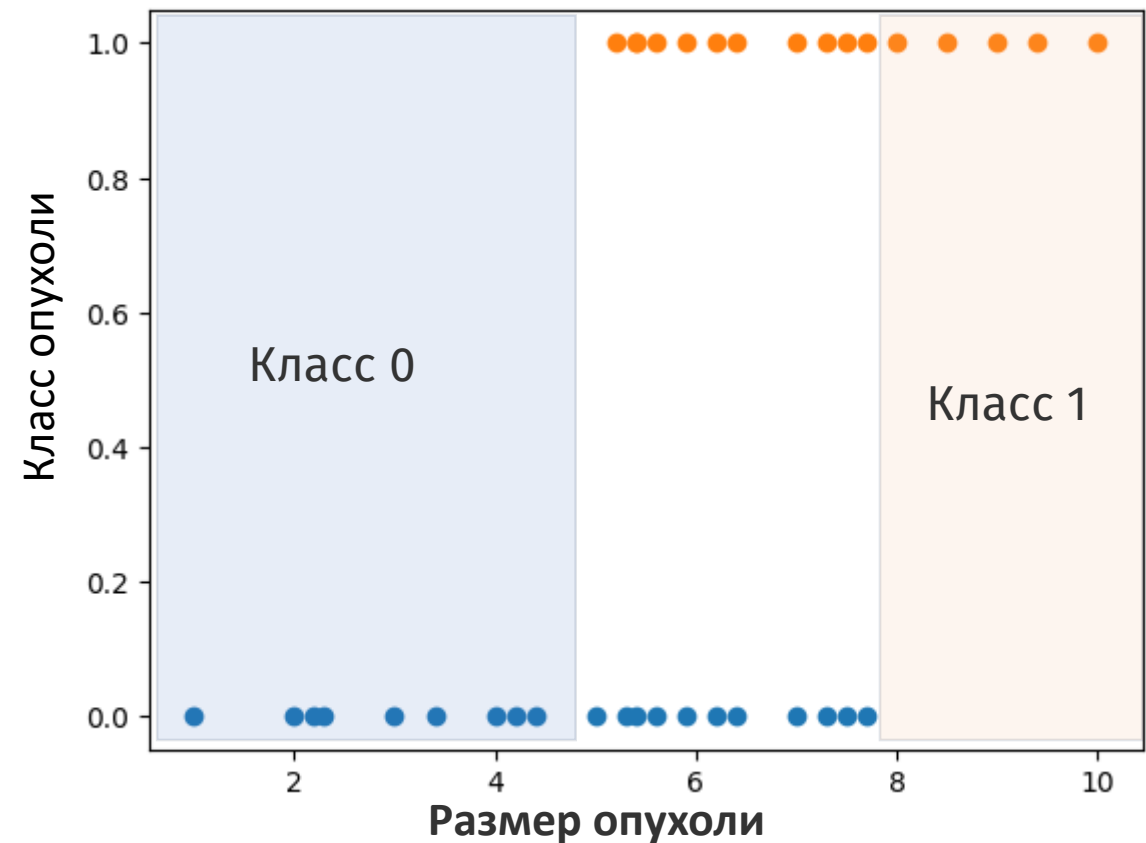
Классификация (Classification)



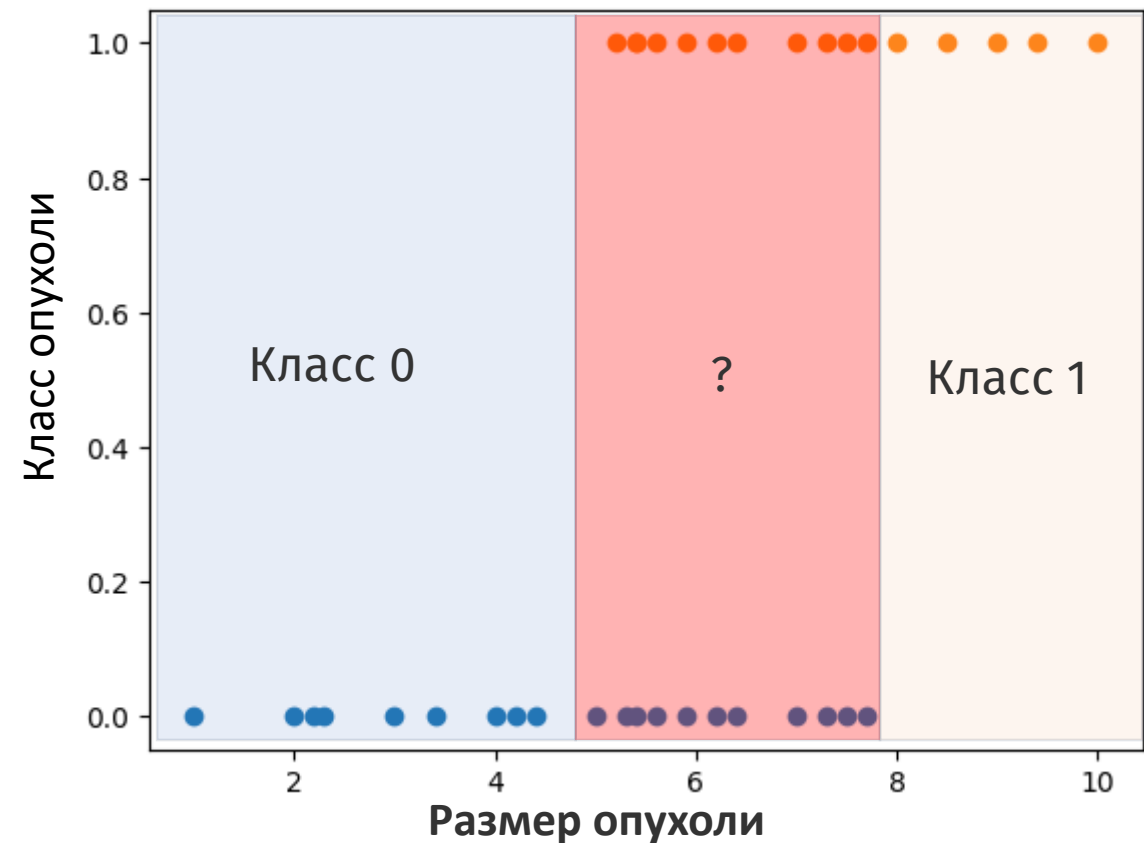
Классификация (Classification)



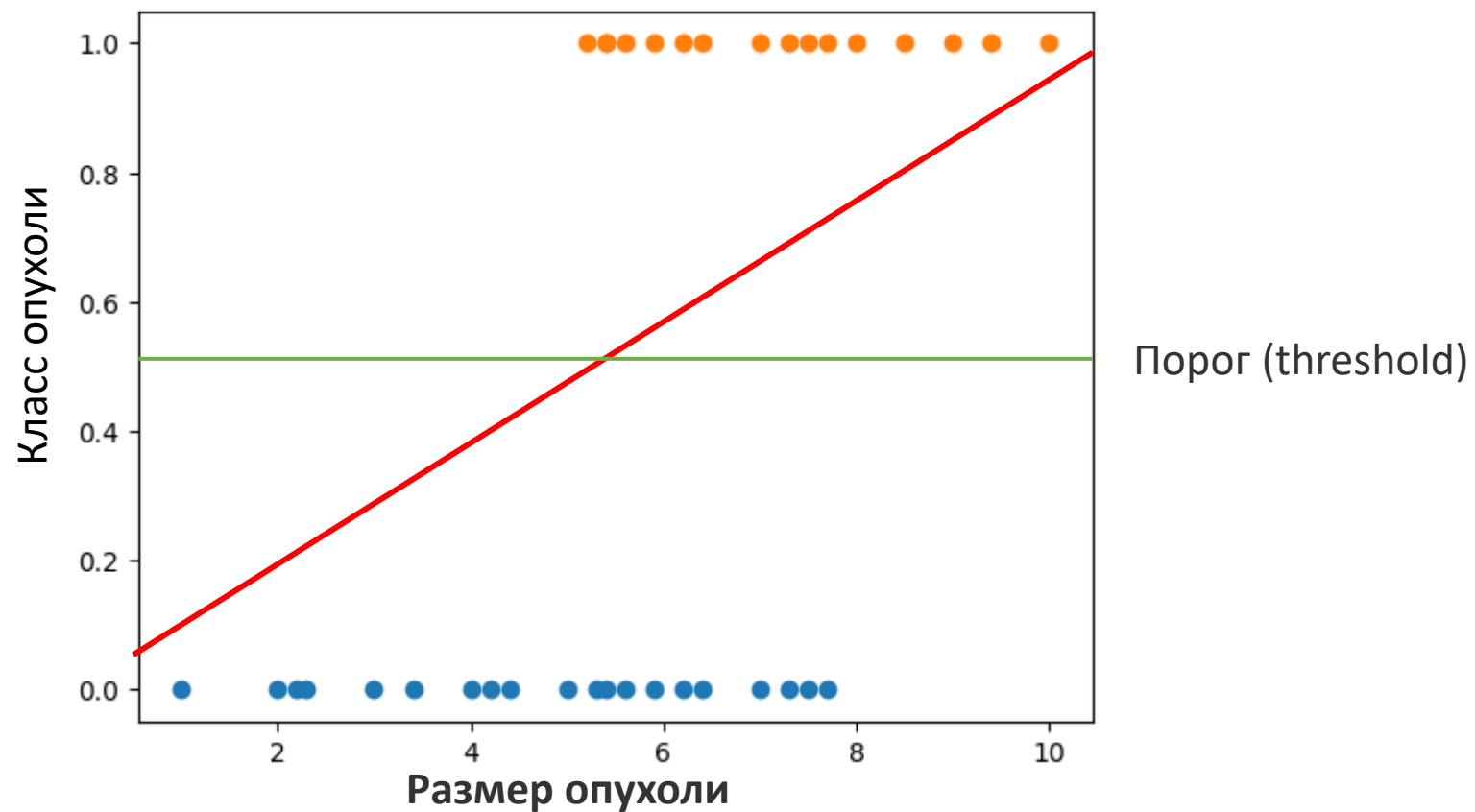
Классификация (Classification)



Классификация (Classification)



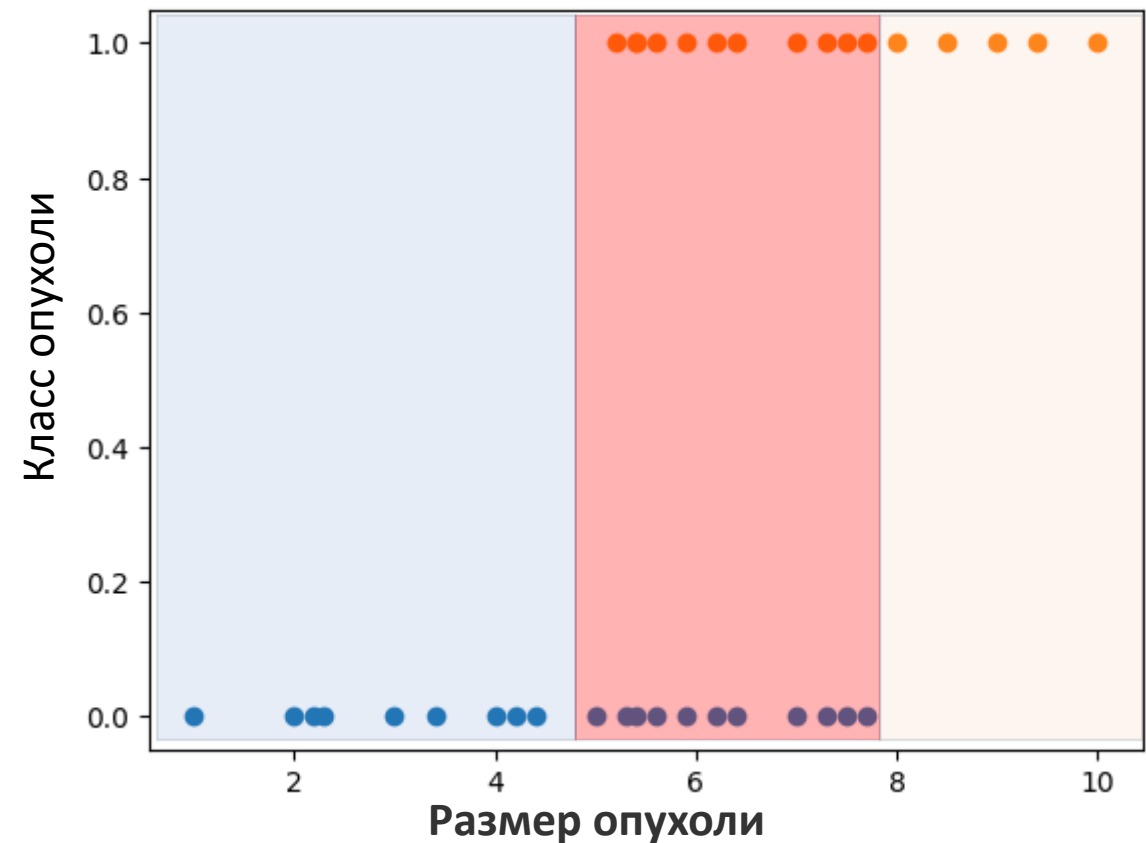
Классификация (Classification)



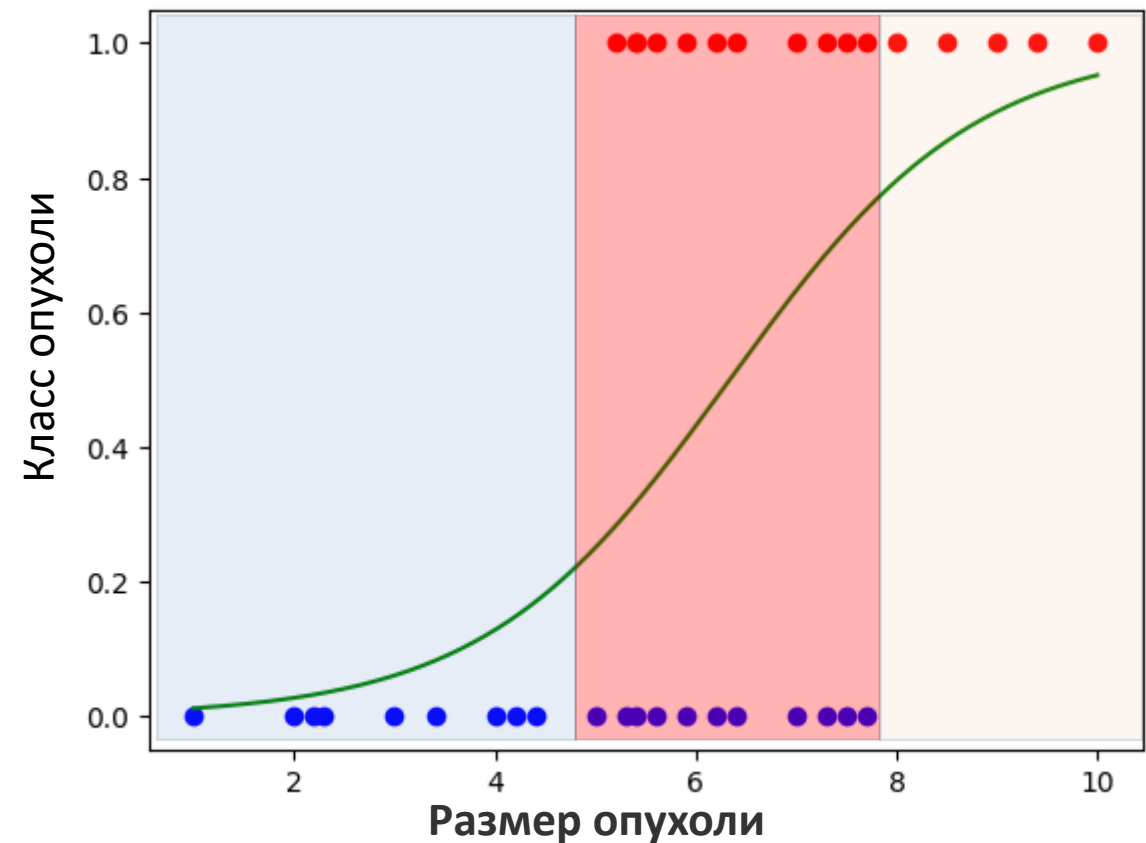
Если $a(x) = \text{sign}(\langle \theta, x \rangle - t) \geq 0.5 \rightarrow y = 1$

Если $a(x) = \text{sign}(\langle \theta, x \rangle - t) \leq 0.5 \rightarrow y = 0$

Классификация (Classification)



Классификация (Classification)



Классификация (Classification)

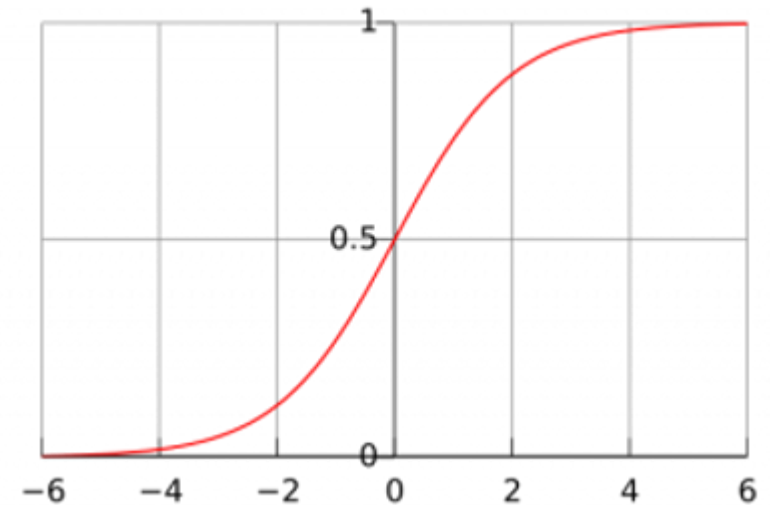
$$a(x) = \textit{sign}(\langle \theta, x \rangle)$$

Вероятность - это число на отрезке $[0;1]$

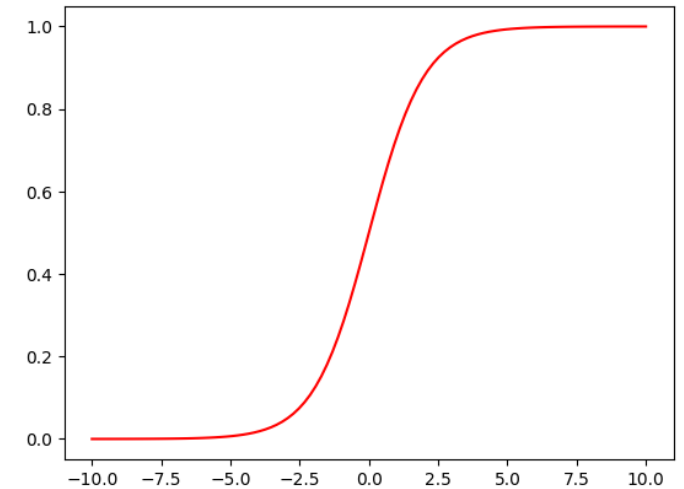
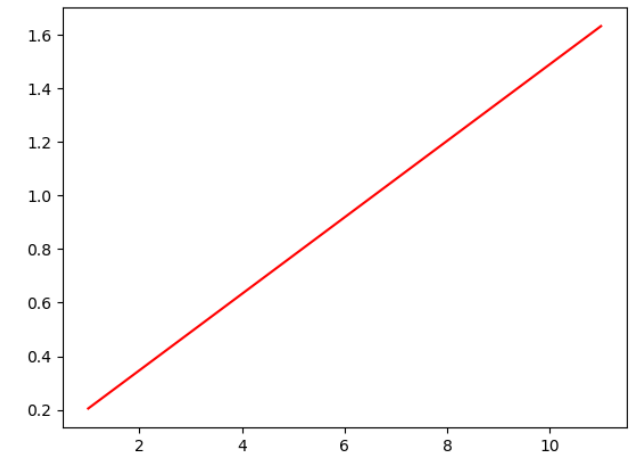
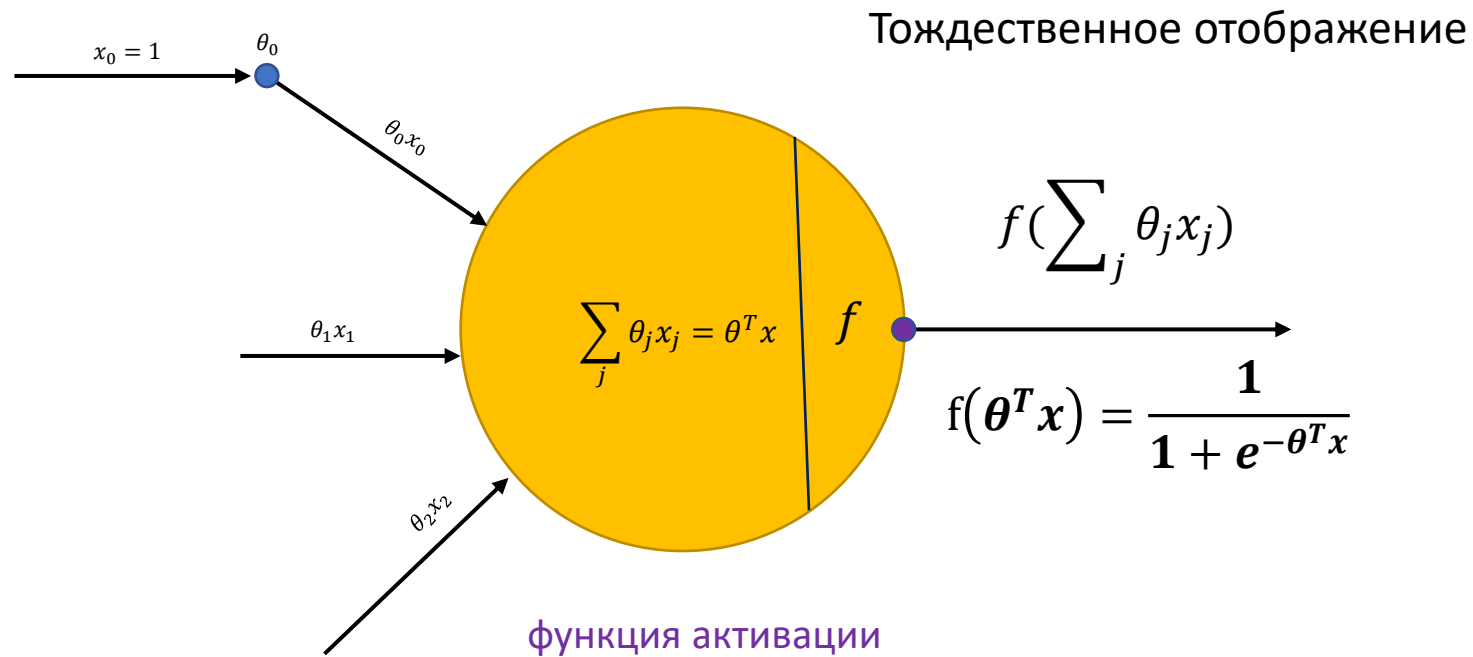
В логистической регрессии для предсказания вероятности используется **сигмоида**

$$a(x) = \sigma(\langle \theta, x \rangle)$$

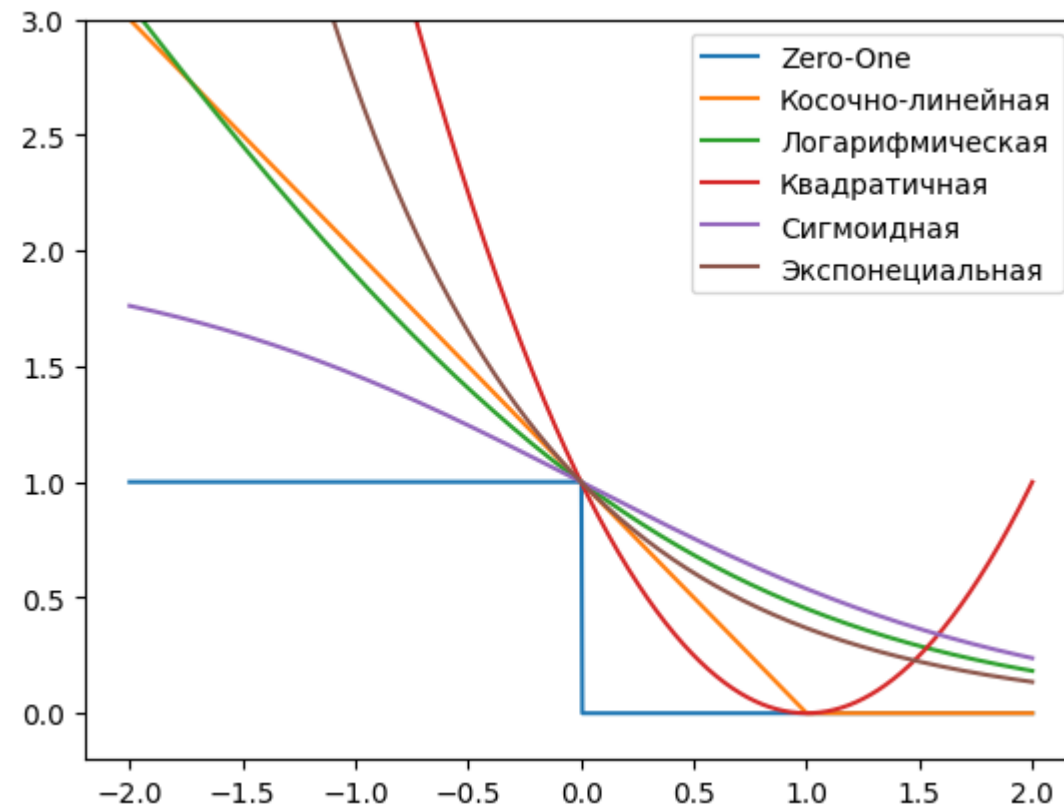
где $\sigma = \frac{1}{1 + e^{-\langle \theta, x \rangle}}$



Логистическая регрессия



Обучение линейных классификаторов



Square loss

Модель

$$a(x) = \sigma(\theta_0 x_0 + \theta_1 x_1 + \dots + \theta_d x_d) = \sigma(\langle \theta, x \rangle)$$

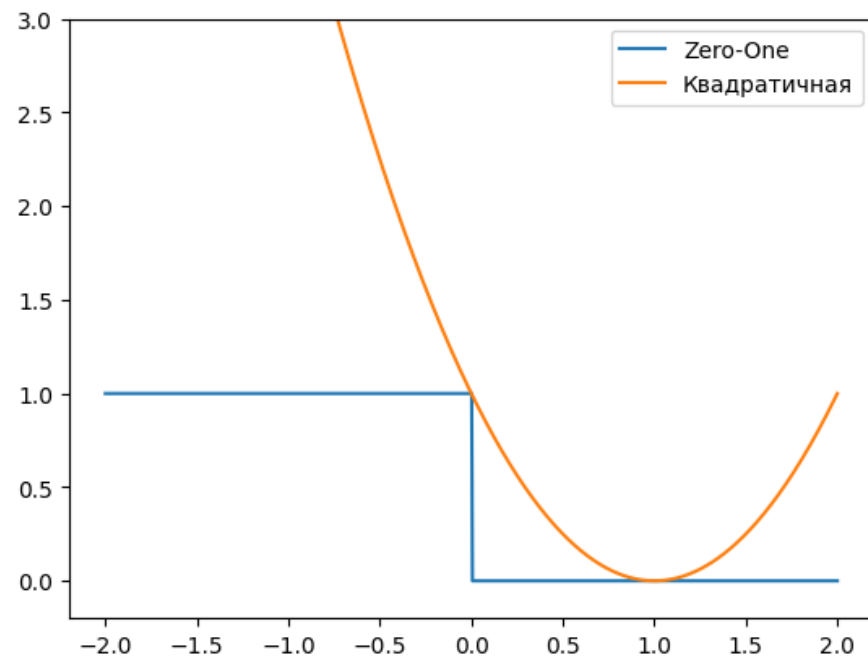
Функция потерь

$L(a(x), y) = (a(x) - y(x))^2$ – квадратичная ошибка

$$L_{MSE} = (y_i - x_i^T \theta)^2 = \frac{(y_i^2 - y_i \cdot x_i^T \theta)^2}{y_i^2} = (1 - y_i \cdot x_i^T \theta)^2$$

$$Q(M) = (1 - M)^2$$

Square loss

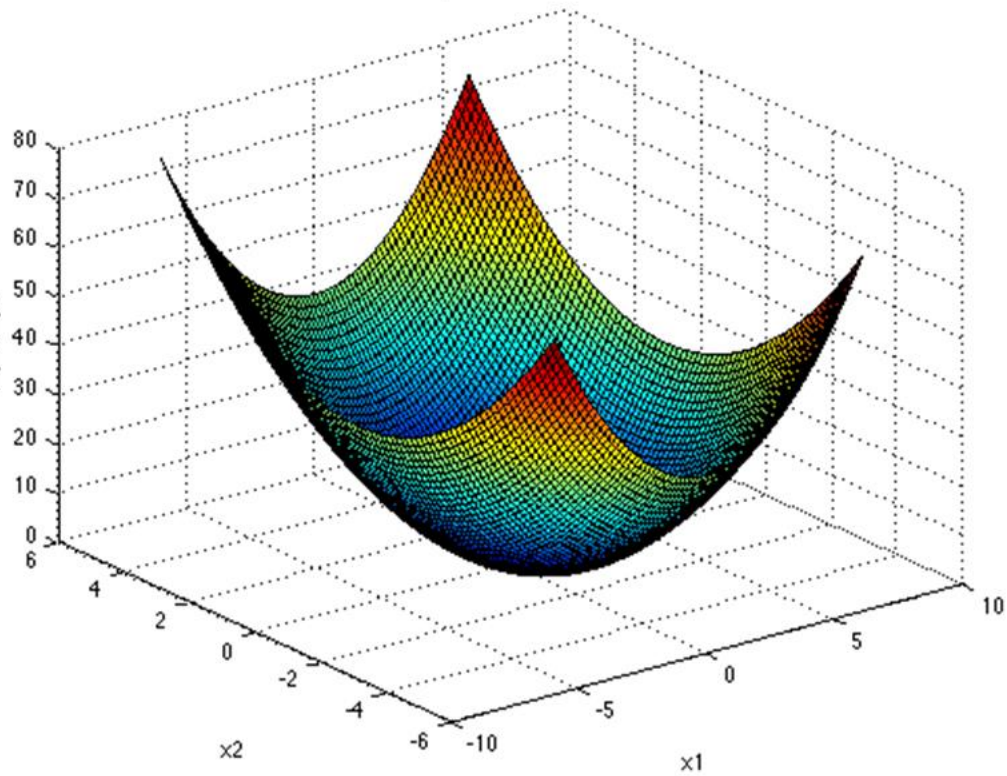


$$Q(M) = (1 - M)^2$$

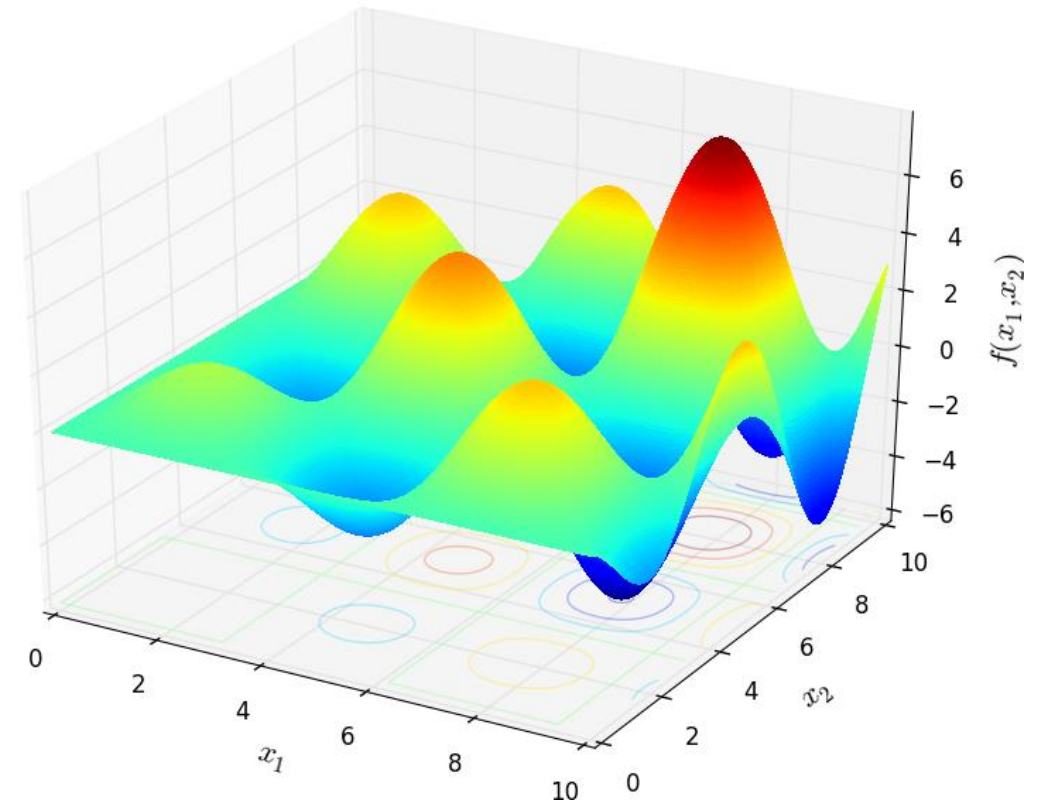
Логистическая регрессия

Функция потерь:

Convex



Non-Convex



Логистическая регрессия

Задача кредитного скоринга

Доход (x_1)	Наличие жилья(x_2)	Наличие работы(x_3)	Число детей(x_4)	(Y)
100 000	1	1	2	1
100 000	1	1	2	-1
100 000	1	1	2	-1
100 000	1	1	2	1
100 000	1	1	2	1
150 000	1	0	3	-1
150 000	1	0	3	-1
150 000	1	0	3	-1
150 000	1	0	3	-1
150 000	1	0	3	-1
150 000	1	0	3	1

$$P(y|x, \theta)$$

$$x = [100\ 000, 1, 1, 2]$$

$$P(y = +1 | x, \theta) = \frac{3}{5} = 0.6$$

$$P(y = -1 | x, \theta) = 1 - P(y = +1 | x, \theta)$$

$$= 1 - \frac{3}{5} = 0.4$$

Бинарная классификация

$$a(x) = \sigma(\langle \theta, x \rangle)$$

Метод максимального правдоподобия

$$P(y|x, \theta)$$

$$a(x) = \underset{\theta}{\operatorname{argmax}} P(y|x, \theta)$$

$$L(\theta) = \prod_{i=1}^n P(y_i|x_i, \theta)$$

Метод максимального правдоподобия

$$a(x) = \underset{\theta}{\operatorname{argmax}} P(y|x, \theta)$$

В общем случае для всей обучающей выборки

$$P(y_1, y_2, \dots, y_n | x_1, x_2, \dots, x_n, \theta_1, \theta_2, \dots, \theta_d)$$

Наивный байесовский классификатор (Naïve Bayes classifier)

$$L(\theta) = \prod_{i=1}^n P(y_i | x_i, \theta)$$

Распределение Бернулли

$$x_i = \begin{cases} 1, \text{ если любит кофе} \\ 0, \text{ если не любит кофе} \end{cases}$$

x_i	0	1
$\mathbb{P}(x_i = k)$	$1 - \rho$	ρ

$$x_1 = 1, x_2 = 0, x_3 = 1, \dots, x_n = 0$$

Задача: найти ML-оценку для ρ

$$\begin{aligned} L(\rho | x_1, \dots, x_n) &= \mathbb{P}(x_1, \dots, x_n | \rho) \\ &= \mathbb{P}(x_1 | \rho) \cdot \mathbb{P}(x_2 | \rho) \cdot \mathbb{P}(x_3 | \rho) \cdot \dots \cdot \mathbb{P}(x_n | \rho) \\ &= \rho \cdot (1 - \rho) \cdot \rho \cdot \dots \cdot (1 - \rho) = \\ &\rho^{\sum x_i} \cdot (1 - \rho)^{n - \sum x_i} \rightarrow \max_{\rho} \end{aligned}$$

Прологарифмируем:

$$\ln L = \sum x_i \cdot \ln \rho + \left(n - \sum x_i \right) \cdot \ln(1 - \rho) \rightarrow \max_{\rho}$$

Логистическая регрессия

Интерпретация вывода гипотезы

$$p(y = 1|x; \theta) = \sigma(\langle \theta, x \rangle)$$

$$p(y = 0|x; \theta) = 1 - \sigma(\langle \theta, x \rangle)$$

$$\left. \begin{array}{l} p(y = 1|x; \theta) = \sigma(\langle \theta, x \rangle) \\ p(y = 0|x; \theta) = 1 - \sigma(\langle \theta, x \rangle) \end{array} \right\} p(y|x; \theta) = \sigma(\langle \theta, x \rangle)^y (1 - \sigma(\langle \theta, x \rangle))^{1-y}$$

Функция правдоподобия

$$L(\theta) = p(Y|X; \theta) = \prod_{i=0}^m p(y^{(i)} | x^{(i)}; \theta) = \prod_{i=1}^m \sigma(\theta^T x^{(i)})^{y^{(i)}} (1 - \sigma(\theta^T x^{(i)}))^{1-y^{(i)}}$$

Логистическая регрессия

Логарифм функции правдоподобия

$$L(\theta) = \log L(\theta) = \log \prod_{i=1}^n \sigma(\theta^T \mathbf{x}^{(i)})^{y^{(i)}} (1 - \sigma(\theta^T \mathbf{x}^{(i)}))^{1-y^{(i)}}$$

$$\log_a bc = \log_a b + \log_a c$$

Логарифм функции правдоподобия

$$\begin{aligned} L(\theta) &= \log L(\theta) = \log \prod_{i=1}^n \sigma(\theta^T \mathbf{x}^{(i)})^{y^{(i)}} (1 - \sigma(\theta^T \mathbf{x}^{(i)}))^{1-y^{(i)}} \\ &= \sum_{i=1}^n \log(\sigma(\theta^T \mathbf{x}^{(i)})^{y^{(i)}} + \log(1 - \sigma(\theta^T \mathbf{x}^{(i)}))^{1-y^{(i)}}) \end{aligned}$$

$$\log_a b^c = c \cdot \log_a b$$

Логарифм функции правдоподобия

$$L(\theta) = \log L(\theta) = \log \prod_{i=1}^n \sigma(\theta^T \mathbf{x}^{(i)})^{y^{(i)}} (1 - \sigma(\theta^T \mathbf{x}^{(i)}))^{1-y^{(i)}}$$

$$= \sum_{i=1}^n \log(\sigma(\theta^T \mathbf{x}^{(i)})^{y^{(i)}} + \log(1 - \sigma(\theta^T \mathbf{x}^{(i)}))^{1-y^{(i)}})$$

$$= \sum_{i=1}^m y^{(i)} \log \sigma(\theta^T \mathbf{x}^{(i)}) + (1 - y^{(i)}) \log(1 - \sigma(\theta^T \mathbf{x}^{(i)}))$$

Логарифм функции правдоподобия

$$\mathbf{L}(\boldsymbol{\theta}) = \sum_{i=1}^m y^{(i)} \log \sigma(\boldsymbol{\theta}^T \mathbf{x}^{(i)}) + (1 - y^{(i)}) \log(1 - \sigma(\boldsymbol{\theta}^T \mathbf{x}^{(i)}))$$

$$-\mathbf{L}(\boldsymbol{\theta}) = \sum_{i=1}^m -y^{(i)} \log \sigma(\boldsymbol{\theta}^T \mathbf{x}^{(i)}) - (1 - y^{(i)}) \log(1 - \sigma(\boldsymbol{\theta}^T \mathbf{x}^{(i)}))$$

Логарифм функции правдоподобия

$$\mathbf{L}(\boldsymbol{\theta}) = \sum_{i=1}^m -y^{(i)} \log \sigma(\boldsymbol{\theta}^T \mathbf{x}^{(i)}) - (1 - y^{(i)}) \log(1 - \sigma(\boldsymbol{\theta}^T \mathbf{x}^{(i)}))$$

Функция потерь

$$\frac{\partial L}{\partial \boldsymbol{\theta}} = \sum_{i=1}^m y^{(i)} \mathbf{x}^{(i)} (1 - \sigma(\boldsymbol{\theta}^T \mathbf{x}^{(i)})) \quad \frac{\partial L}{\partial \boldsymbol{\theta}} = \sum_{i=1}^m (y^{(i)} - 1) \mathbf{x}^{(i)} \sigma(\boldsymbol{\theta}^T \mathbf{x}^{(i)})$$

$$\frac{\partial L}{\partial \boldsymbol{\theta}} = \sum_{i=1}^m y^{(i)} \mathbf{x}^{(i)} (1 - \sigma(\boldsymbol{\theta}^T \mathbf{x}^{(i)})) + \sum_{i=1}^m (y^{(i)} - 1) \mathbf{x}^{(i)} \sigma(\boldsymbol{\theta}^T \mathbf{x}^{(i)})$$

$$\frac{\partial L}{\partial \boldsymbol{\theta}} = \sum_{i=1}^m \mathbf{x}^{(i)} (y^{(i)} - \sigma(\boldsymbol{\theta}^T \mathbf{x}^{(i)}))$$

Логарифм функции правдоподобия

$$\mathbf{L}(\boldsymbol{\theta}) = \sum_{i=1}^m -y^{(i)} \log \sigma(\boldsymbol{\theta}^T \mathbf{x}^{(i)}) - (1 - y^{(i)}) \log(1 - \sigma(\boldsymbol{\theta}^T \mathbf{x}^{(i)}))$$

Градиент

$$\nabla Q(\boldsymbol{\theta}) = \frac{1}{\ell} \sum_{i=1}^{\ell} \mathbf{x}^{(i)} (y^{(i)} - \sigma(\boldsymbol{\theta}^T \mathbf{x}^{(i)}))$$

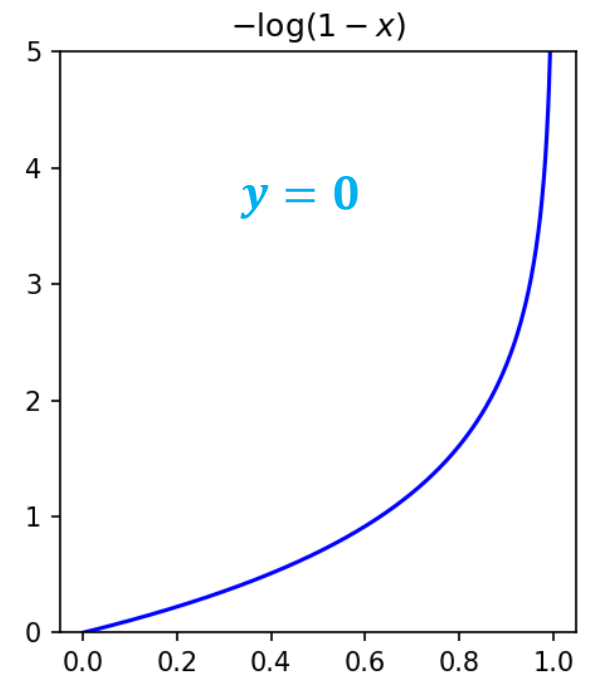
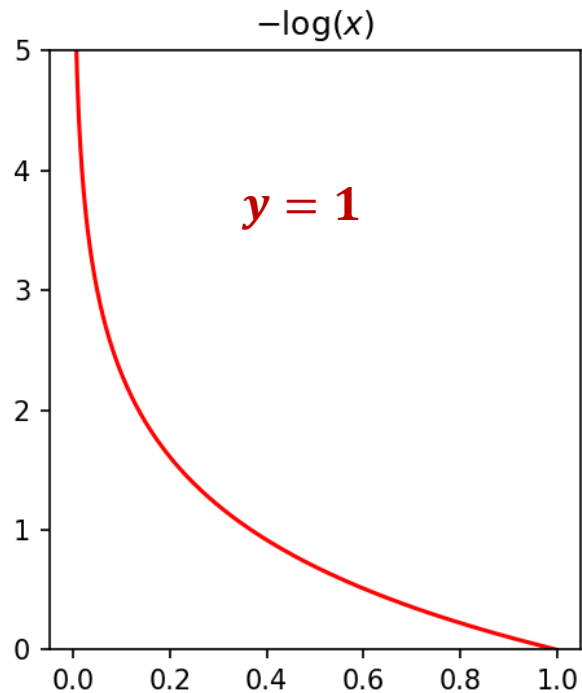
Сдвигаемся по антиградиенту

$$\boldsymbol{\theta}^t = \boldsymbol{\theta}^{t-1} - \alpha \nabla Q(\boldsymbol{\theta}^{t-1})$$

Логистическая регрессия

Функция потерь:

$$Cost(a(x), y) = \begin{cases} -\log(a(x)), & y = 1 \\ -\log(1 - a(x)), & y = 0 \end{cases}$$

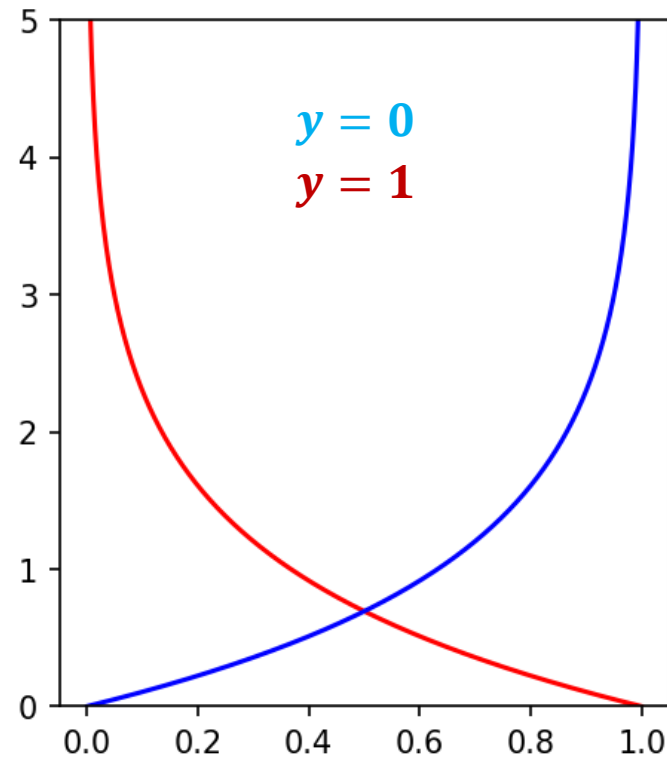


Логистическая регрессия

Функция потерь:

$$Q(\theta, X) = \sum_{i=1}^n Q(a(x^{(i)}), y^{(i)})$$

$$Q(a(x^{(i)}), y^{(i)}) = -y^{(i)}\log(a(x^{(i)})) + (1 - y^{(i)})\log(1 - a(x^{(i)}))$$



Метрики качества классификации

Accuracy

Accuracy - это доля правильных ответов модели.

$$accuracy = \frac{1}{\ell} \sum_{i=1}^{\ell} [a(x_i) = y_i]$$

$a(x_i)$ — предсказание класса на объекте x_i

y_i - правильный ответ

$[a(x_i) = y_i]$ — индикатор, то есть величина, равная 1, если $a(x_i) = y_i$, и 0 иначе

Несбалансированы

Пример : Fraud detection

Пусть в тренировочных данных 1000 транзакций, и 50 из них мошеннические.

Класс +1: 50 (мошеннические)

Класс -1: 590 (нормально)

Модель: $a(x) = -1$

$$accuracy = \frac{1}{\ell} \sum_{i=1}^{\ell} [a(x_i) = y_i]$$

Доля правильных ответов : 0.95



Сбалансированы

Пример: кредитный скоринг

в тестовых данных в задаче скоринга 1000 клиентов

Доля правильных ответов: 0.8

Можно ли сказать, что модель имеет хорошее качество?

Сбалансированы

Пример: кредитный скоринг

в тестовых данных в задаче скоринга 1000 клиентов

Доля правильных ответов: 0.8

Можно ли сказать, что модель имеет хорошее качество?

- модель выдает кредит клиентам, которые его не вернут
- модель не выдает кредит клиентам, которые его вернут

Матрица ошибок

	Y=1	Y=-1
a(x)=1 срабатывание	True Positive (TP)	False Positive (FP)
a(x)=-1 пропуск	False Negative (FN)	True Negative (TN)

Матрица ошибок

модель предсказывает, выдавать клиенту кредит или нет.

данные сбалансированы: всего 200 клиентов, по 100 клиентов каждого класса

Модель 1: $a_1(x)$

	Y=1	Y=-1
a(x)=1 Получили кредит	80	20
a(x)=-1 Не получили кредит	20	80

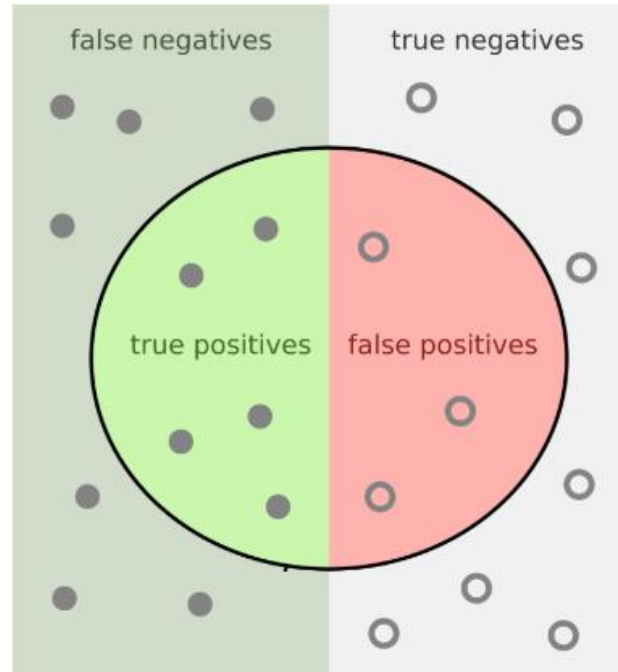
Модель 2: $a_2(x)$

	Y=1	Y=-1
a(x)=1 Получили кредит	48	2
a(x)=-1 Не получили кредит	52	98

Точность (precision)

$$\text{precision}(a, X) = \frac{TP}{TP + FP}$$

$$\text{Precision} = \frac{\text{green semi-circle}}{\text{green + red semi-circles}}$$



Можно ли доверять классификатору, когда он относит объект к положительному классу?

Матрица ошибок

Модель 1: $a_1(x)$

	Y=1	Y=-1
a(x)=1 Получили кредит	80	20
a(x)=-1 Не получили кредит	20	80

$\text{precision}(a_1, X) = 0.8$

Модель 2: $a_2(x)$

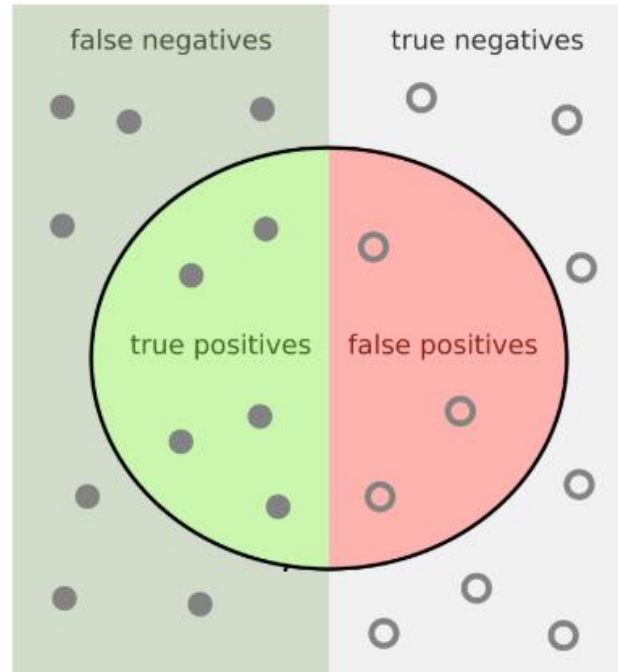
	Y=1	Y=-1
a(x)=1 Получили кредит	48	2
a(x)=-1 Не получили кредит	52	98

$\text{precision}(a_2, X) = 0.96$

Полнота (recall)

$$\text{recall}(a, X) = \frac{TP}{TP + FN}$$

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$



Какую долю положительного класса модель смогла выявить?

Матрица ошибок

Модель 1: $a_1(x)$

	Y=1	Y=-1
a(x)=1 Получили кредит	80	20
a(x)=-1 Не получили кредит	20	80

$\text{recall}(a_1, X) = 0.8$

Модель 2: $a_2(x)$

	Y=1	Y=-1
a(x)=1 Получили кредит	48	2
a(x)=-1 Не получили кредит	52	98

$\text{recall}(a_2, X) = 0.48$

Медицинская диагностика

	Y=1	Y=-1
a(x)=1	10	20
a(x)=-1	90	10000

$$accuracy(a,X)=0.99$$

$$precision(a_1,X)=0.33$$

$$recall(a_1,X)=0.1$$

доля правильных ответов

$$accuracy = \frac{1}{\ell} \sum_{i=1}^{\ell} [a(x_i) = y_i]$$

Точность

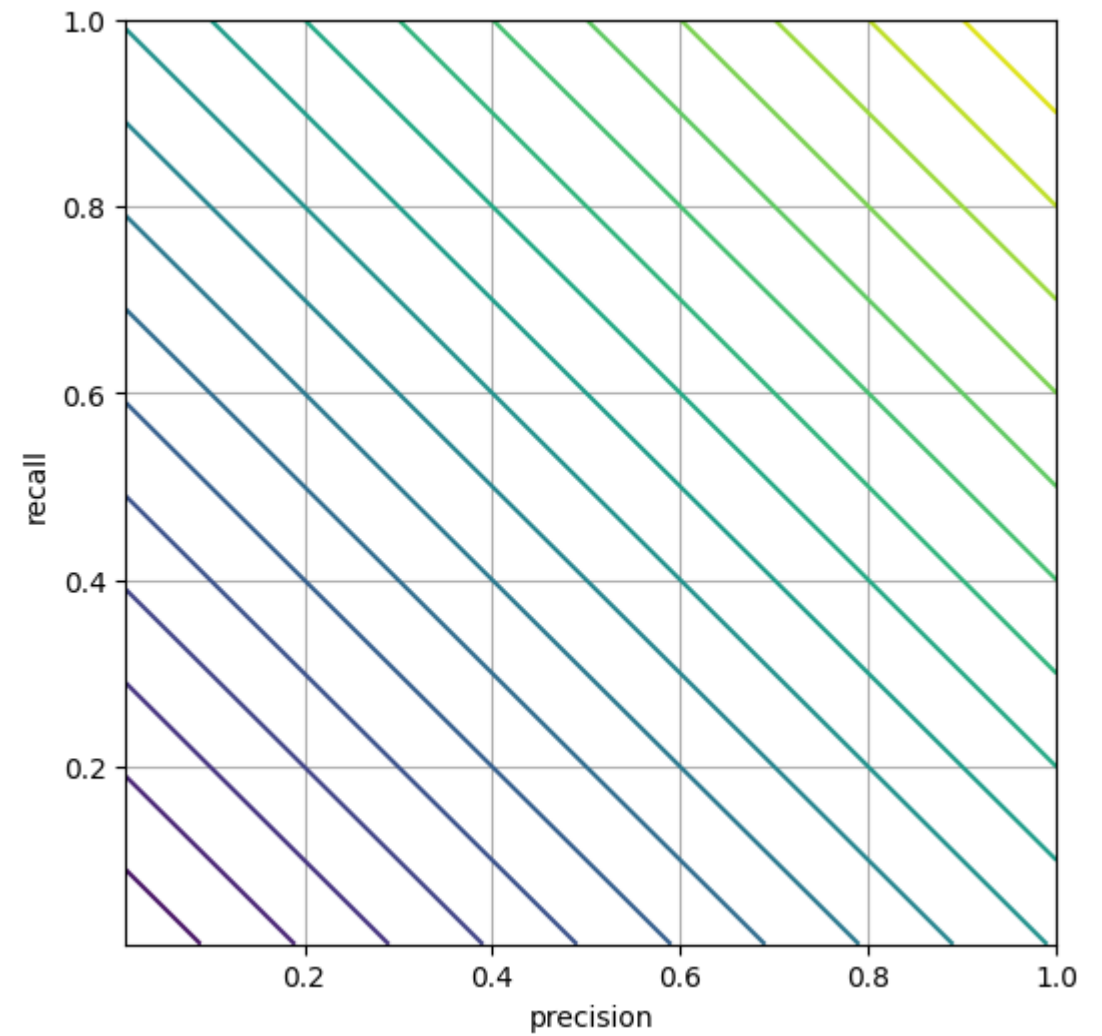
$$precision(a, X) = \frac{TP}{TP + FP}$$

Полнота

$$recall(a, X) = \frac{TP}{TP + FN}$$

Арифметическое среднее

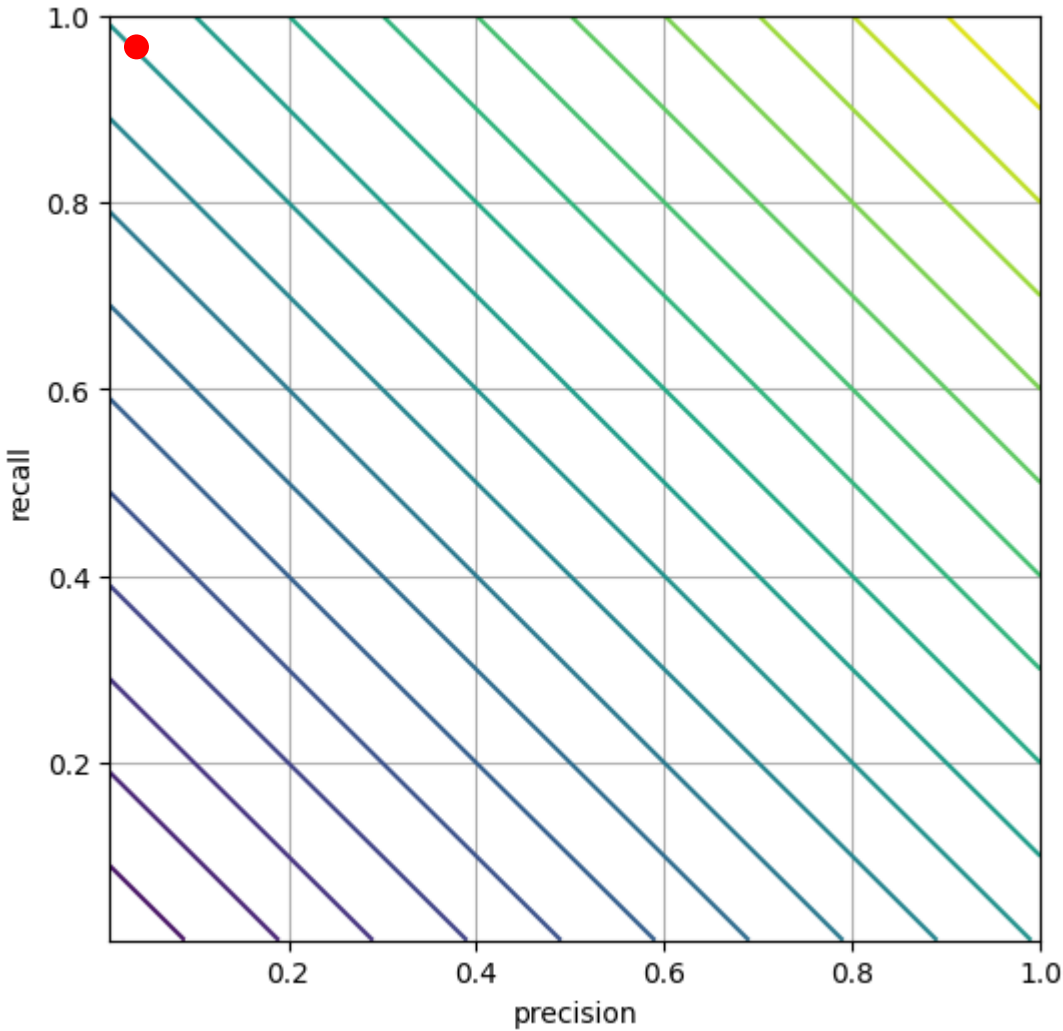
$$A = \frac{1}{2}(\textit{precision} + \textit{recall})$$



Арифметическое среднее

$$A = \frac{1}{2}(precision + recall)$$

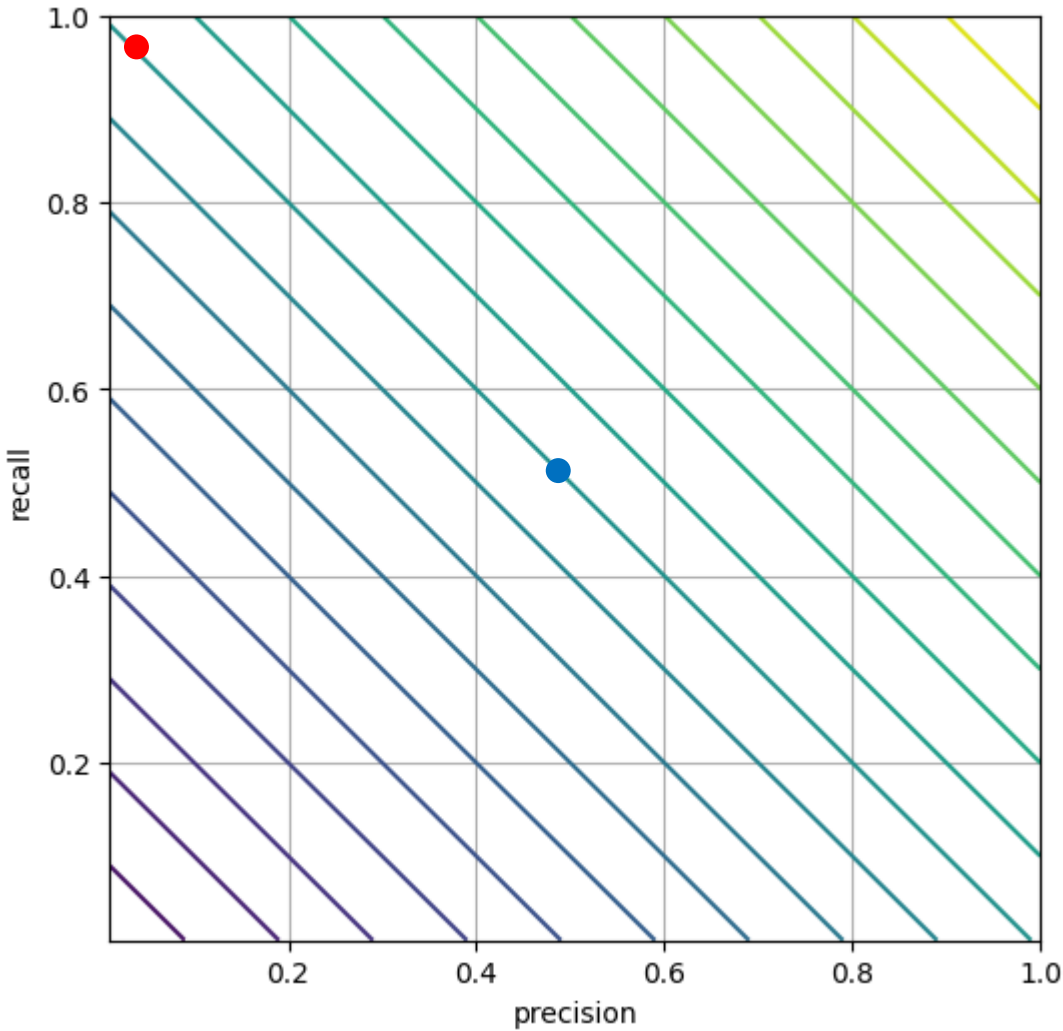
Модель	Precision	recall	A
$a_1(x)$	0.1	1	0.55



Арифметическое среднее

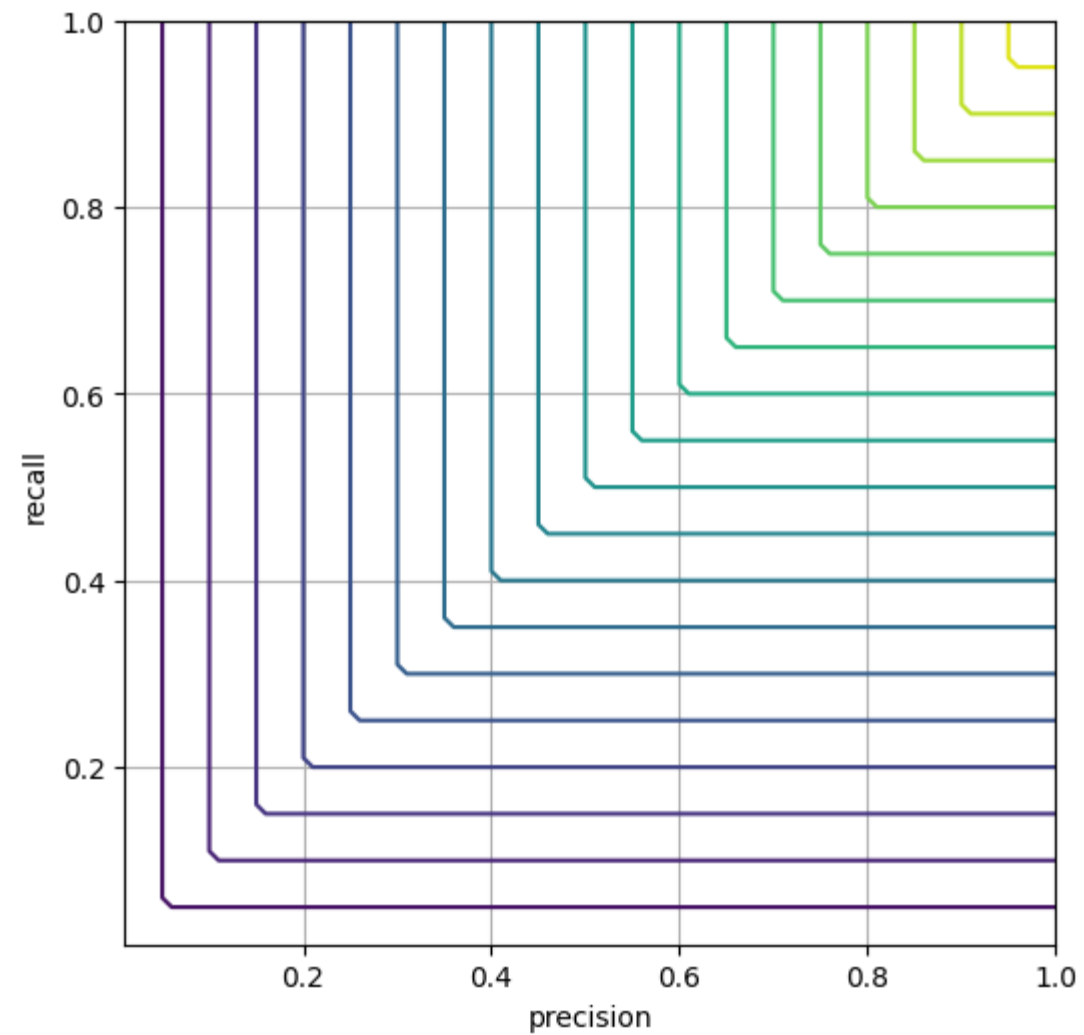
$$A = \frac{1}{2}(precision + recall)$$

Модель	Precision	recall	A
$a_1(x)$	0.1	1	0.55
$a_2(x)$	0.55	0.55	0.55



Минимальное значение

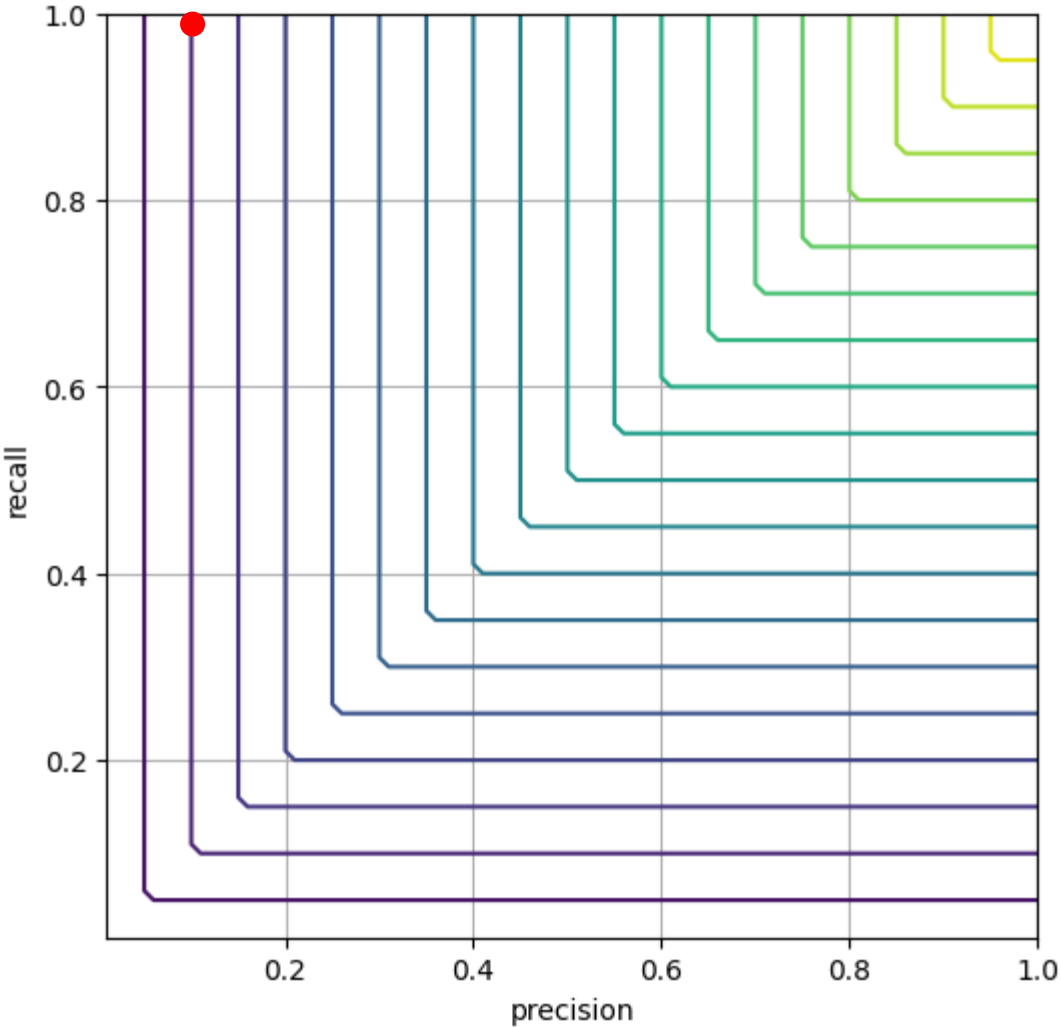
$$M = \min(\text{precision}, \text{recall})$$



Минимальное значение

$M = \min(\text{precision}, \text{recall})$

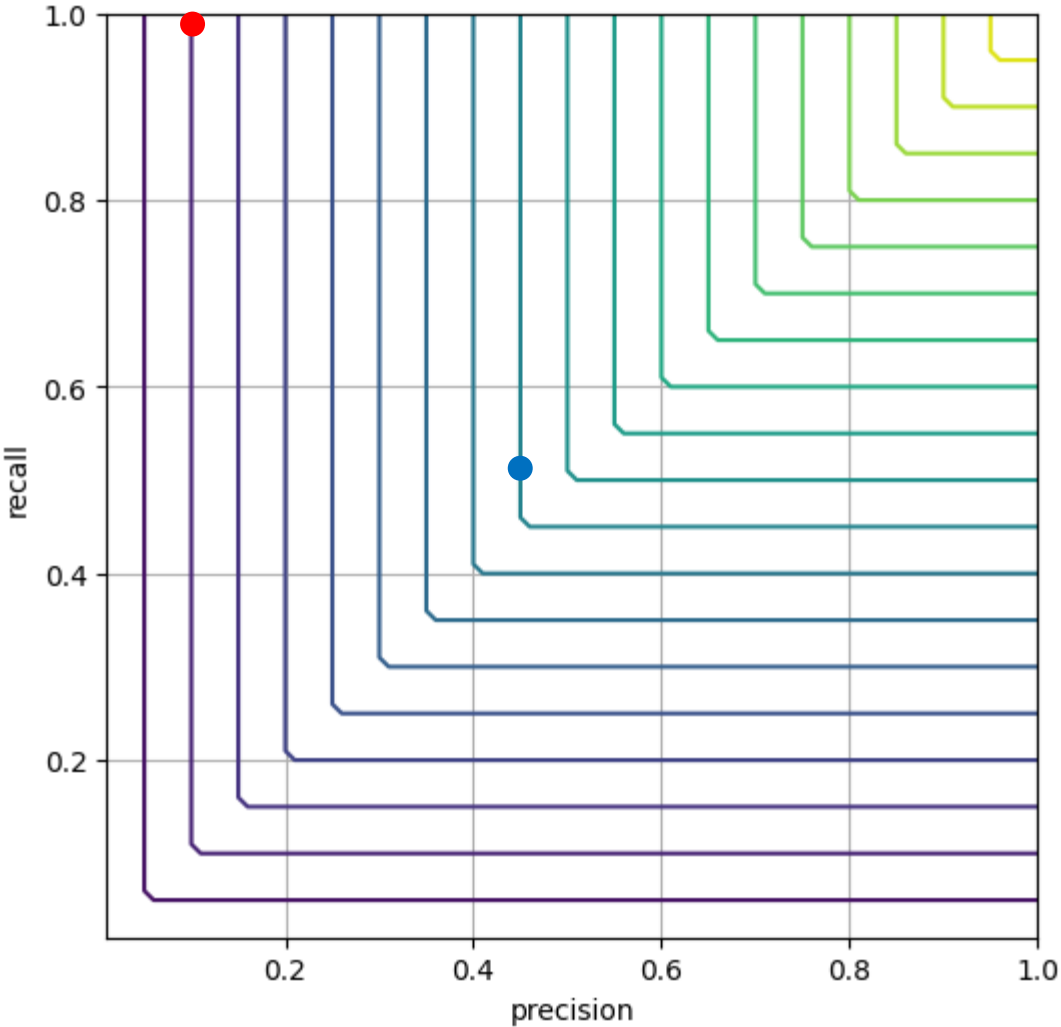
Модель	Precision	recall	M
$a_1(x)$	0.1	1	0.1



Минимальное значение

$M = \min(\text{precision}, \text{recall})$

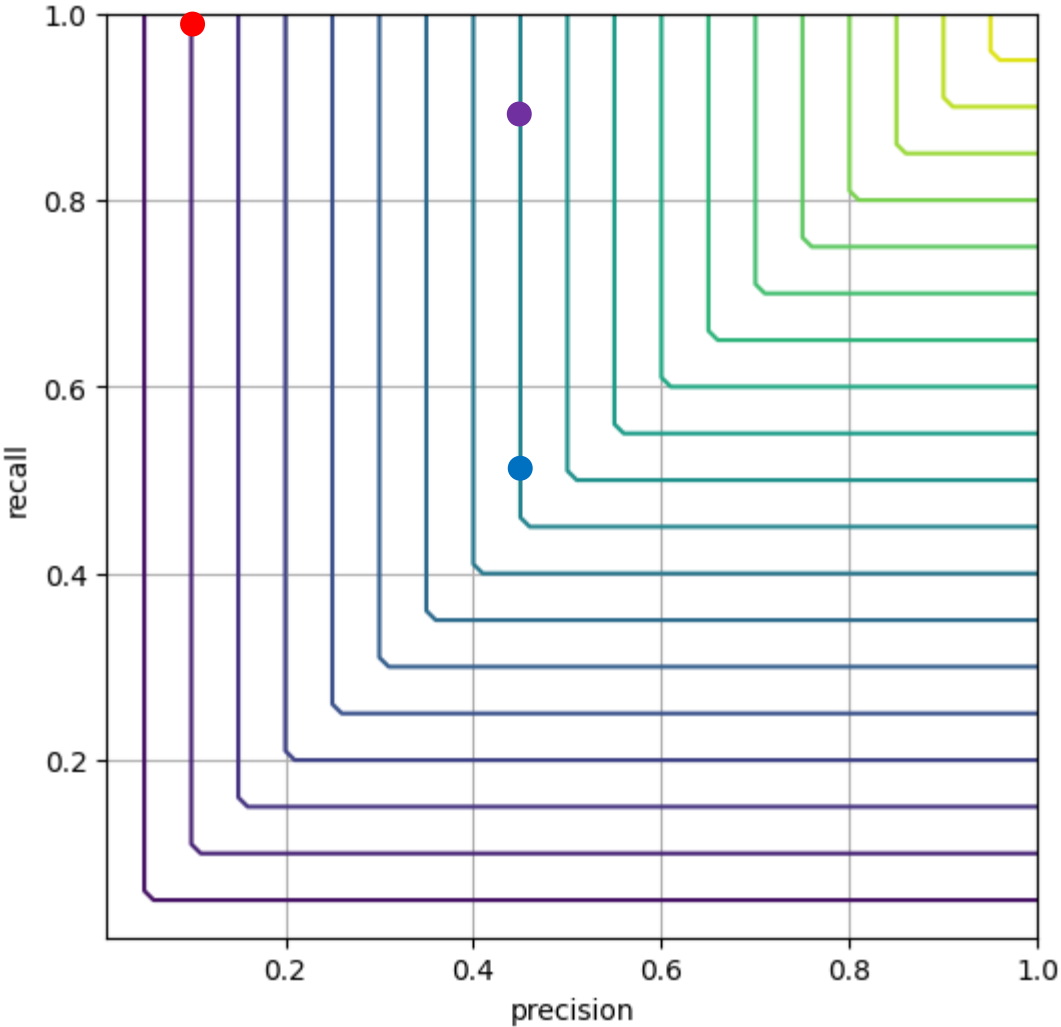
Модель	Precision	recall	M
$a_1(x)$	0.1	1	0.1
$a_2(x)$	0.55	0.55	0.55



Минимальное значение

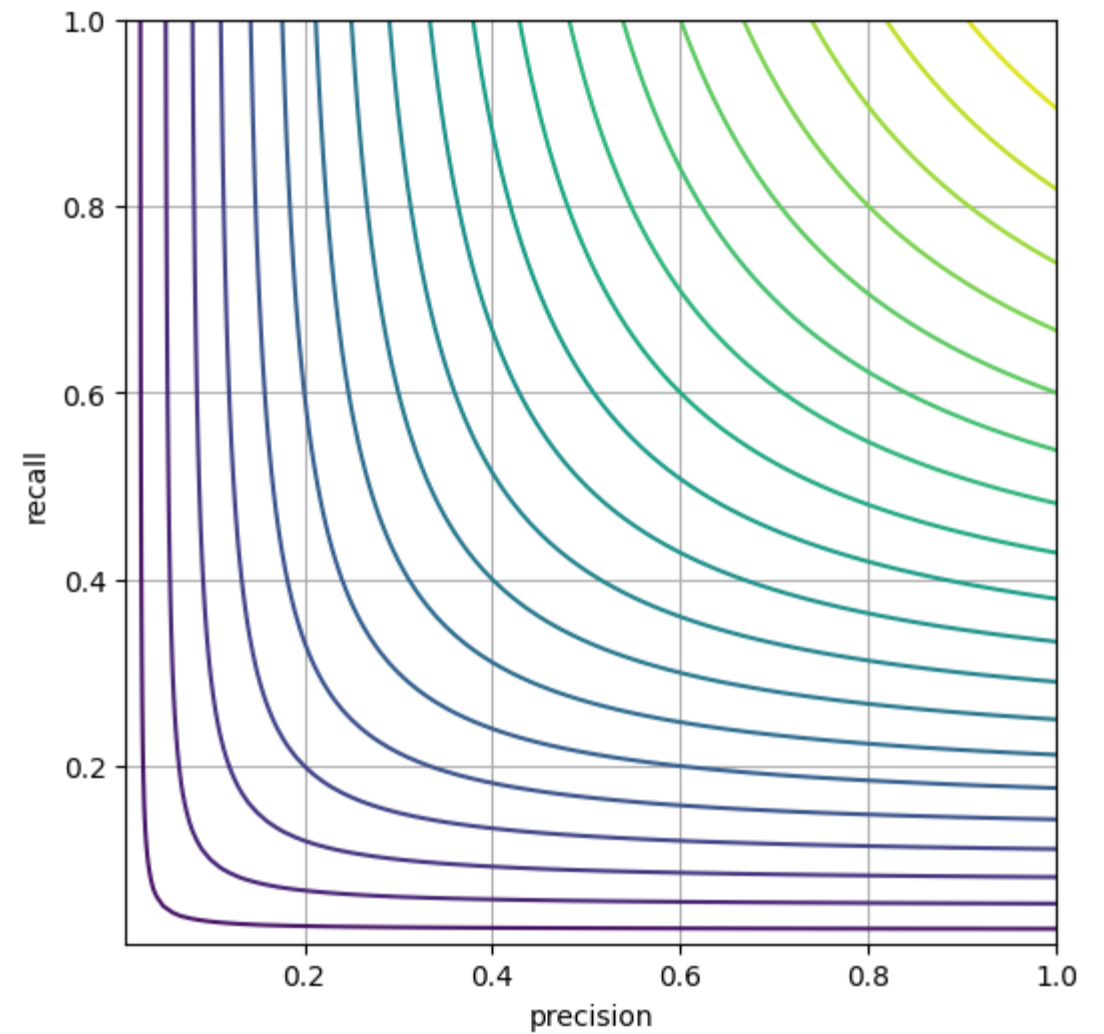
$M = \min(\text{precision}, \text{recall})$

Модель	Precision	recall	M
$a_1(x)$	0.1	1	0.1
$a_2(x)$	0.55	0.55	0.55
$a_2(x)$	0.55	0.90	0.55



Гармоническое среднее (F-мера)

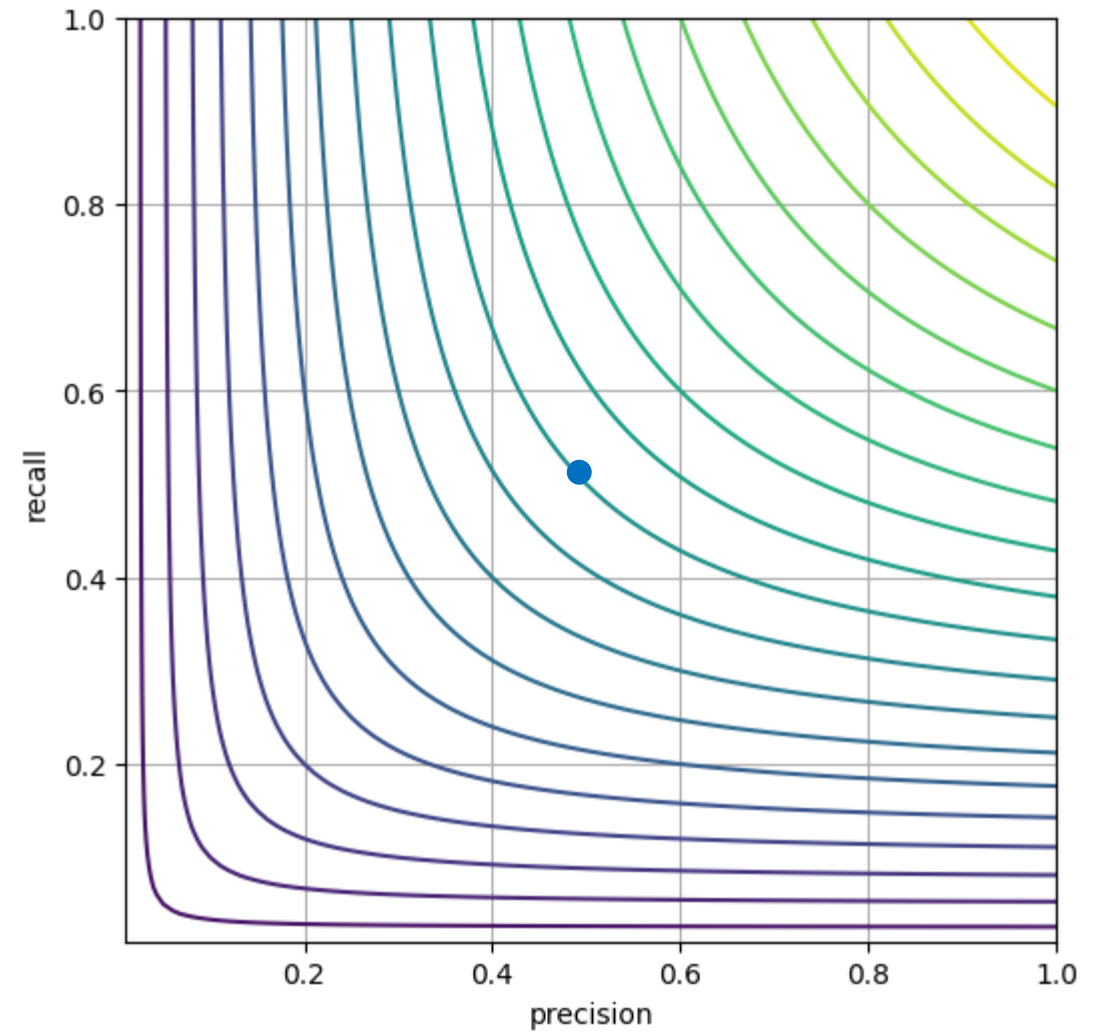
$$F = \frac{2 * precision * recall}{precision + recall}$$



Гармоническое среднее (F-мера)

$$F = \frac{2 * precision * recall}{precision + recall}$$

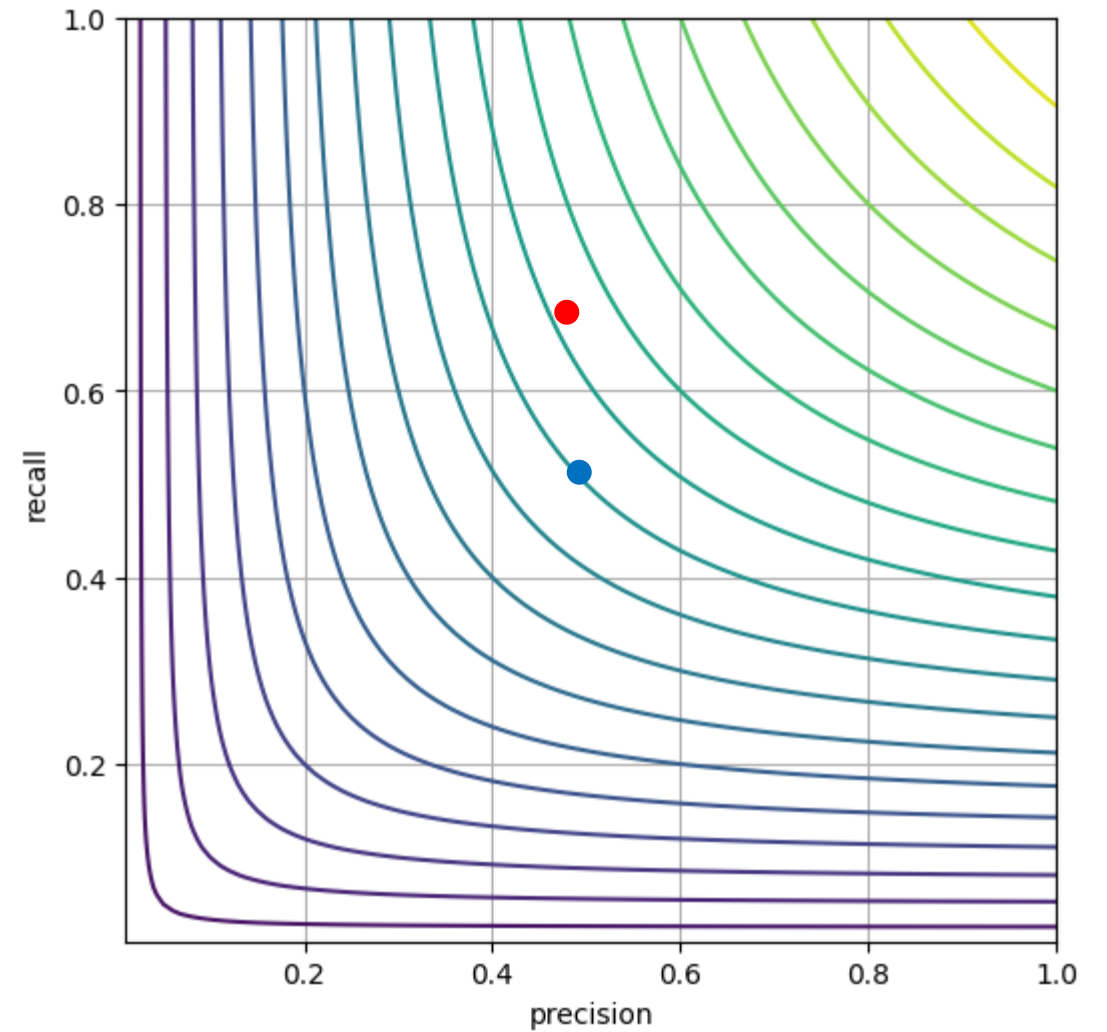
Модель	Precision	recall	F
$a_1(x)$	0.55	0.55	0.55



Гармоническое среднее (F-мера)

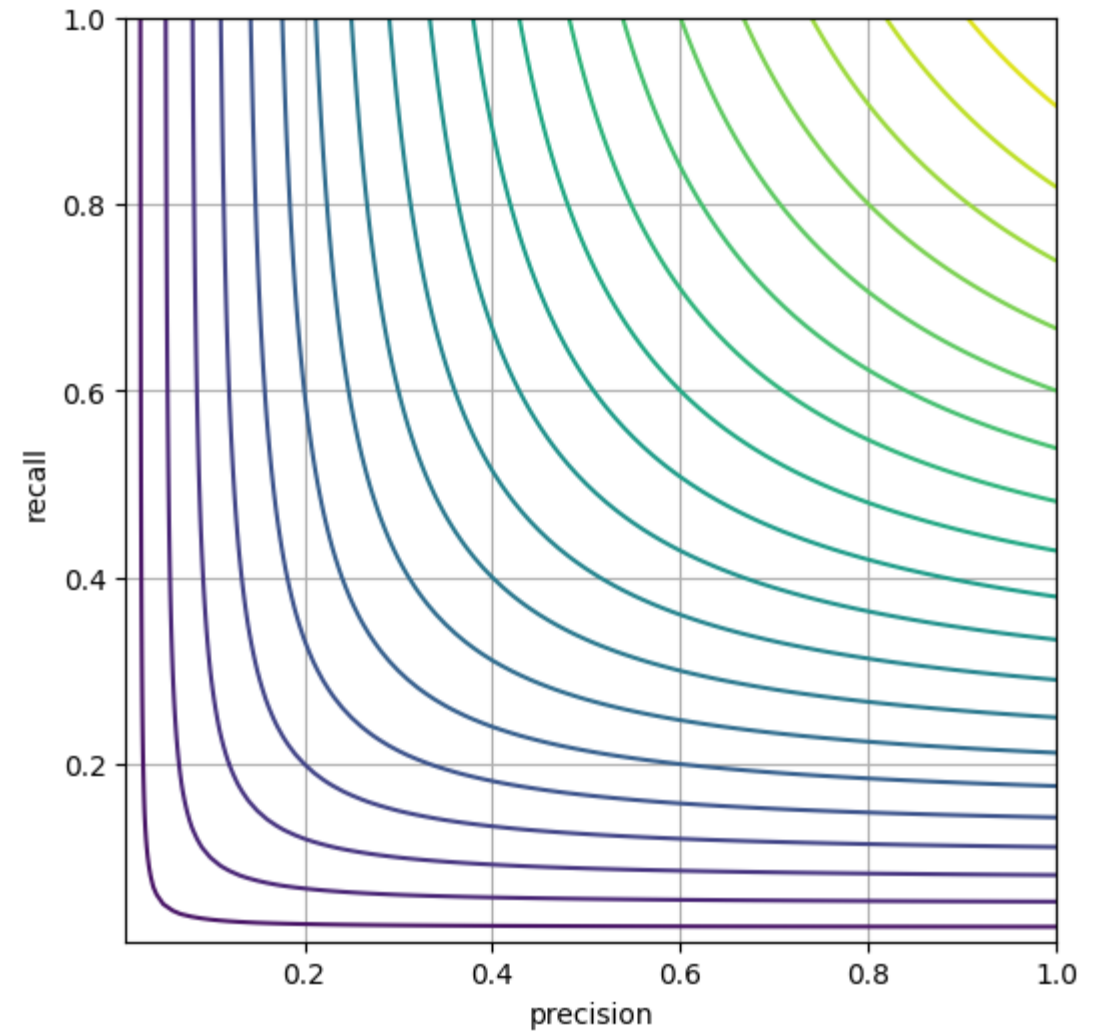
$$F = \frac{2 * precision * recall}{precision + recall}$$

Модель	Precision	recall	F
$a_1(x)$	0.55	0.55	0.55
$a_2(x)$	0.55	0.90	0.68



Гармоническое среднее (F-мера)

$$F = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall}$$

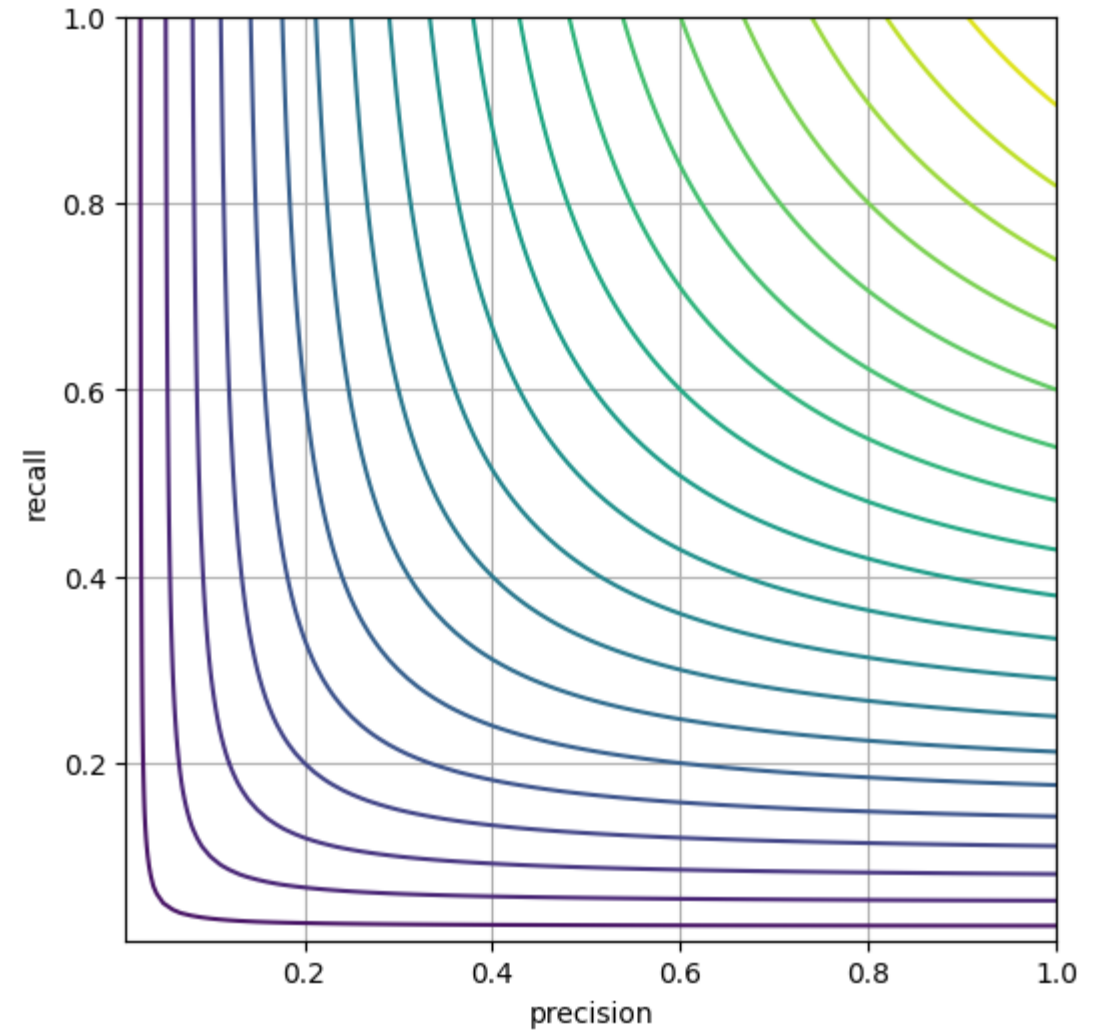


Гармоническое среднее (F-мера)

$$F = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall}$$

если $\beta=1$ Важнее recall и *precision*

$$F = \frac{(1+1^2)*precision*recall}{1^2*precision+recall} = \frac{2*precision*recall}{precision+recall}$$

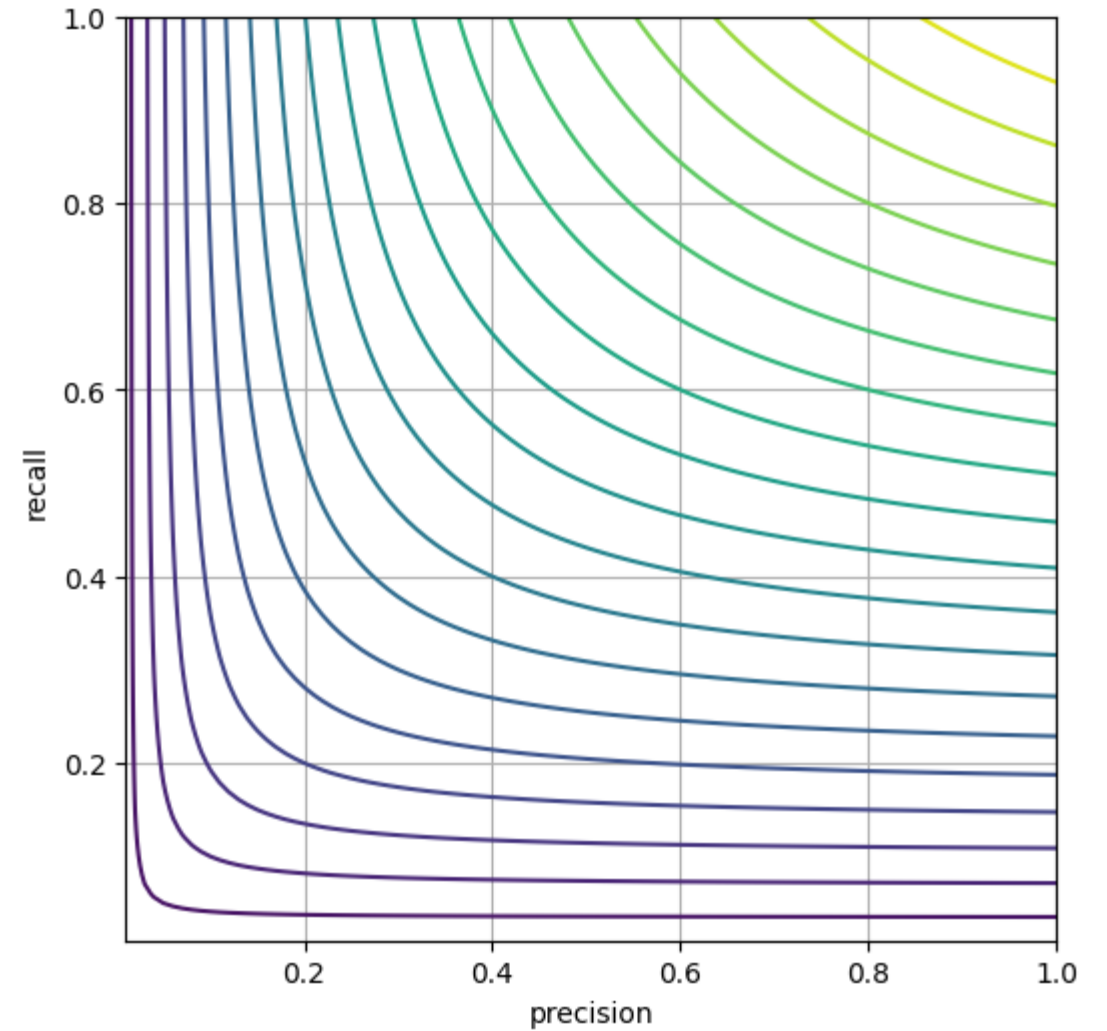


Гармоническое среднее (F-мера)

$$F = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall}$$

если $\beta > 1$ Важнее recall

$$\beta = 1.5$$

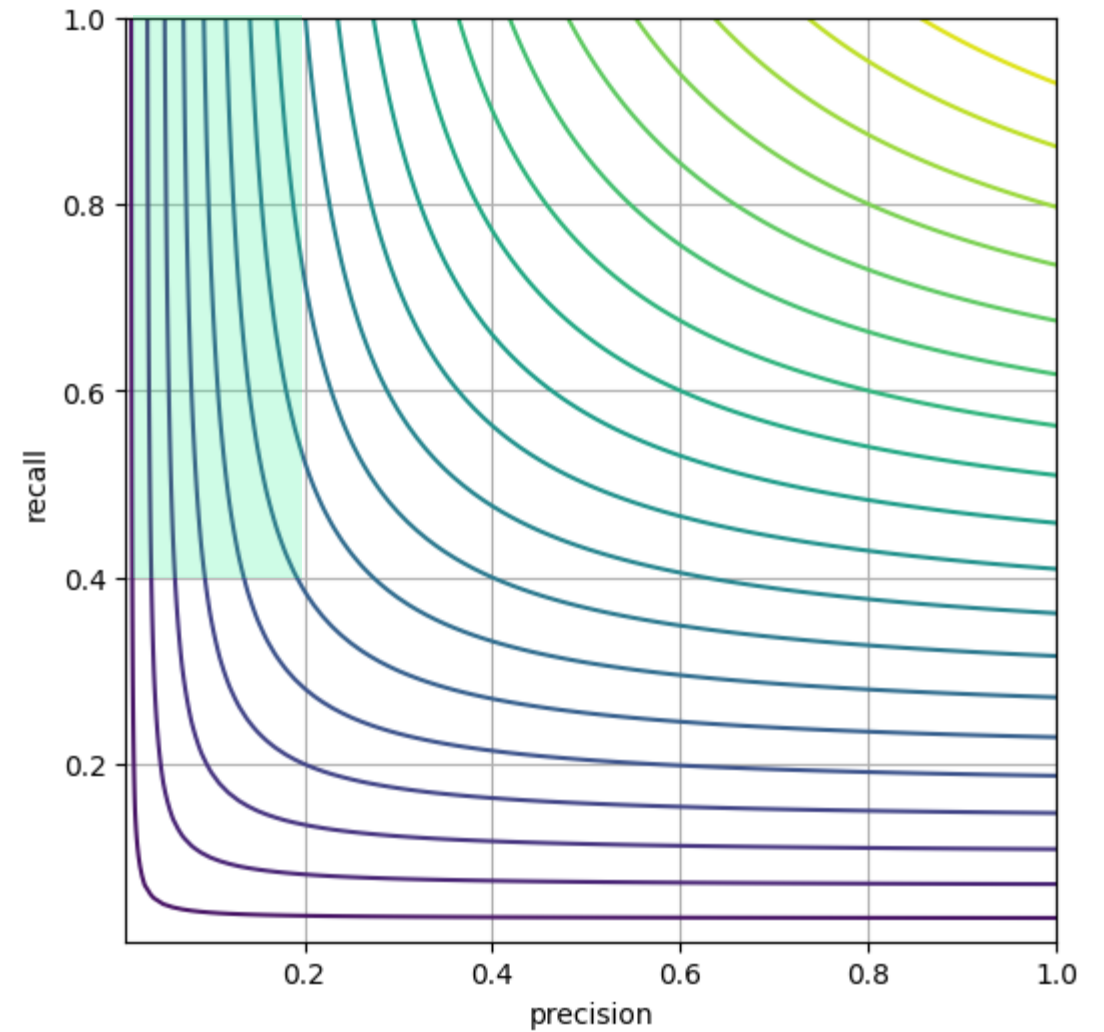


Гармоническое среднее (F-мера)

$$F = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall}$$

если $\beta > 1$ Важнее recall

$$\beta = 1.5$$

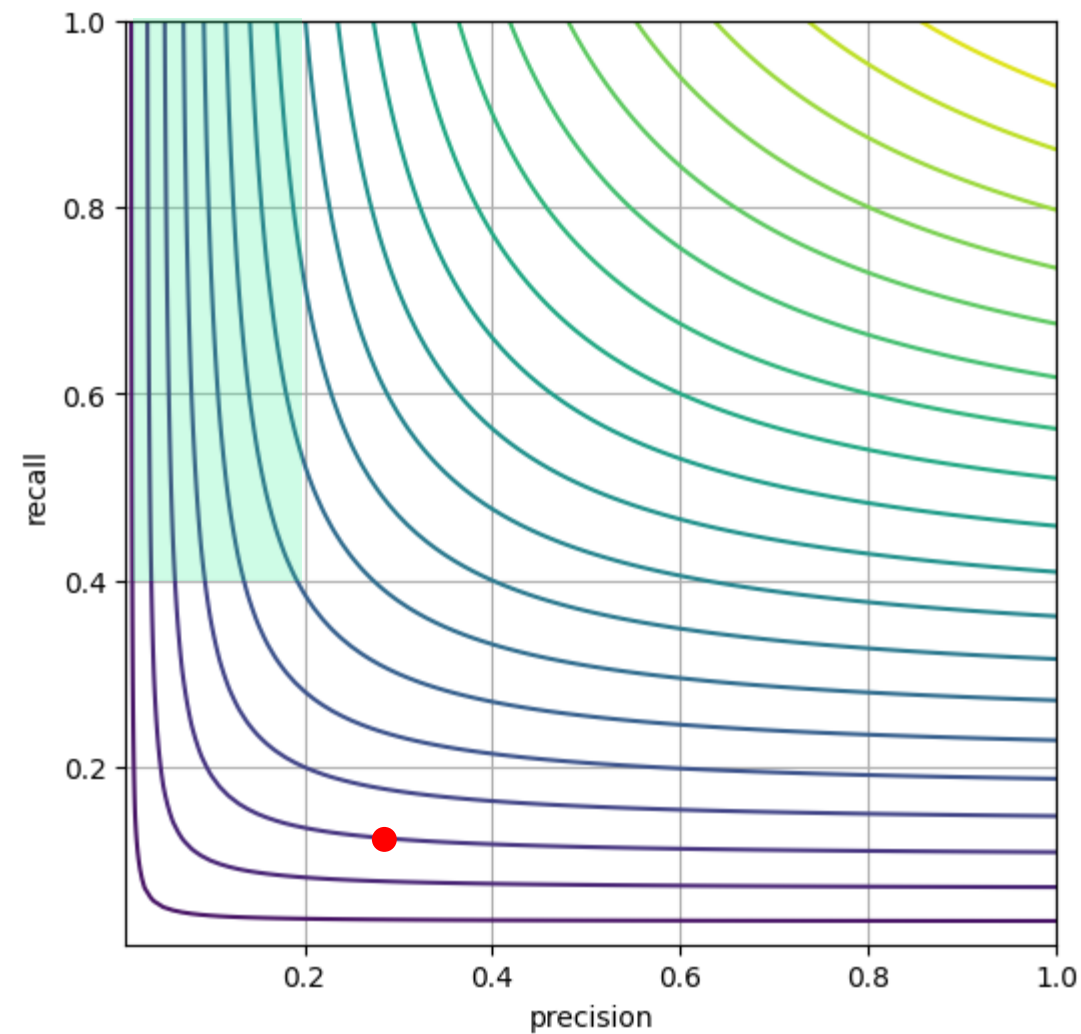


Гармоническое среднее (F-мера)

$$F = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall}$$

если $\beta > 1$ Важнее recall

$$\beta = 1.5$$

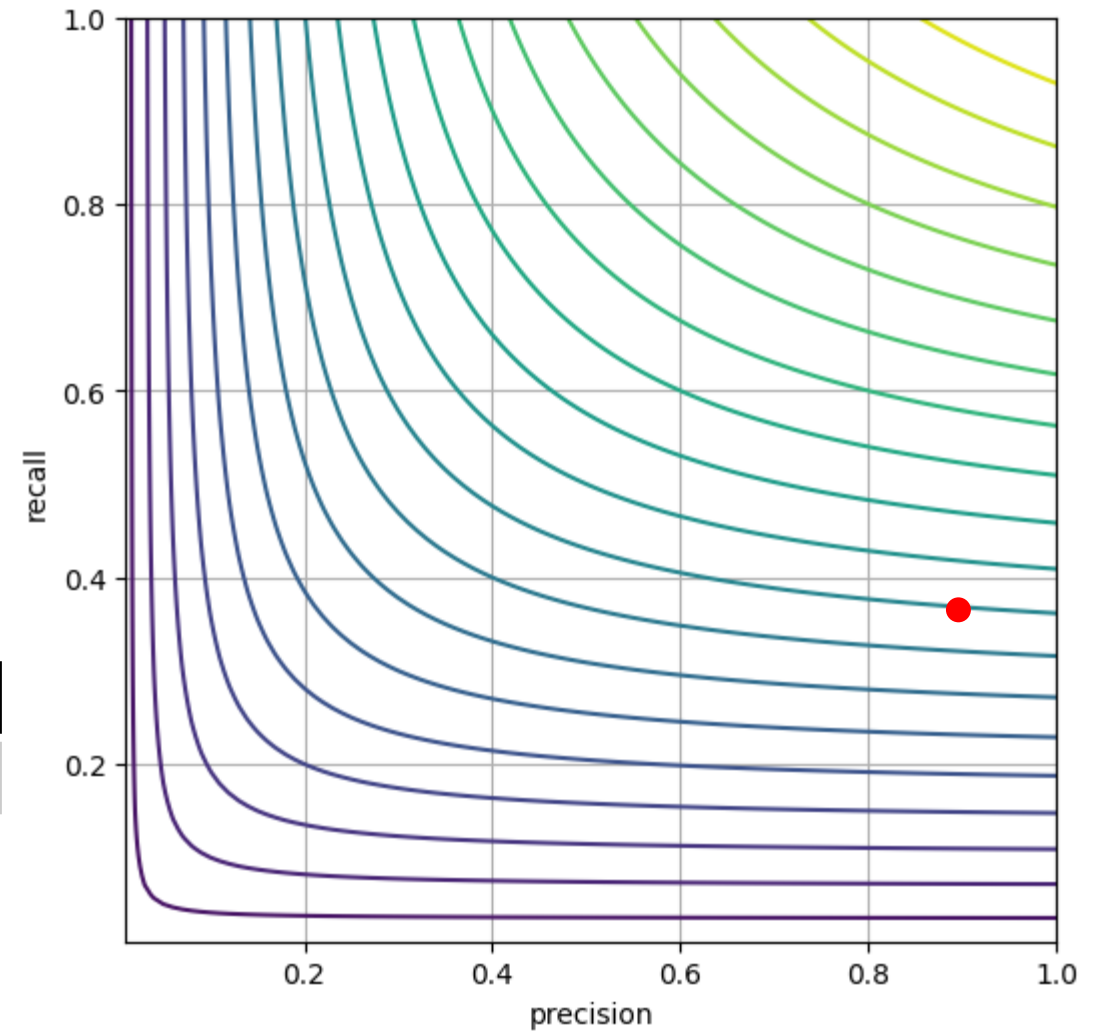


Гармоническое среднее (F-мера)

$$F = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall}$$

если $\beta > 1$ Важнее recall

Модель	Precision	recall	F1-Мера	F1.5-мера
$a_1(x)$	0.9	0.4	0.55	0.48

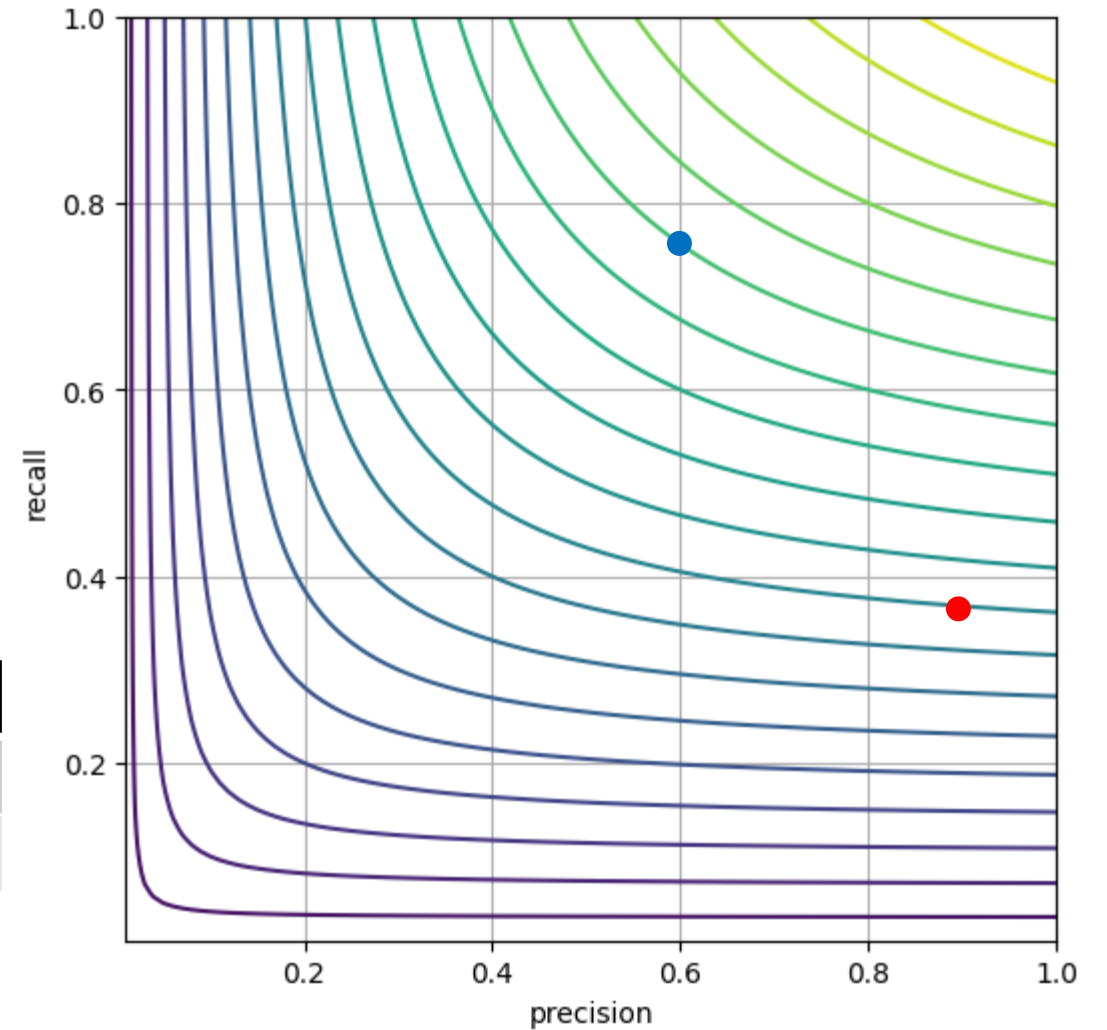


Гармоническое среднее (F-мера)

$$F = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall}$$

если $\beta > 1$ Важнее recall

Модель	Precision	recall	F1-Мера	F1.5-мера
$a_1(x)$	0.9	0.4	0.55	0.48
$a_2(x)$	0.6	0.9	0.55	0.78



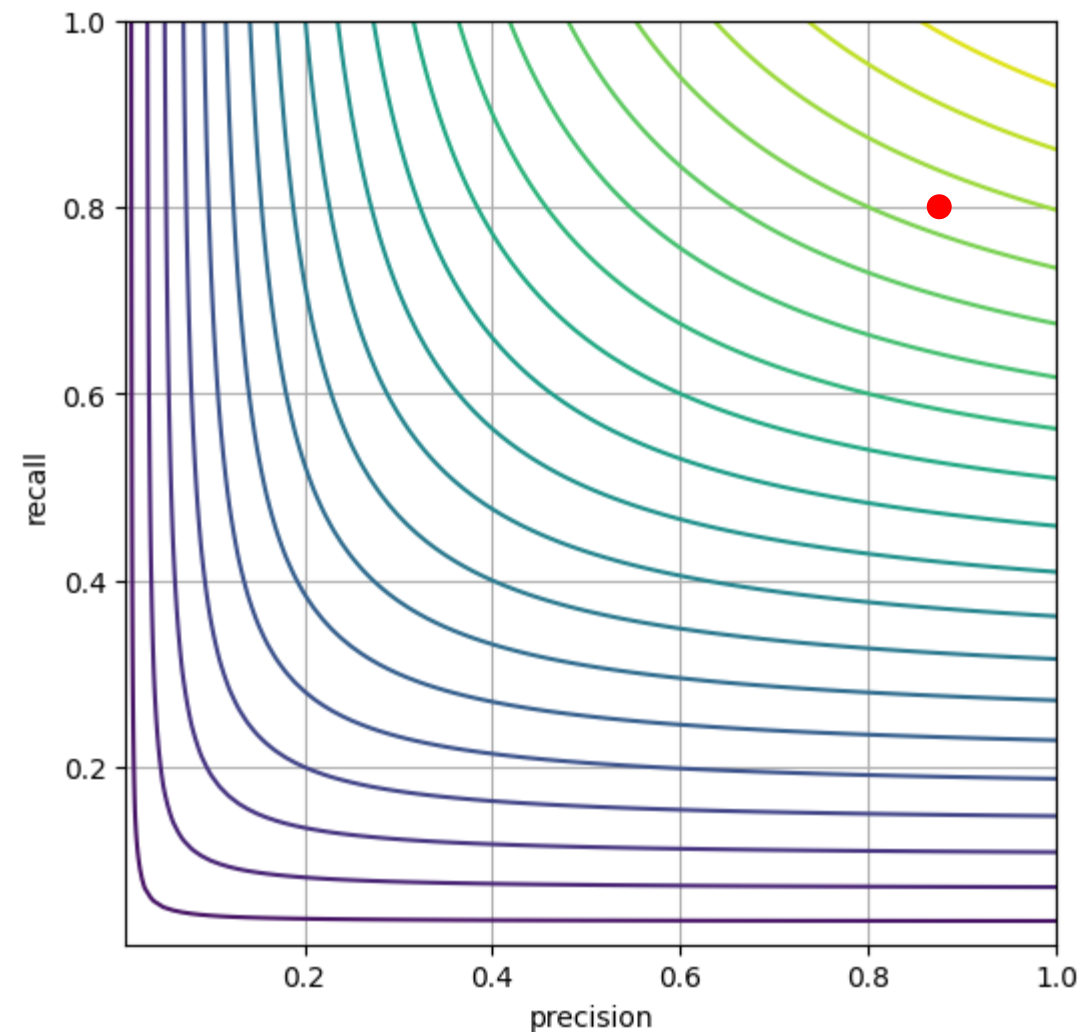
Гармоническое среднее (F-мера)

$$F = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall}$$

если $\beta > 1$ Важнее recall

$$\beta = 100$$

Модель	Precision	recall	F100-Мера
$a_1(x)$	0.9	0.8	0.8

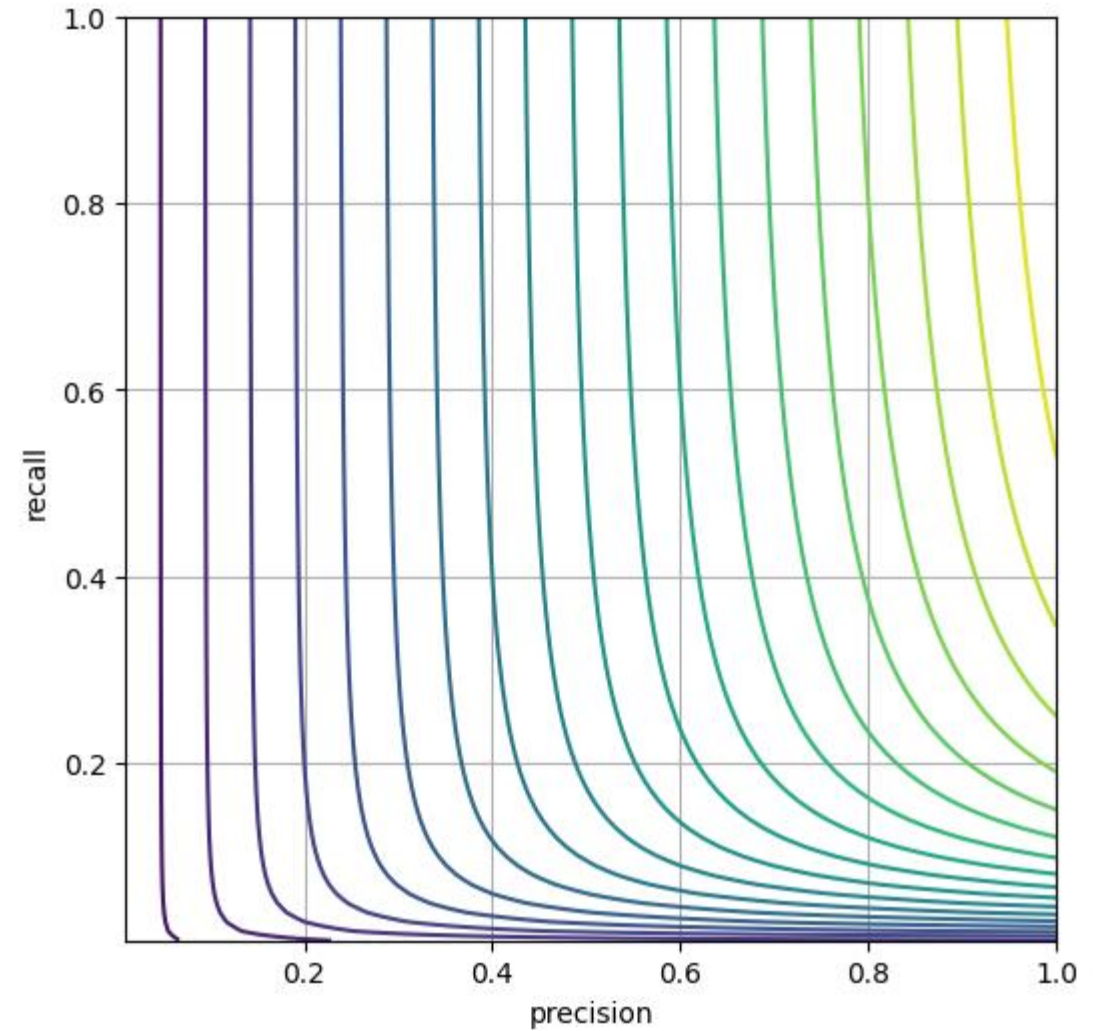


Гармоническое среднее (F-мера)

$$F = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall}$$

если $\beta < 1$ Важнее *precision*

$$\beta = 0.25$$

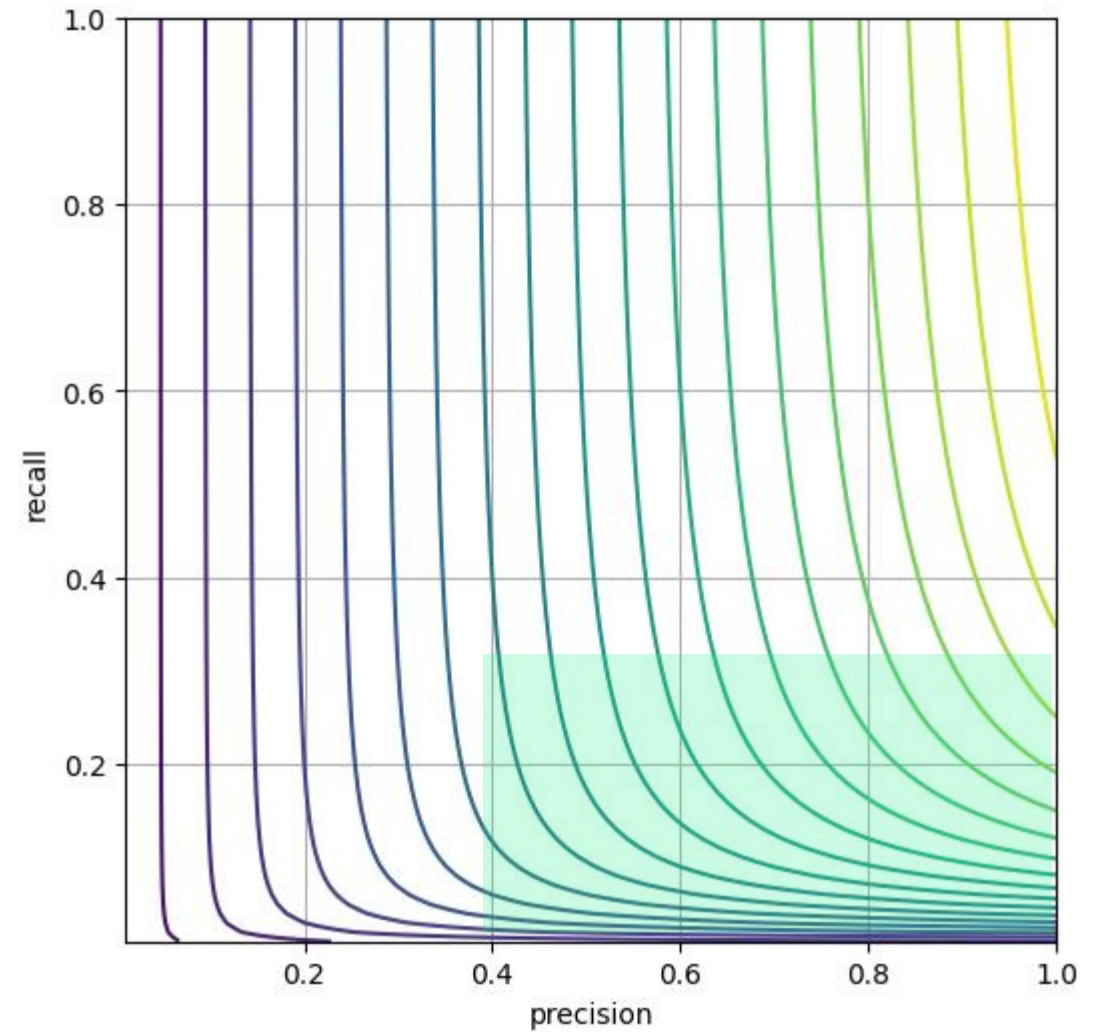


Гармоническое среднее (F-мера)

$$F = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall}$$

если $\beta < 1$ Важнее *precision*

$$\beta = 0.25$$

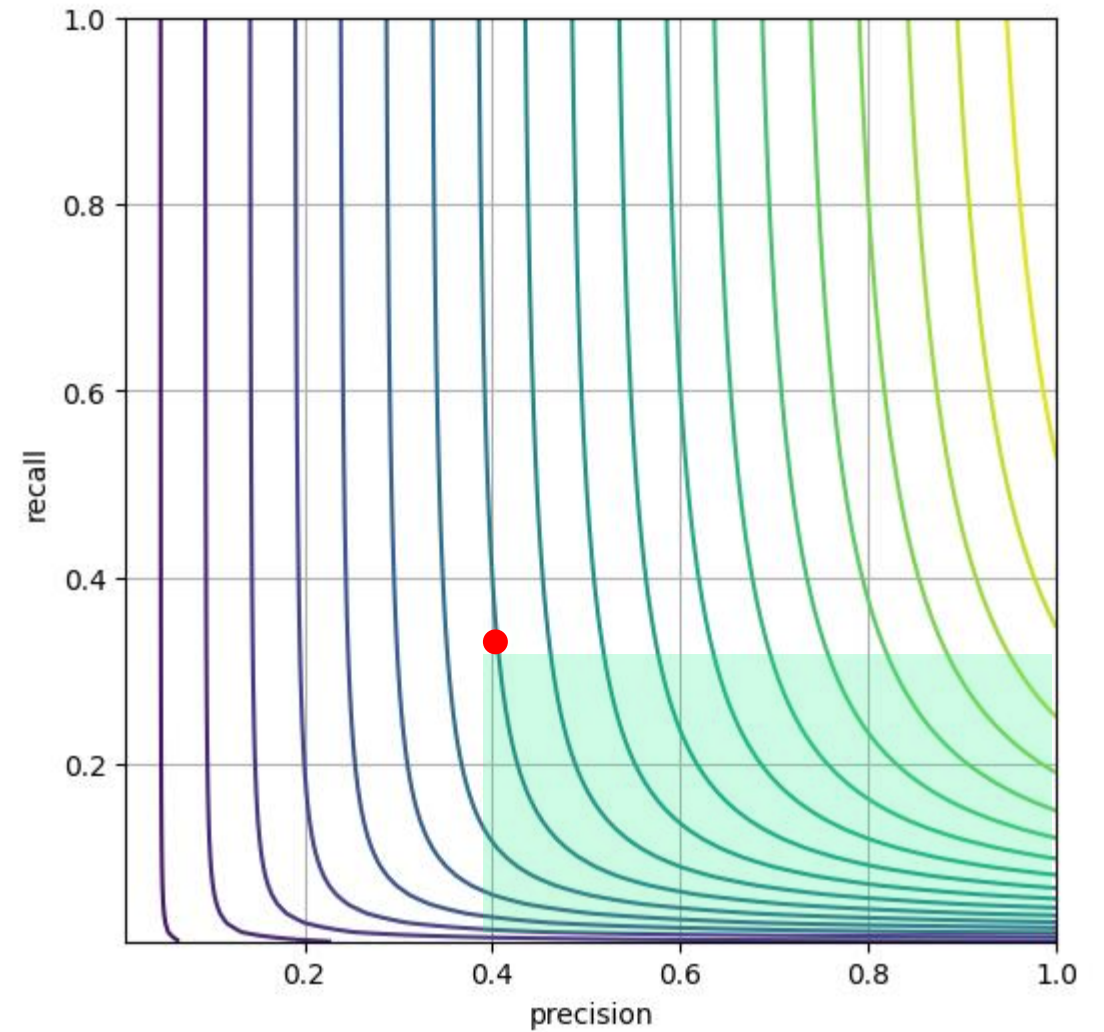


Гармоническое среднее (F-мера)

$$F = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall}$$

если $\beta < 1$ Важнее *precision*

$$\beta = 0.25$$

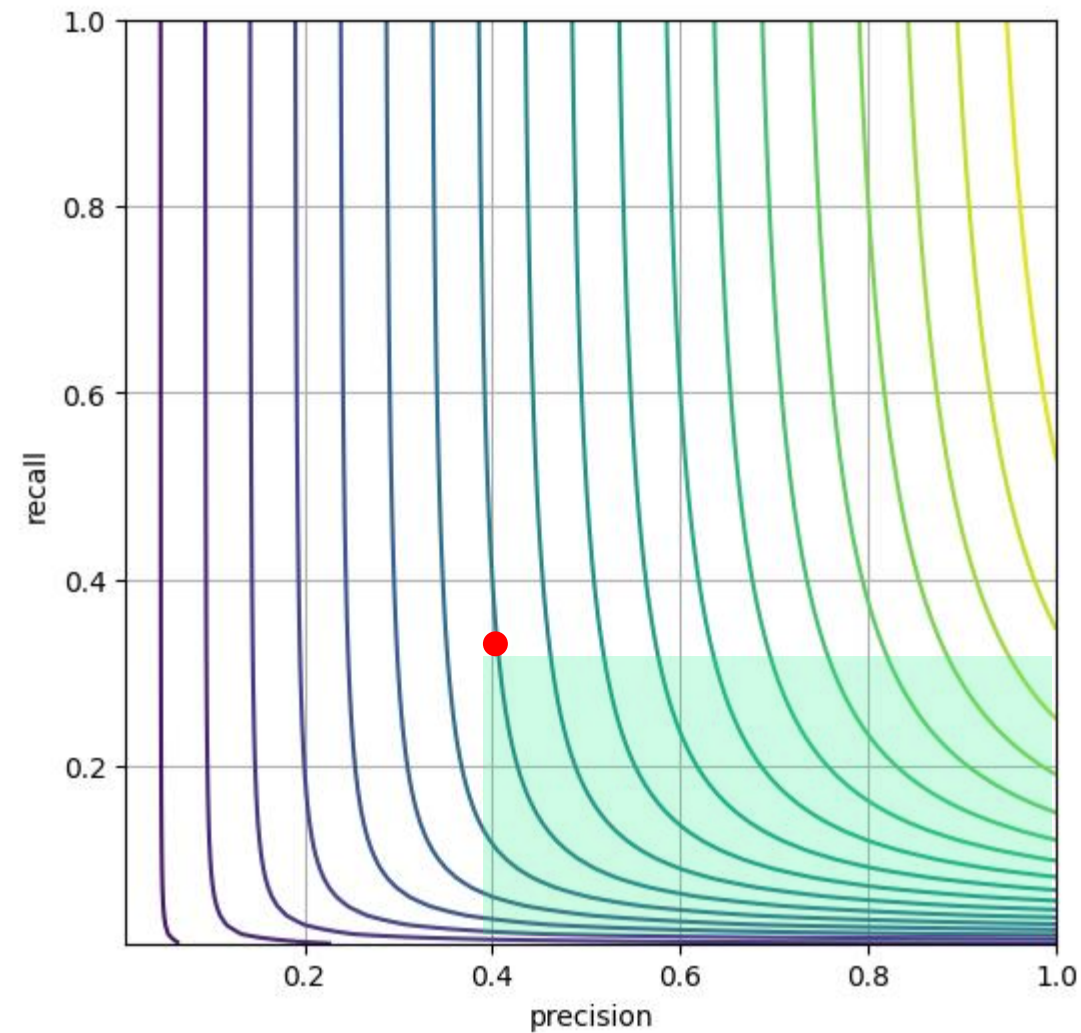


Гармоническое среднее (F-мера)

$$F = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall}$$

если $\beta < 1$ Важнее *precision*

$$\beta = 0.25$$

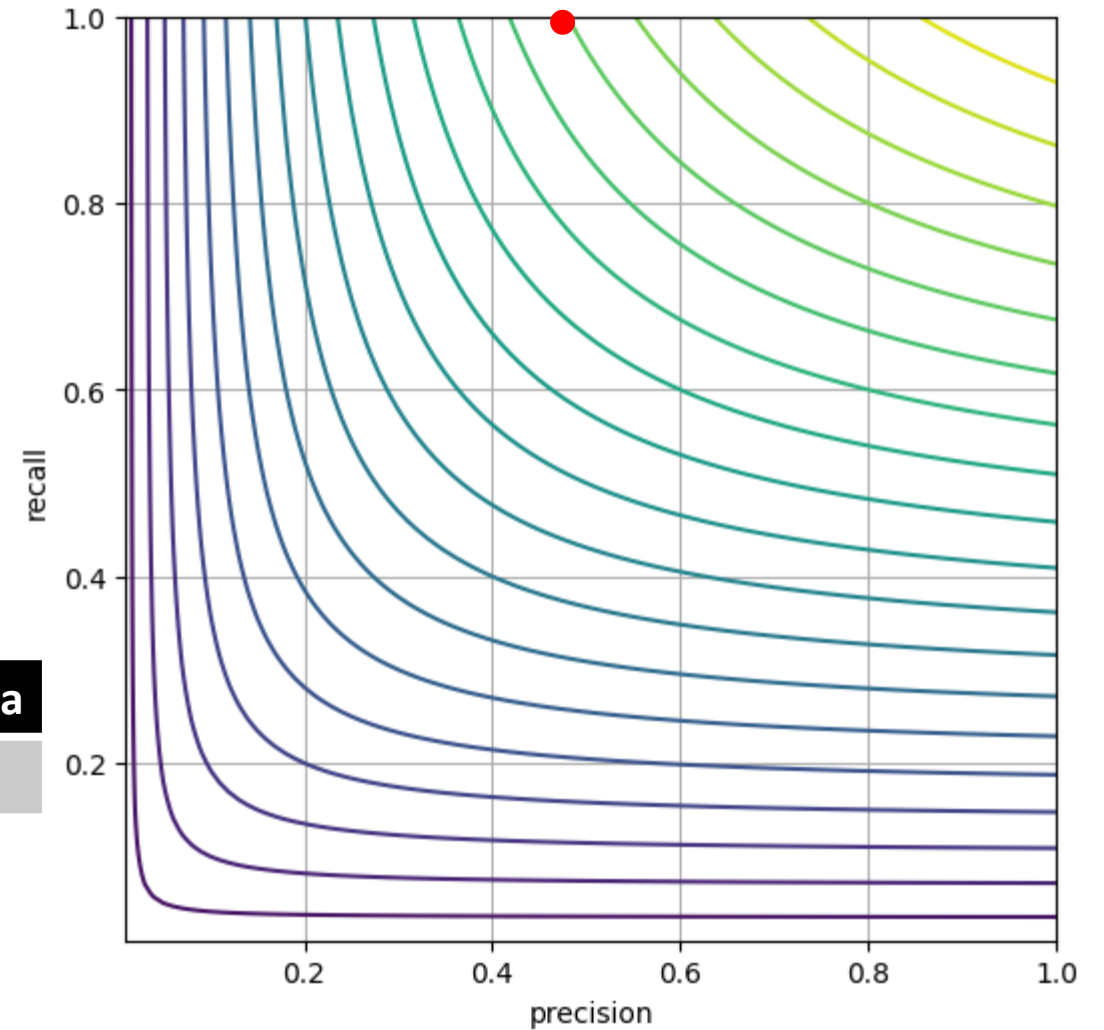


Гармоническое среднее (F-мера)

$$F = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall}$$

если $\beta < 1$ Важнее *precision*

Модель	Precision	recall	F1-Мера	F0.25-мера
$a_1(x)$	0.5	1	0.67	0.52

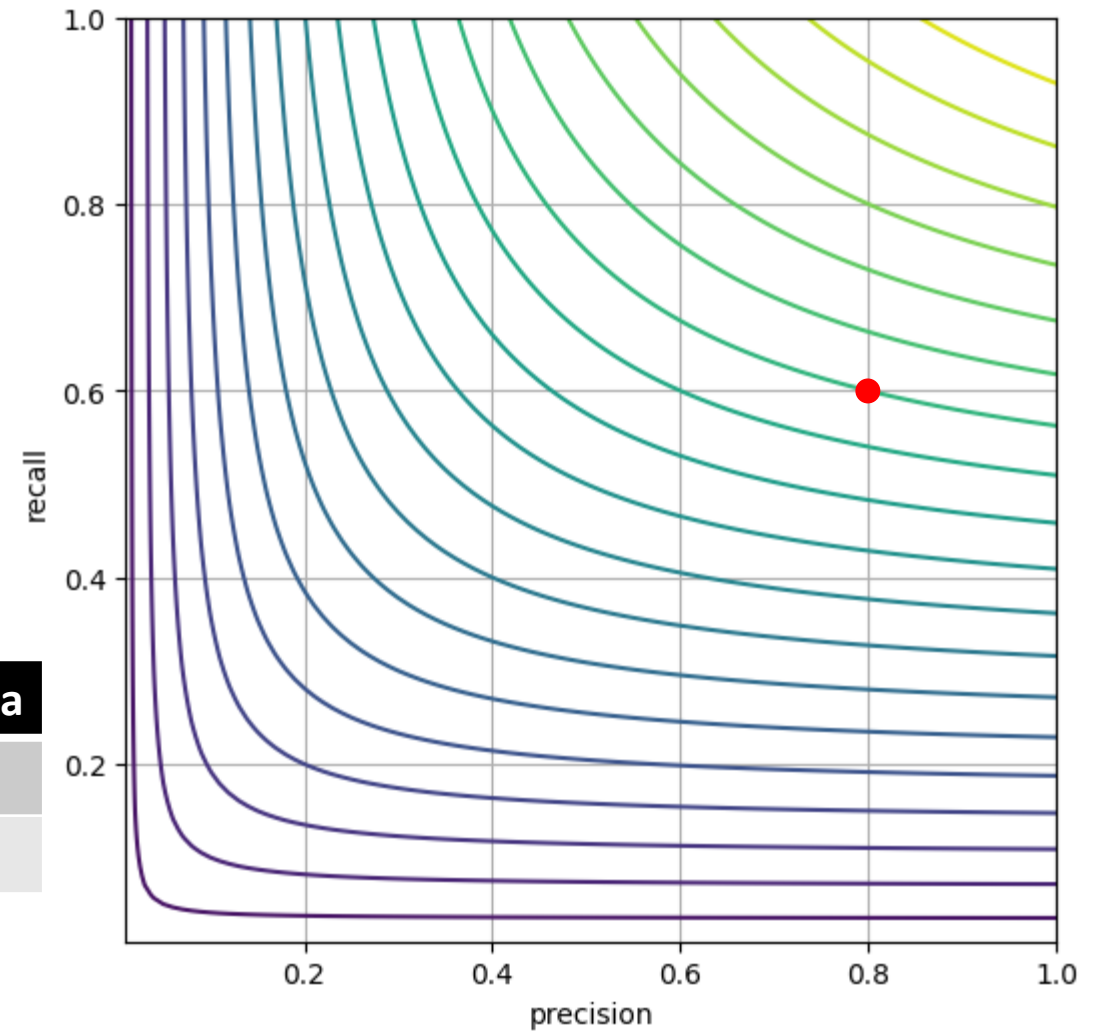


Гармоническое среднее (F-мера)

$$F = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall}$$

если $\beta < 1$ Важнее *precision*

Модель	Precision	recall	F1-Мера	F0.25-мера
$a_1(x)$	0.5	1	0.67	0.52
$a_2(x)$	0.8	0.6	0.69	0.78



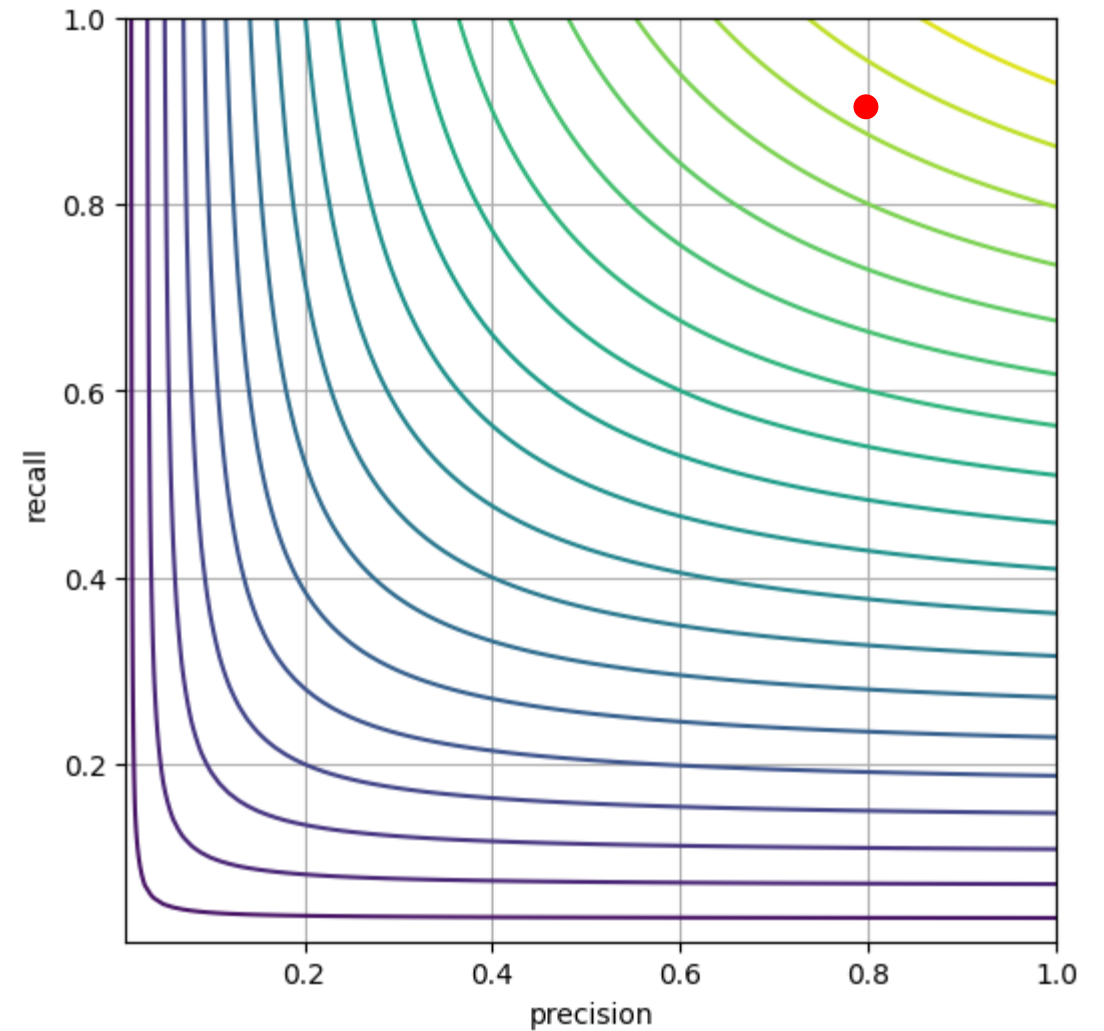
Гармоническое среднее (F-мера)

если $\beta = 0$ Важнее *precision*

$$F_0 = \frac{(1 + 0^2) * precision * recall}{0^2 * precision + recall}$$

$$F_0 = \frac{precision * recall}{recall} = precision$$

Модель	Precision	recall	F0-Мера
$a_1(x)$	0.8	0.9	0.8



WEIGHTED УСРЕДНЕНИЕ

доля правильных ответов

$$accuracy = \frac{1}{\ell} \sum_{i=1}^{\ell} [a(x_i) = y_i]$$

Точность

$$precision(a, X) = \frac{TP}{TP + FP}$$

Полнота

$$recall(a, X) = \frac{TP}{TP + FN}$$

F-мера

$$F = \frac{(1 + \beta^2) * precision * recall}{\beta^2 * precision + recall}$$

WEIGHTED УСРЕДНЕНИЕ метрика

Y	$a_1(x)$	$a_2(x)$
0	0	0
0	0	0
0	1	0
0	1	0
0	1	0
1	1	1
1	1	1
1	1	0
1	1	0
1	1	0

$a_1(x)$	Precision	Recall	F1-Mepa
Класс 0	1.0	0.4	0.57
Класс 1	0.62	1.0	0.77

Micro

Y	$a_1(x)$
0	0
0	0
0	1
0	1
0	1
1	1
1	1
1	1
1	1
1	1

$$accuracy = \frac{1}{\ell} \sum_{i=1}^{\ell} [a(x_i) = y_i]$$

<i>accuracy</i>	Precision	Recall	F1-Mepa
0.7			

Micro

Y	$a_1(x)$
0	0
0	0
0	1
0	1
0	1
1	1
1	1
1	1
1	1
1	1

precision_{micro}

=

Общее количество TP

Общее количество TP + Общее количество FP

<i>accuracy</i>	Precision	Recall	F1-Мера
0.7	0.7		

Micro

Y	$a_1(x)$
0	0
0	0
0	1
0	1
0	1
1	1
1	1
1	1
1	1
1	1

$$F_{micro} = \frac{2 * precision_{micro} * recall_{micro}}{precision_{micro} + recall_{micro}}$$

<i>accuracy</i>	Precision	Recall	F1-Mepa
0.7	0.7	0.7	0.7

Macro

Y	$a_1(x)$
0	0
0	0
0	1
0	1
0	1
1	1
1	1
1	1
1	1
1	1

$$Precision_{macro} = \frac{1}{N} \sum_{i=1}^N Precision_i$$

N — общее количество классов

$Precision_0$	$Precision_1$
0.62	1

$$Precision_{macro} = \frac{Precision_0 + Precision_1}{2} = 0.81$$

Macro

Y	$a_1(x)$
0	0
0	0
0	1
0	1
0	1
1	1
1	1
1	1
1	1
1	1

$$Recall_{macro} = \frac{1}{N} \sum_{i=1}^N Recall_i$$

N — общее количество классов

$Recall_0$	$Recall_1$
1.0	0.4

$$Precision_{macro} = \frac{Recall_0 + Recall_1}{2} = 0.7$$

Macro

Y	$a_1(x)$
0	0
0	0
0	1
0	1
0	1
1	1
1	1
1	1
1	1
1	1

$$F - \text{Мера}_{macro} = \frac{1}{N} \sum_{i=1}^N F - \text{Мера}_i$$

N – общее количество классов

$F - \text{Мера}_0$	$F - \text{Мера}_1$
0.76	0.57

$$F - \text{Мера}_{macro} = \frac{F - \text{Мера}_0 + F - \text{Мера}_1}{2} = 0.67$$

Weighted

$$Precision_{weighted} = \sum_{i=0}^N \frac{\text{Количество элементов в классе } i}{\text{Количество элементов}} * Precision_i$$

$$Recall_{weighted} = \sum_{i=0}^N \frac{\text{Количество элементов в классе } i}{\text{Количество элементов}} * Recall_i$$

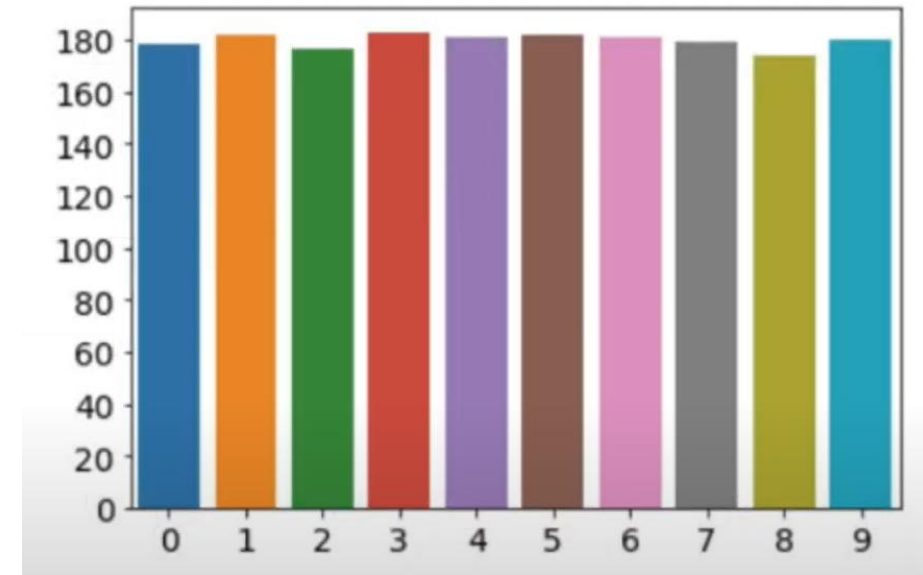
$$F - \text{Мера}_{weighted} = \sum_{i=0}^N \frac{\text{Количество элементов в классе } i}{\text{Количество элементов}} * F - \text{Мера}_i$$

N — общее количество классов

Применение

Micro – когда у нас данные сбалансированы

MNIST- образцы рукописного написания цифр

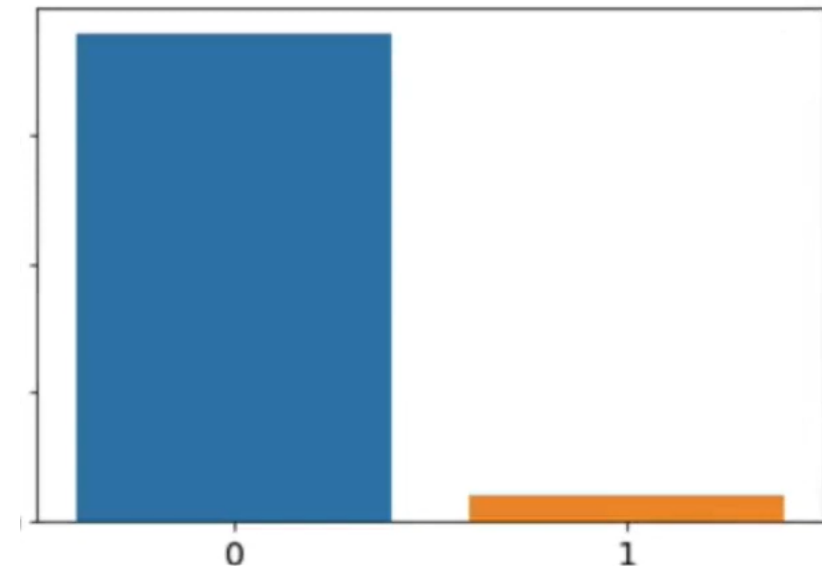


Применение

Macro – когда у нас данные несбалансированы

Антифрод – выявление мошеннических транзакций

	Precision	Recall	F1-Mepa
Класс 0	0.95	1.0	0.97
Класс 1	0	0	0
Micro	0.95	0.95	0.95
Macro	0.47	0.95	0.93

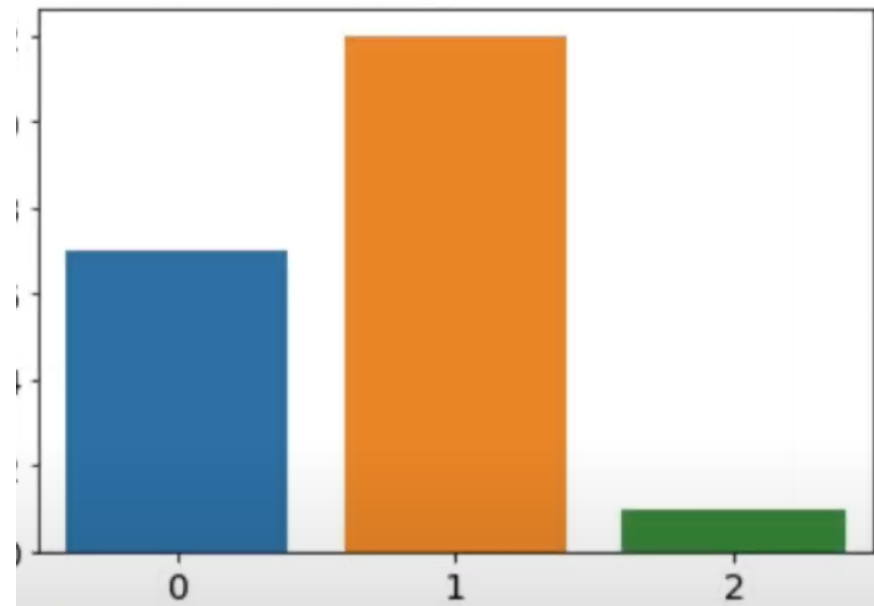


Применение

Weighted – когда необходимо дать классу вес согласно его количеству

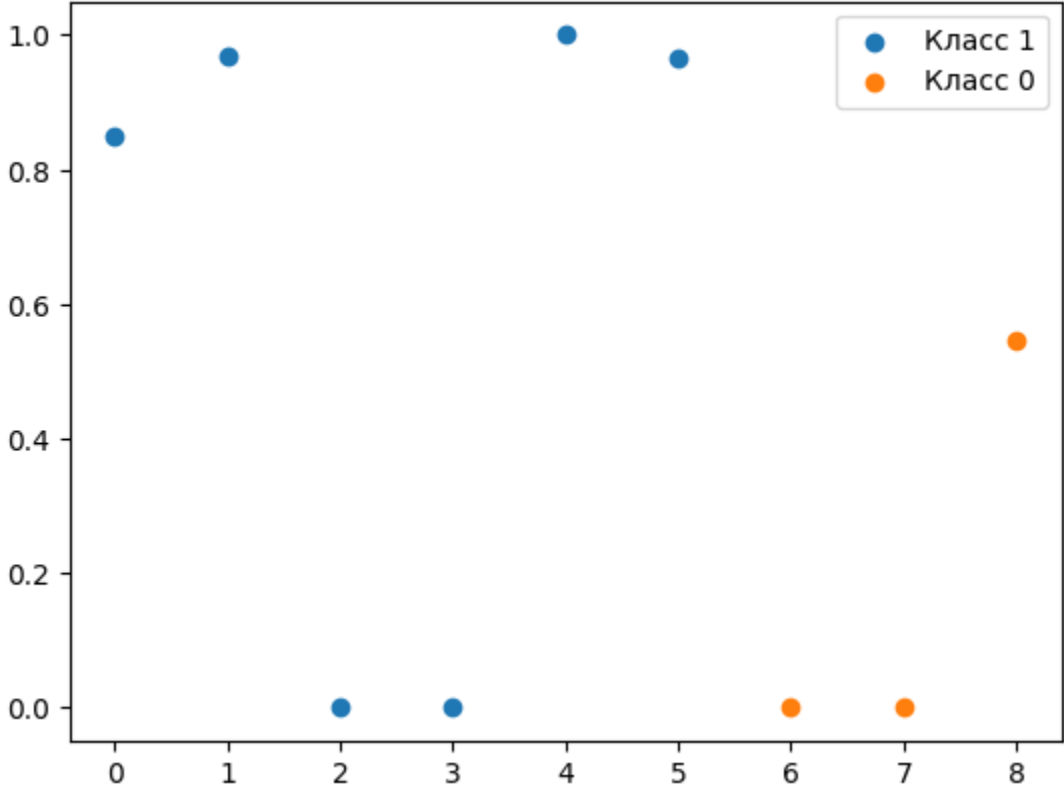
MNIST- образцы рукописного написания цифр

	Precision	Recall	F1-Мера
Класс 0	1	1.0	1
Класс 1	0.92	1.0	0.96
Класс 2	0	0	0
Micro	0.95	0.95	0.95
Macro	0.64	0.67	0.65
Weighted	0.90	0.95	0.93



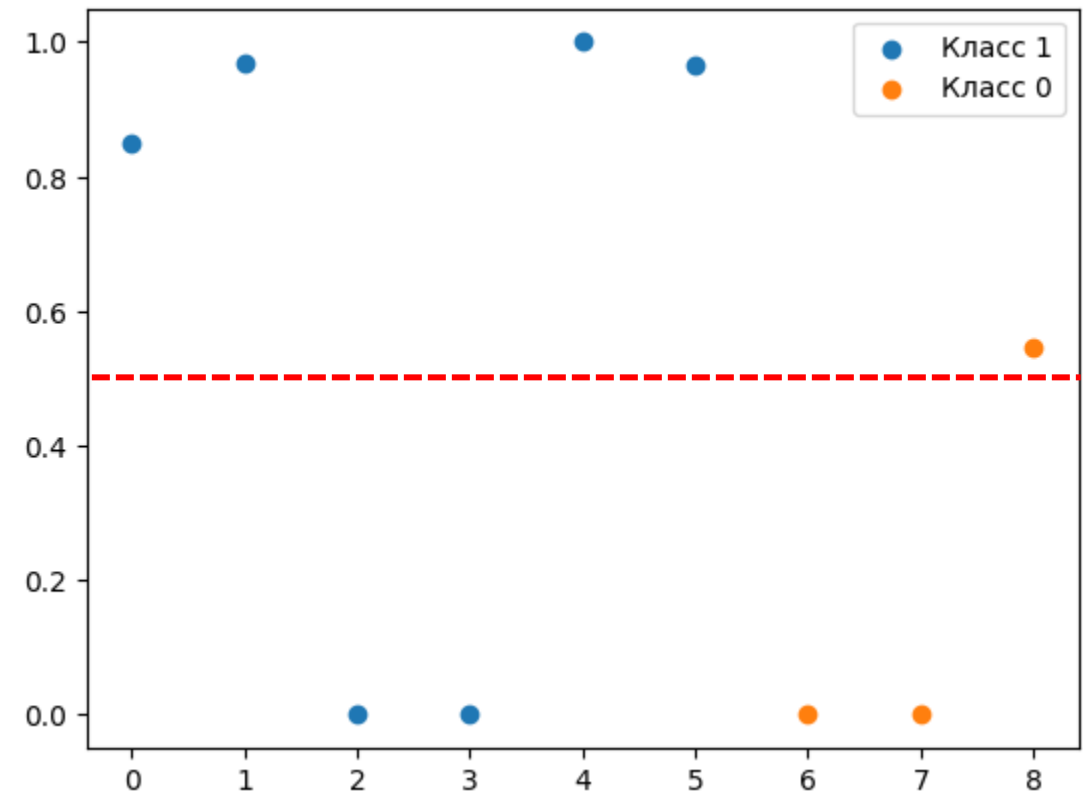
PR-AUC

Класс 0	Класс 1
0.1509434	0.8490566
0.03125	0.96875
1.	0.
1.	0.
1.	0.
1	0
0.	1.
0.03370787	0.96629213
0.45454545	0.54545455



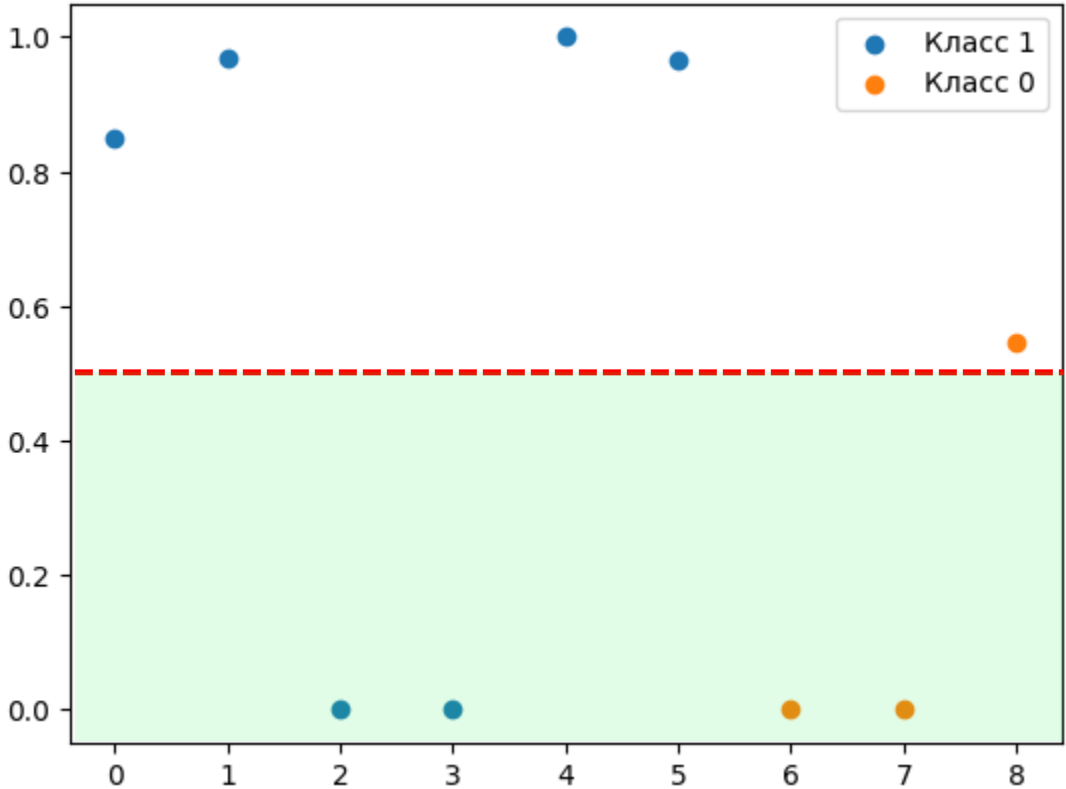
Класс 0	Класс 1
0.1509434	0.8490566
0.03125	0.96875
1.	0.
1.	0.
1.	0.
1	0
0.	1.
0.03370787	0.96629213
0.45454545	0.54545455

Порог = 0.5



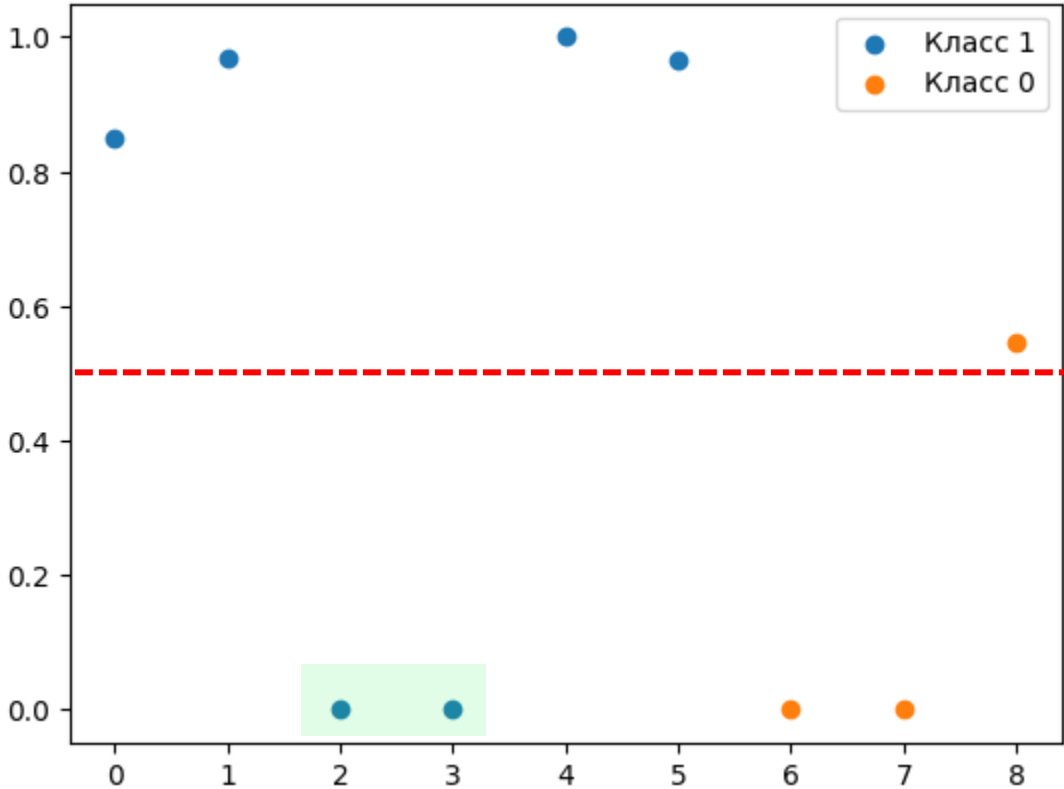
Класс 0	Класс 1
0.1509434	0.8490566
0.03125	0.96875
1.	0.
1.	0.
1.	0.
1	0
0.	1.
0.03370787	0.96629213
0.45454545	0.54545455

Порог = 0.5



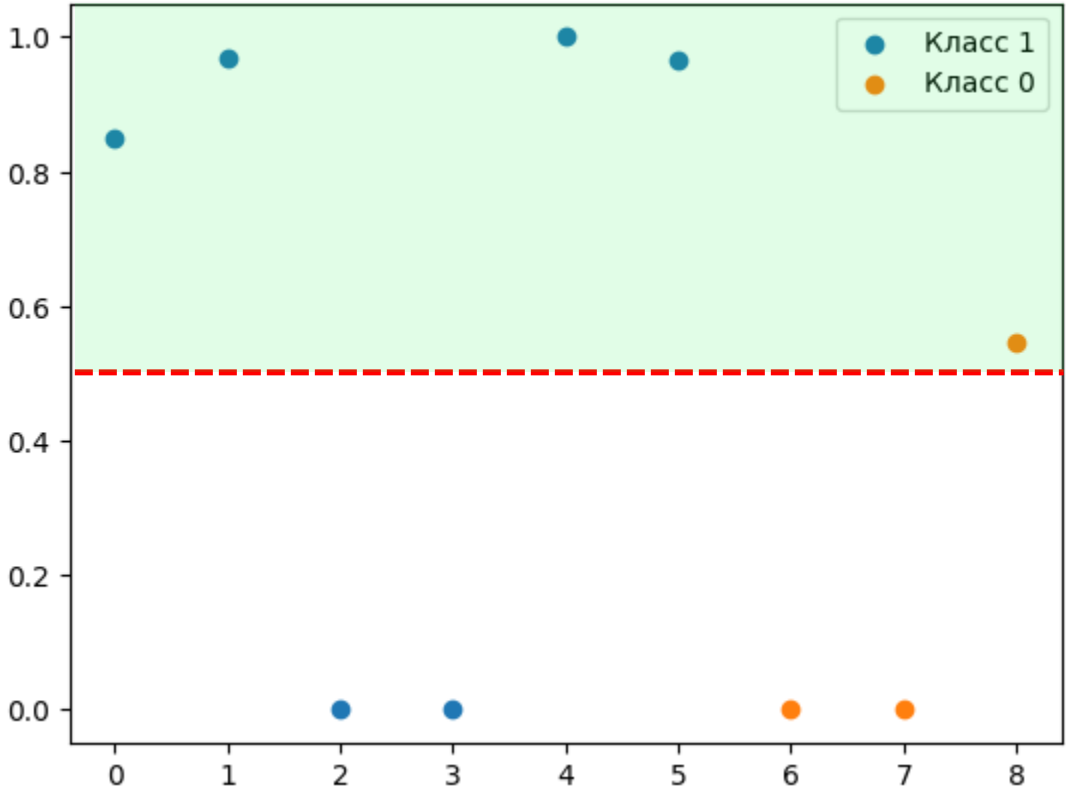
Класс 0	Класс 1
0.1509434	0.8490566
0.03125	0.96875
1.	0.
1.	0.
1.	0.
1	0
0.	1.
0.03370787	0.96629213
0.45454545	0.54545455

Порог = 0.5



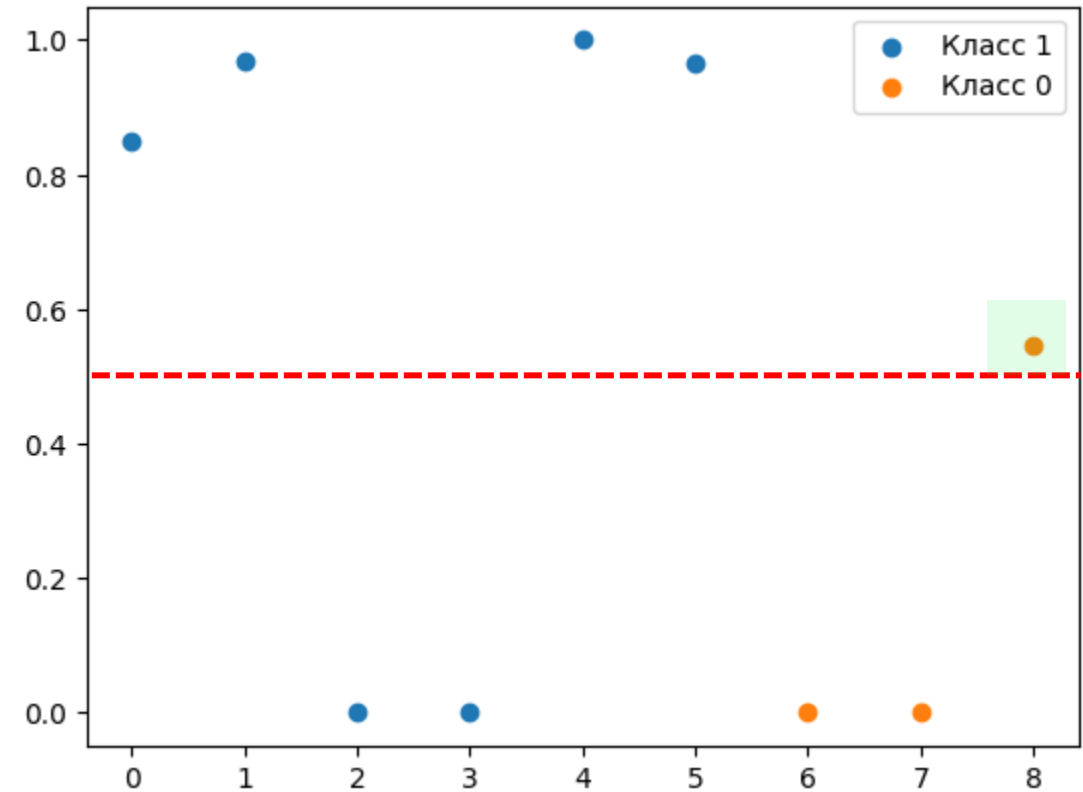
Класс 0	Класс 1
0.1509434	0.8490566
0.03125	0.96875
1.	0.
1.	0.
1.	0.
1	0
0.	1.
0.03370787	0.96629213
0.45454545	0.54545455

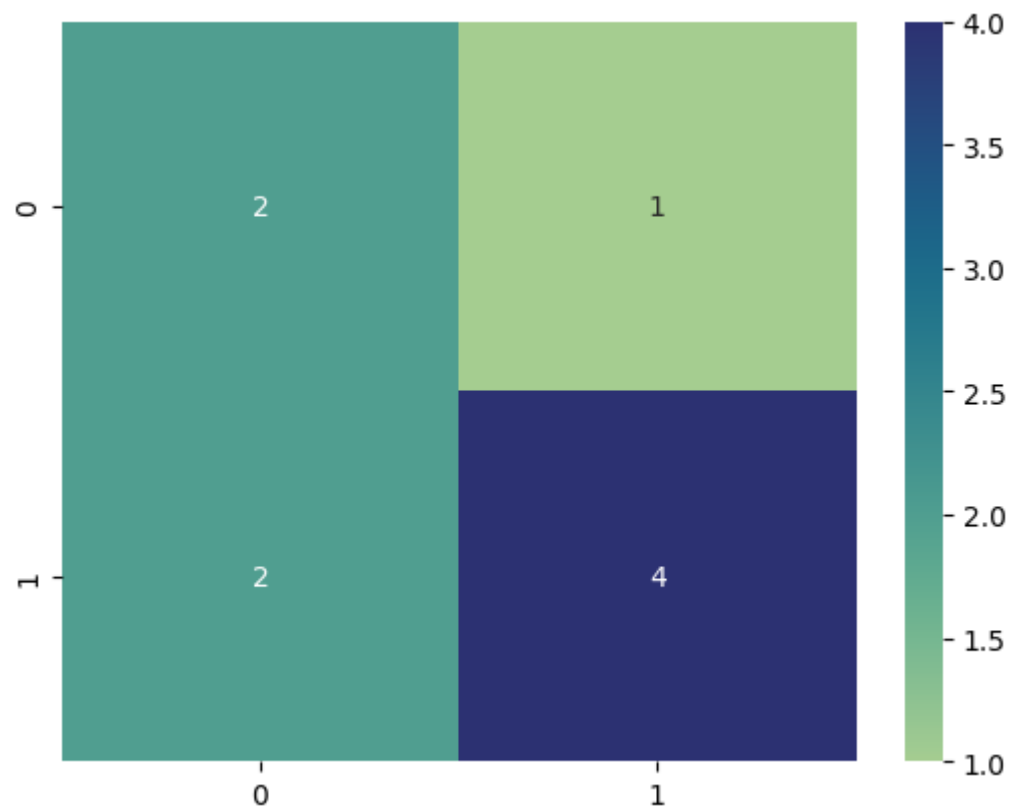
Порог = 0.5



Класс 0	Класс 1
0.1509434	0.8490566
0.03125	0.96875
1.	0.
1.	0.
1.	0.
1	0
0.	1.
0.03370787	0.96629213
0.45454545	0.54545455

Порог = 0.5

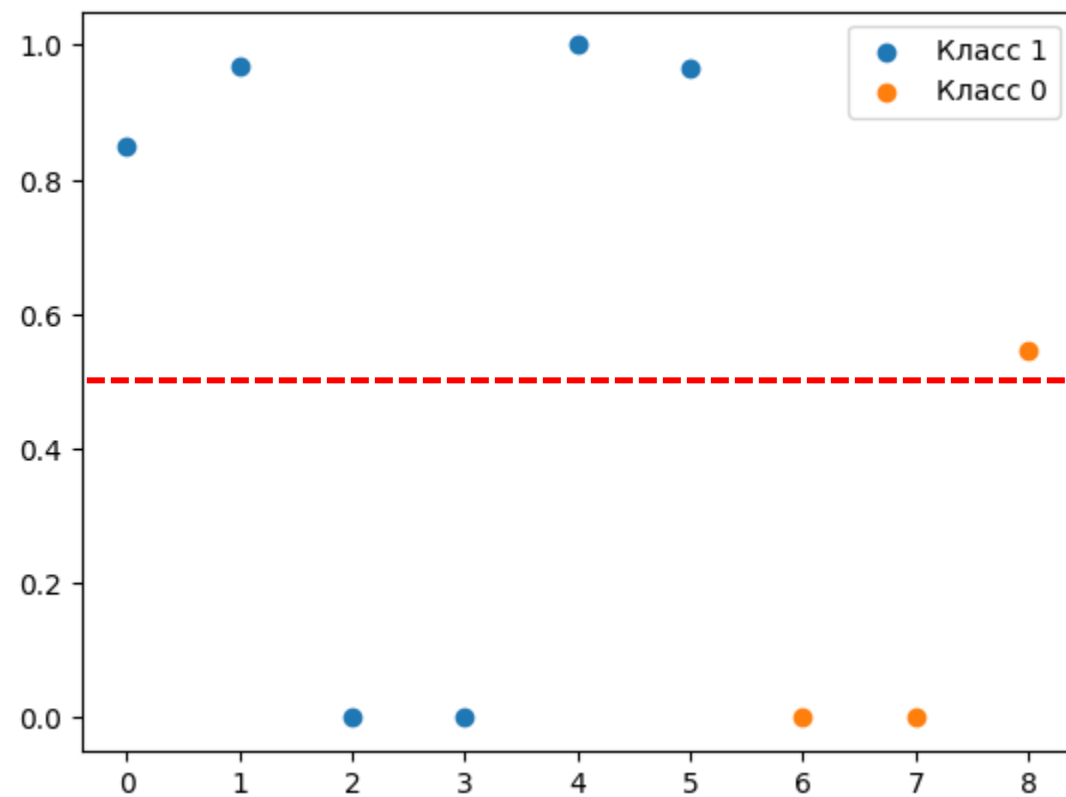


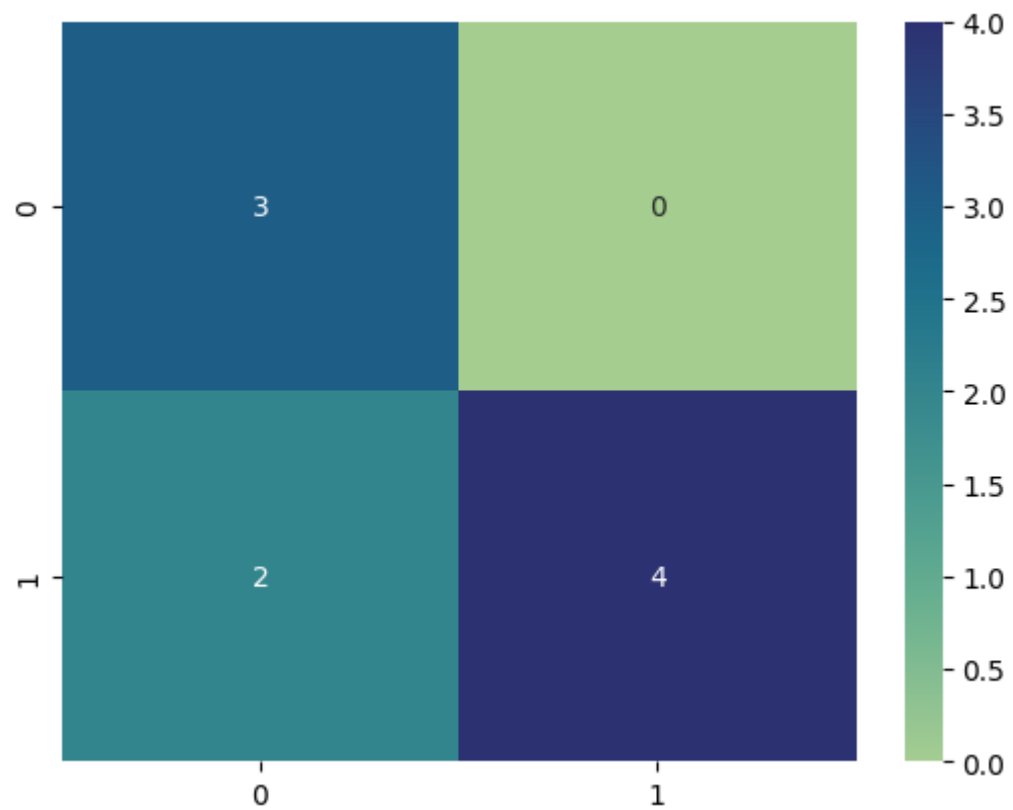


Порог = 0.5

Precision = 0.8

Recall = 0.66

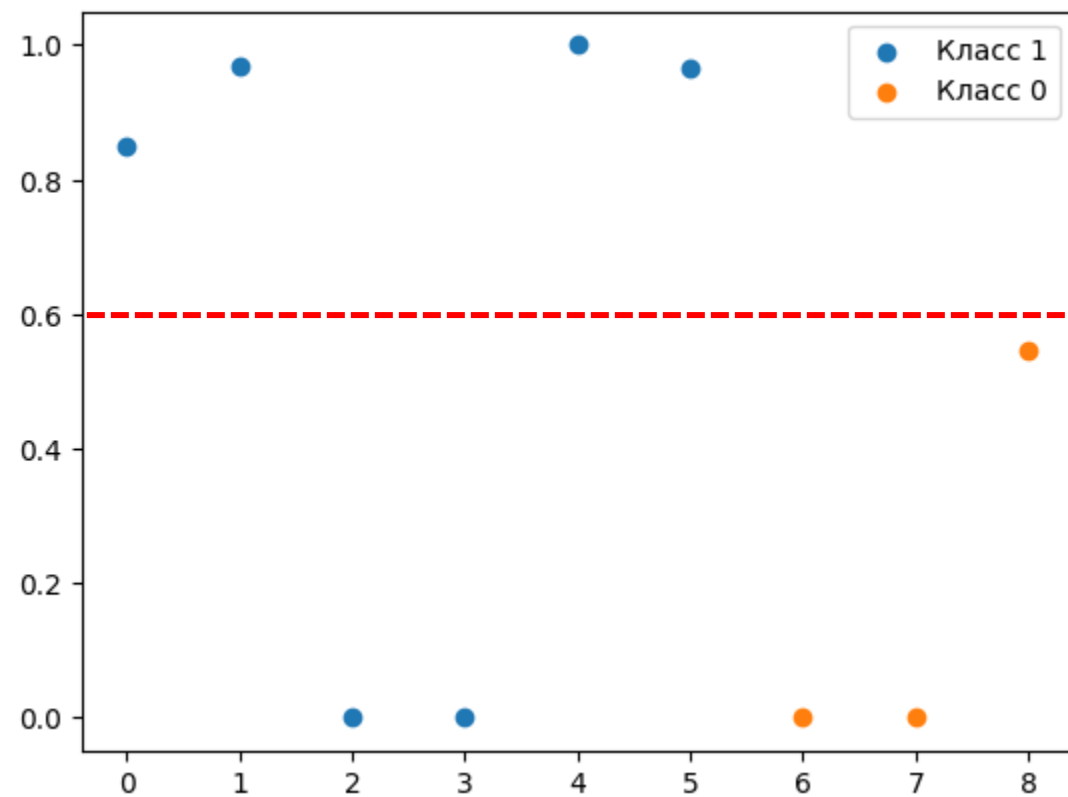


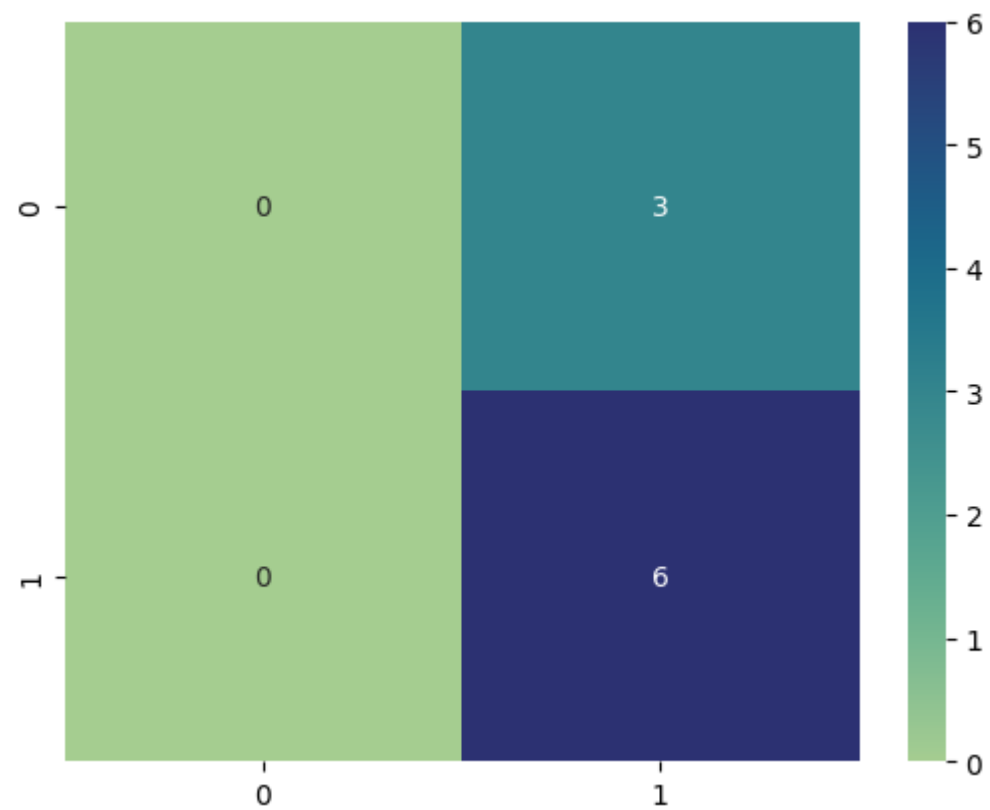


Порог = 0.6

Precision = 1

Recall = 0.66

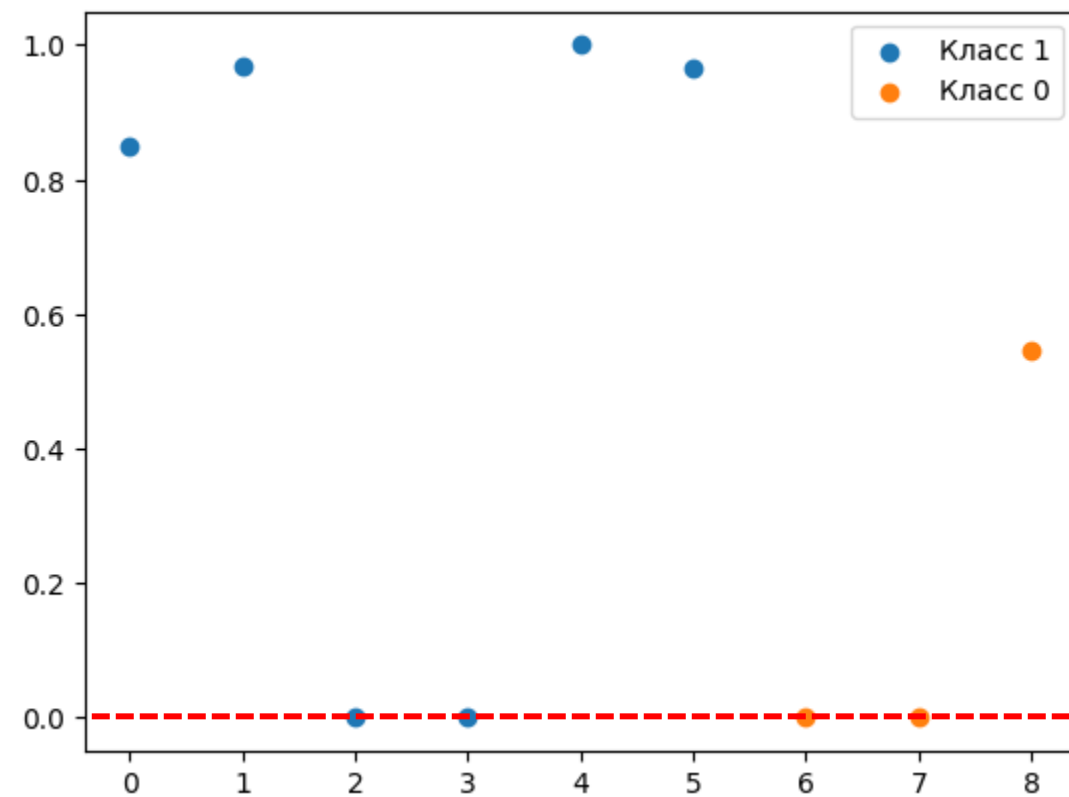




Порог = 0.0

Precision = 0.66

Recall = 1



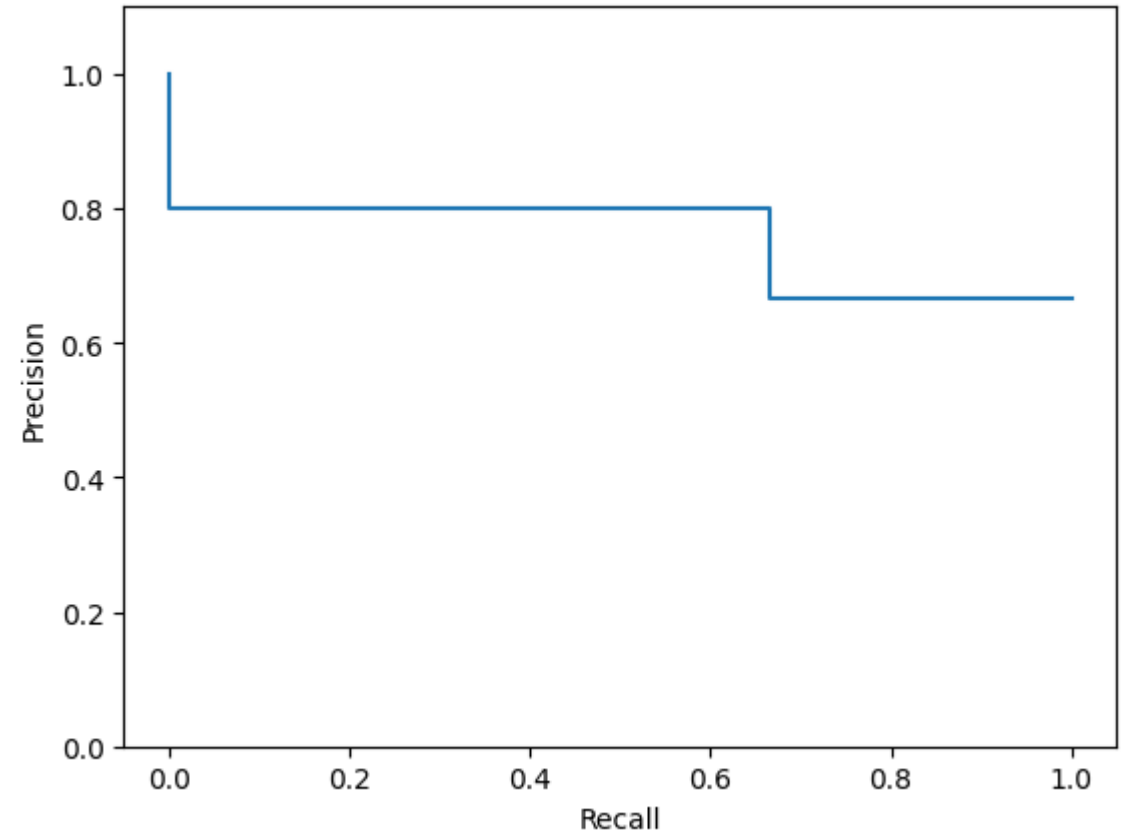
PR-кривая (PR-curve)

Левая точка: $(0,1)$

Правая точка: $(1,r)$

r -доля положительных объектов

Для идеального классификатора проходит через $(1,1)$



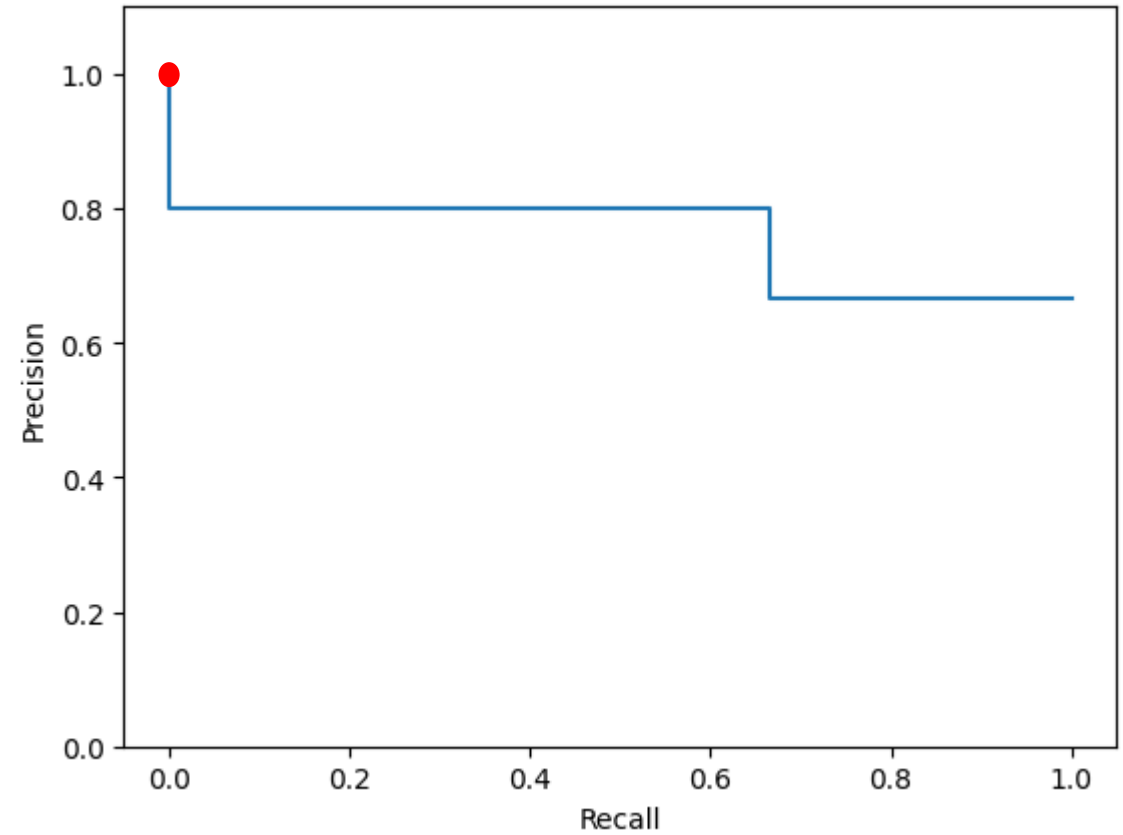
PR-кривая (PR-curve)

Левая точка: $(0,1)$

Правая точка: $(1,r)$

r -доля положительных объектов

Для идеального классификатора проходит через $(1,1)$



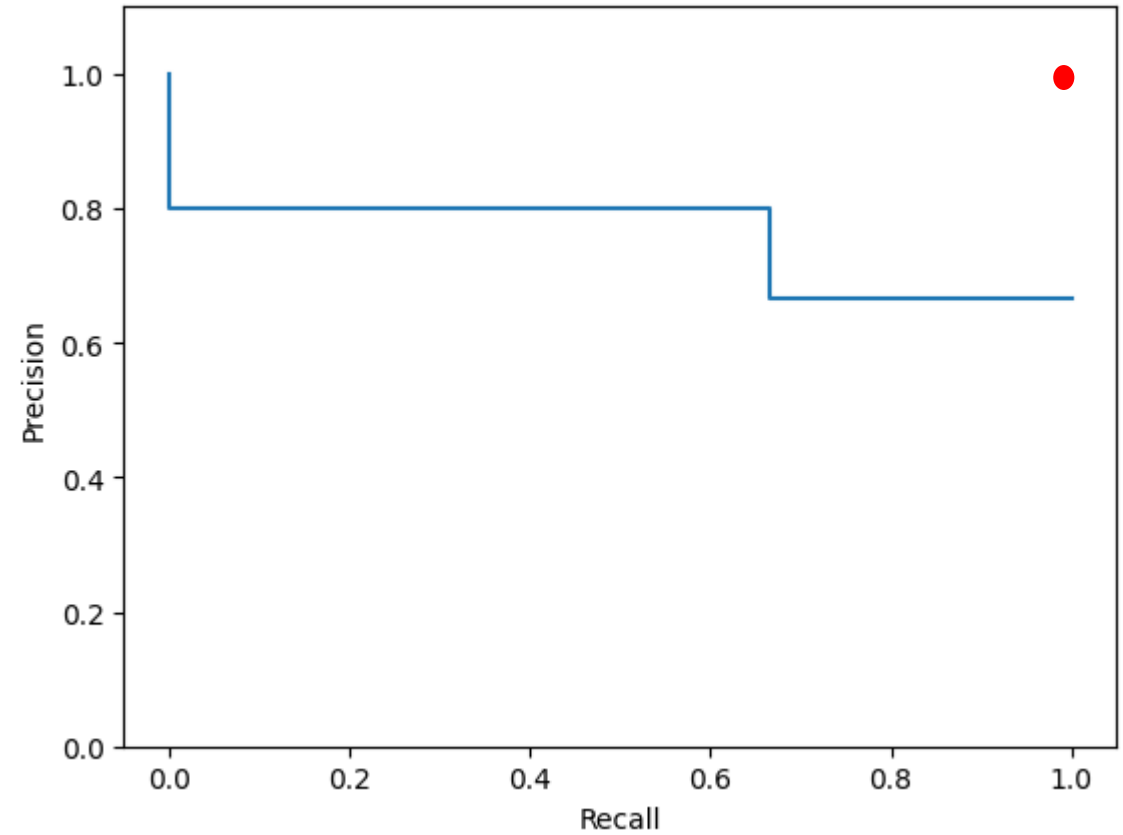
PR-кривая (PR-curve)

Левая точка: $(0,1)$

Правая точка: $(1,r)$

r -доля положительных объектов

Для идеального классификатора проходит через $(1,1)$



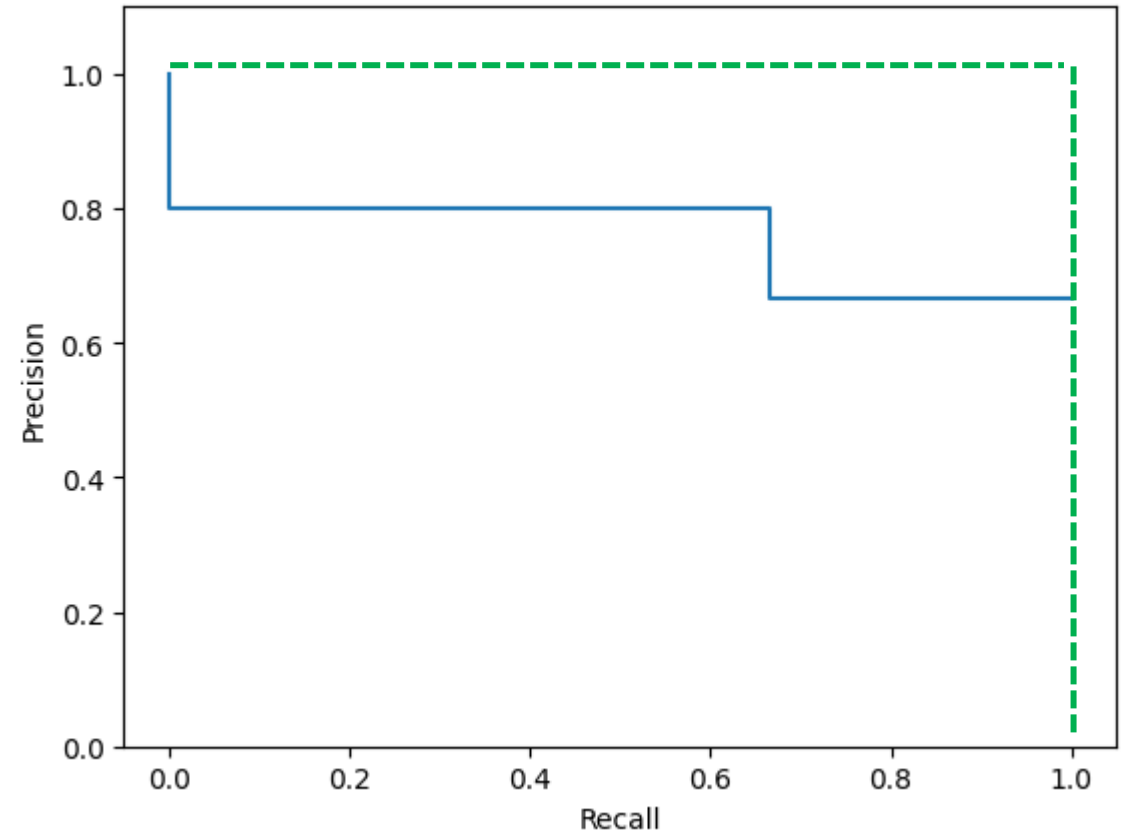
PR-кривая (PR-curve)

Левая точка: $(0,1)$

Правая точка: $(1,r)$

r -доля положительных объектов

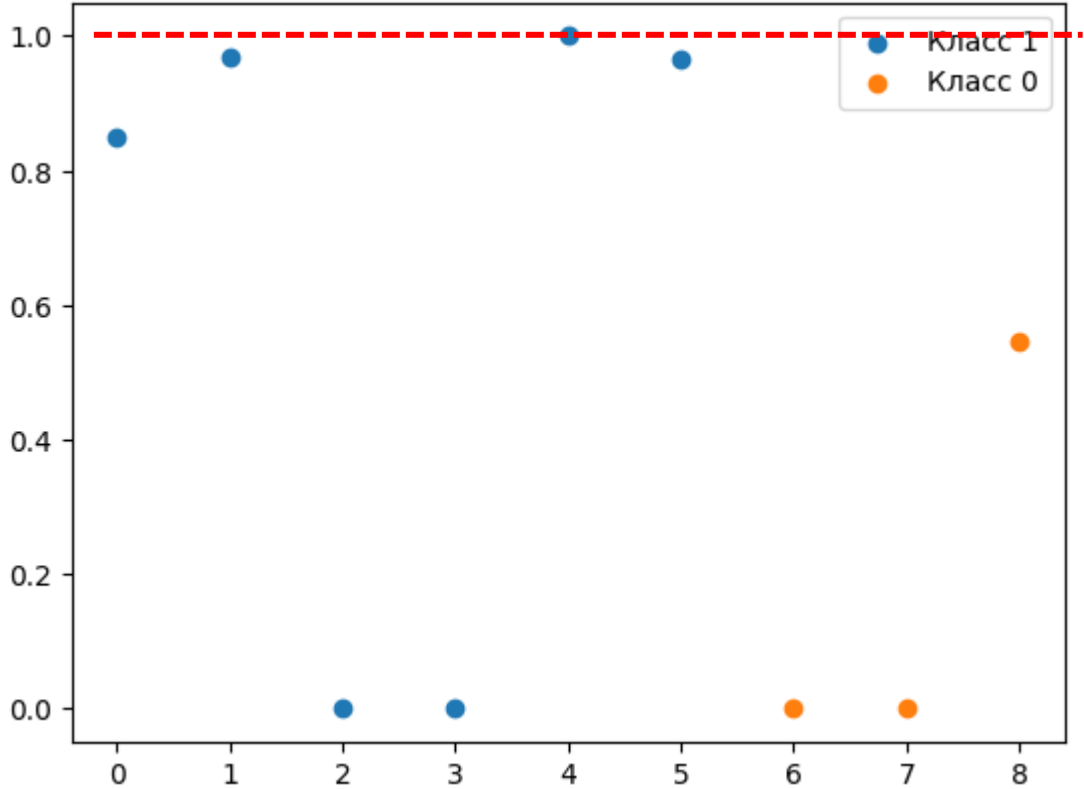
Для идеального классификатора проходит черерз $(1,1)$



PR-кривая (PR-curve)

Proba	Класс	
0.0	0	0
0.0	0	0
0.0	1	0
0.0	1	0
0.54545455	0	0
0.84905661	1	0
0.96629213	1	0
0.98675023	1	0
1.0	1	1

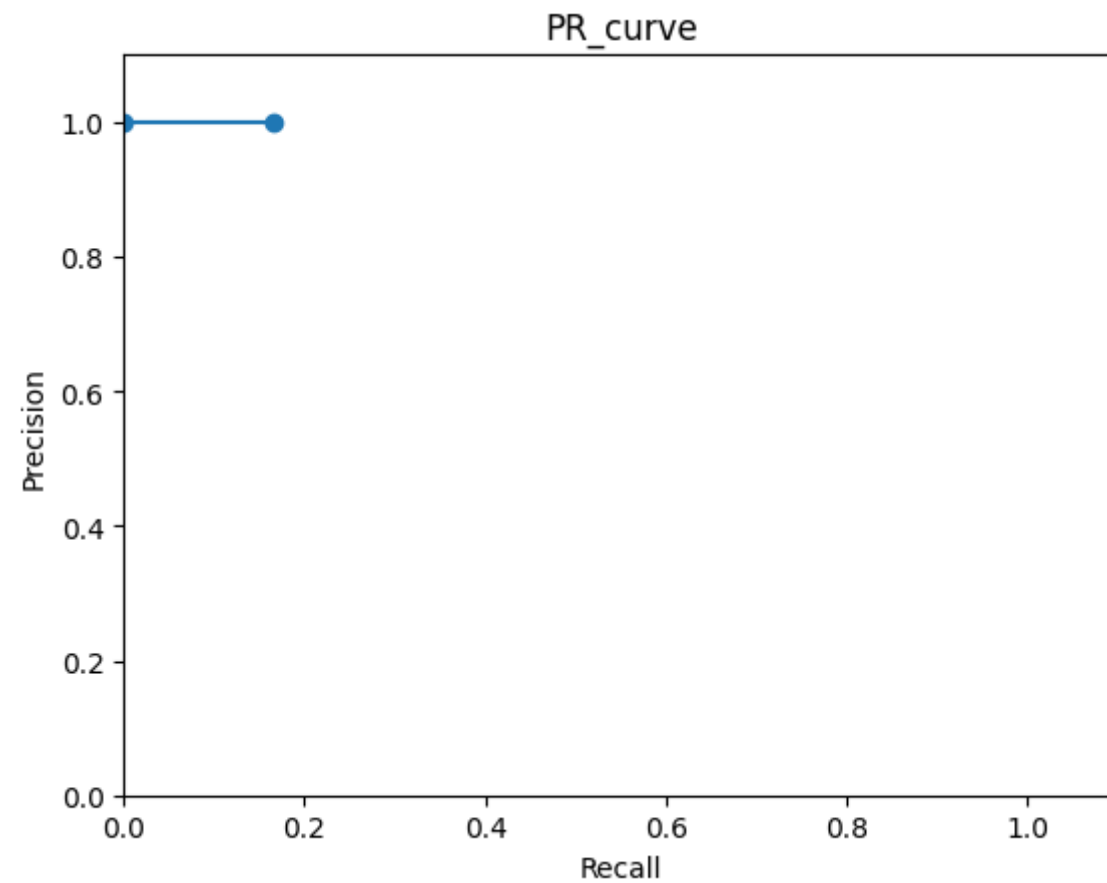
Порог = 1.0



$$\text{Precision} = \frac{TP}{TP+FP} = \frac{1}{1+0} = 1$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{1}{1+5} = 0.167$$

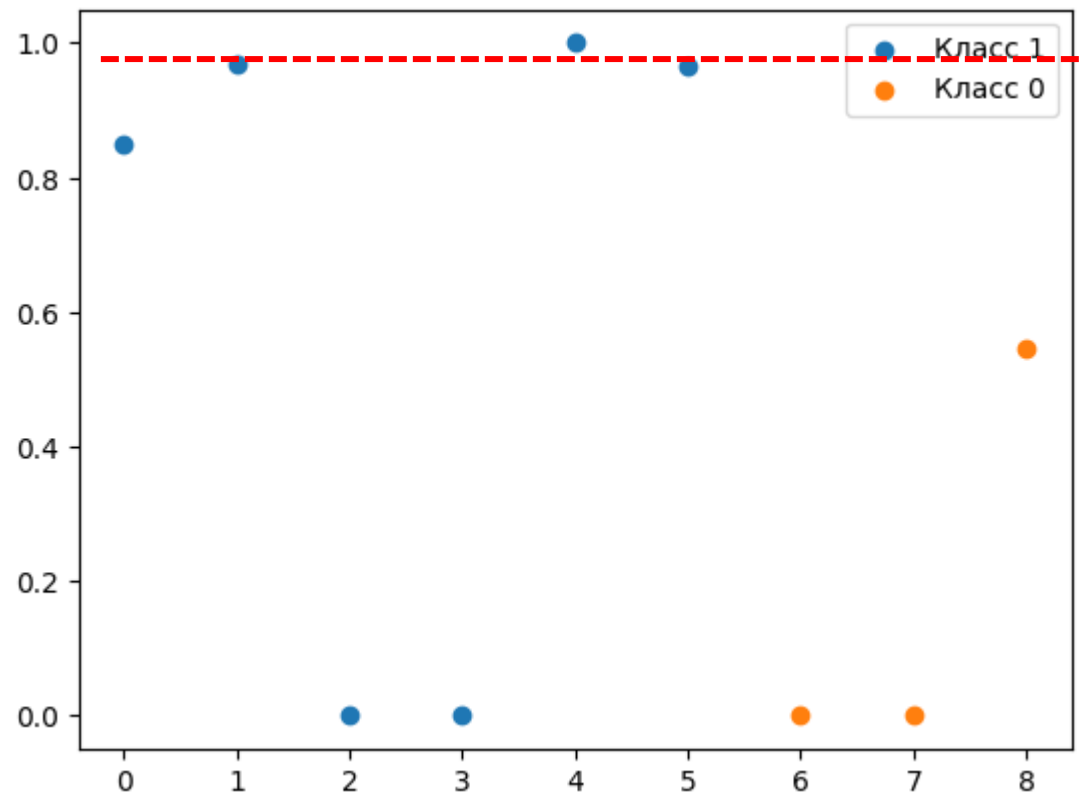
PR-кривая (PR-curve)



PR-кривая (PR-curve)

Proba	Класс	
0.0	0	0
0.0	0	0
0.0	1	0
0.0	1	0
0.54545455	0	0
0.84905661	1	0
0.96629213	1	0
0.98675023	1	1
1.0	1	1

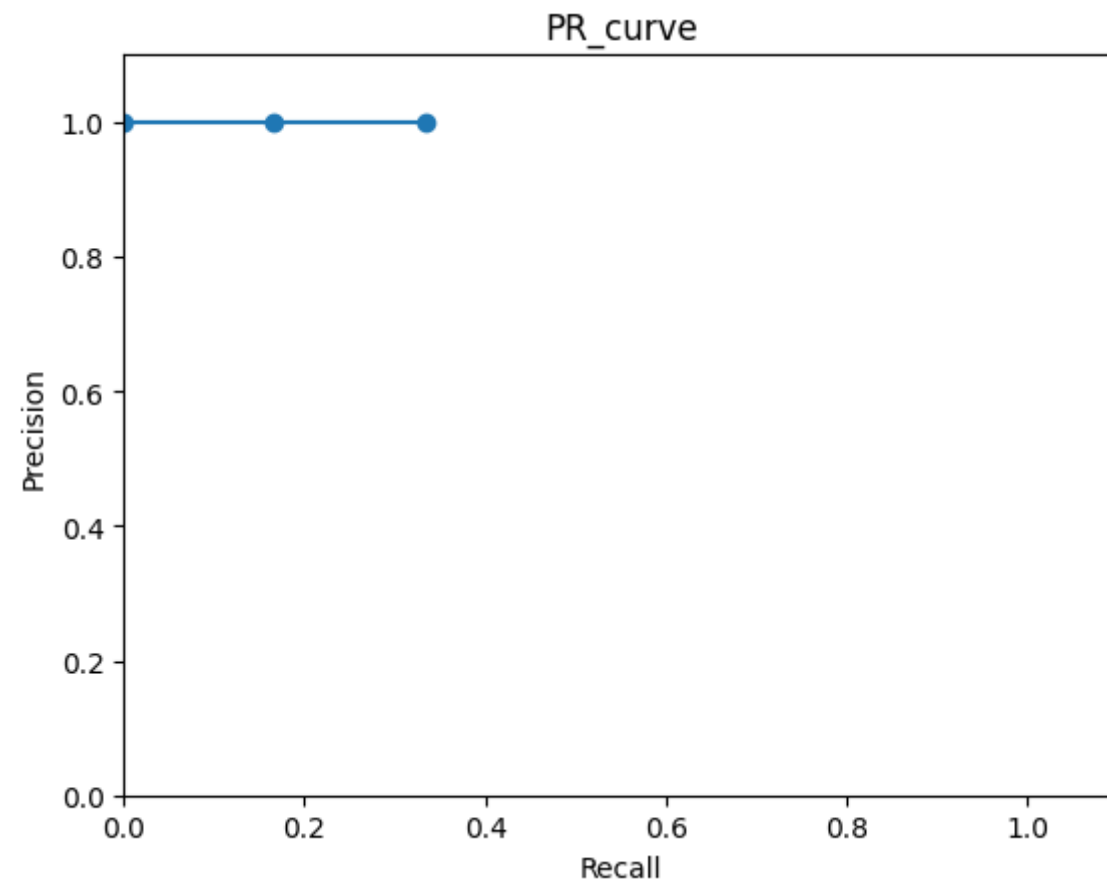
Порог = 0.98



$$\text{Precision} = \frac{TP}{TP+FP} = \frac{2}{2+0} = 1$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{2}{2+4} = 0.334$$

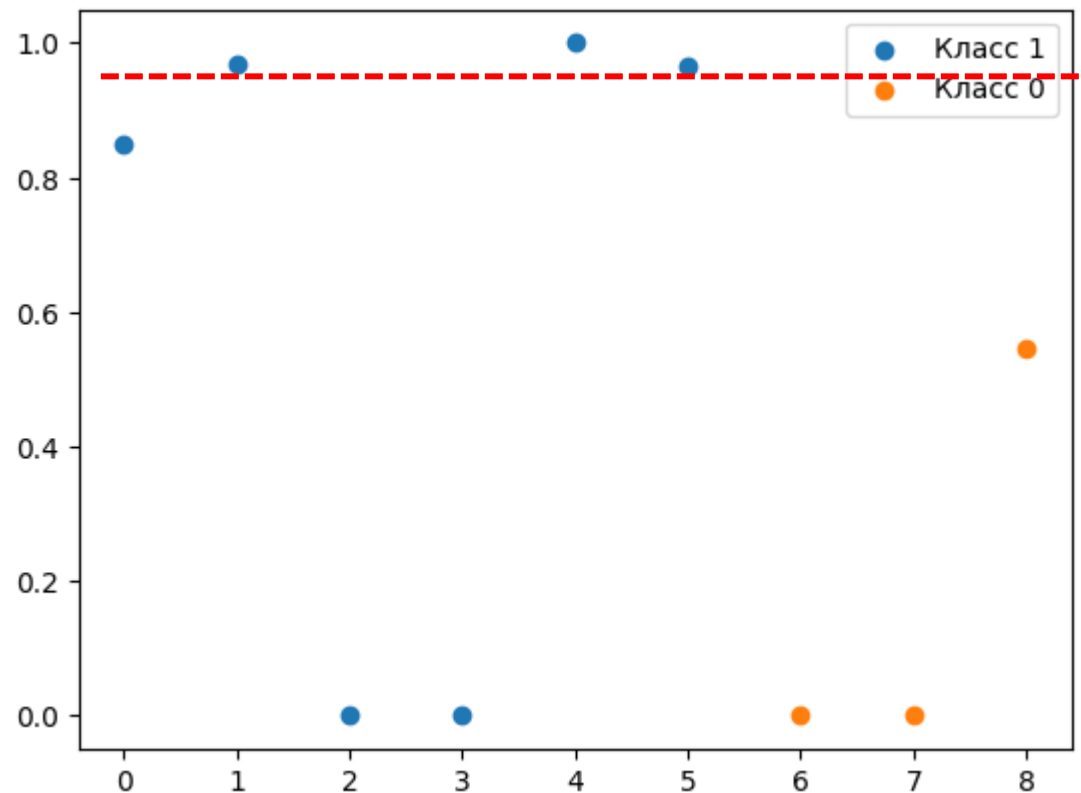
PR-кривая (PR-curve)



PR-кривая (PR-curve)

Proba	Класс	
0.0	0	0
0.0	0	0
0.0	1	0
0.0	1	0
0.54545455	0	0
0.84905661	1	0
0.96629213	1	1
0.98675023	1	1
1.0	1	1

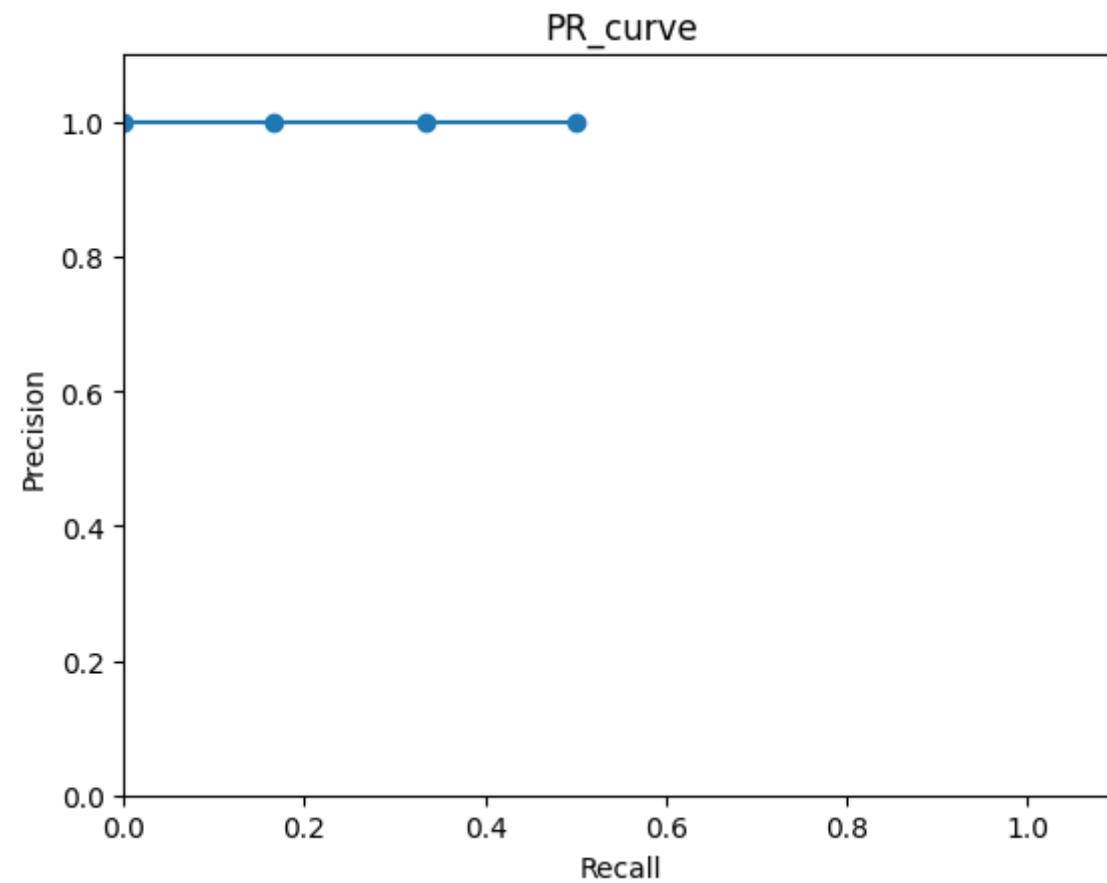
Порог = 0.96



$$\text{Precision} = \frac{TP}{TP+FP} = \frac{3}{3+0} = 1$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{3}{3+3} = 0.5$$

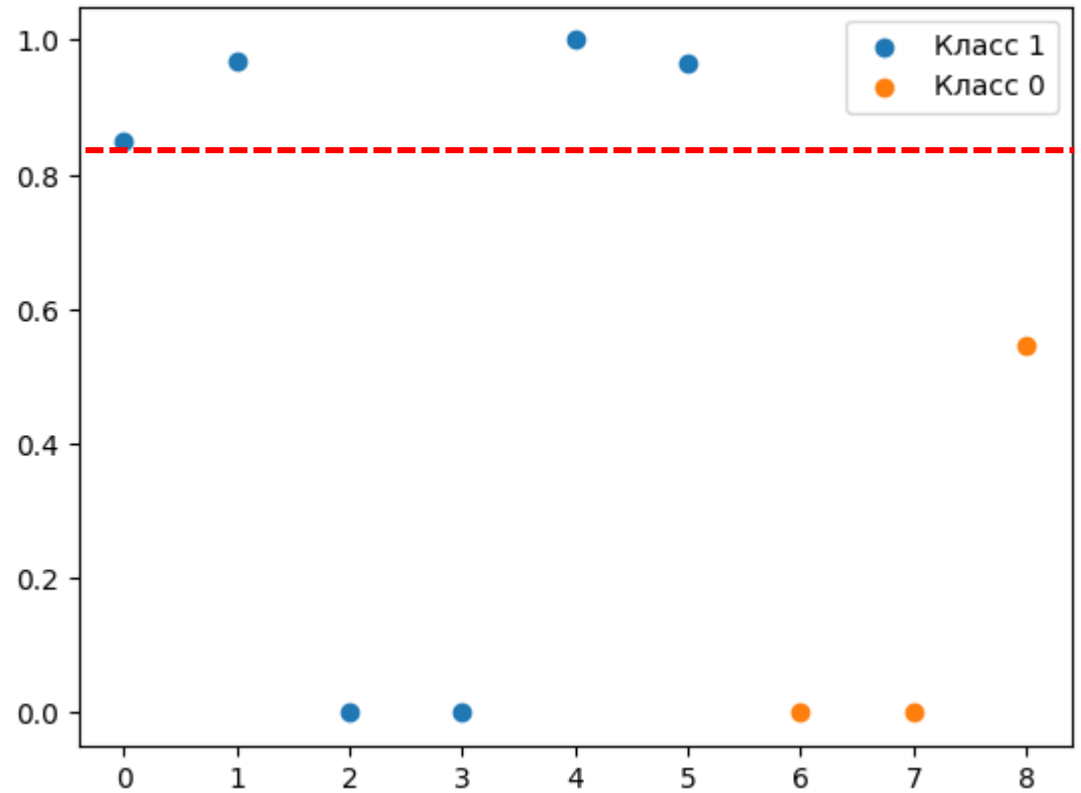
PR-кривая (PR-curve)



PR-кривая (PR-curve)

Proba	Класс	
0.0	0	0
0.0	0	0
0.0	1	0
0.0	1	0
0.54545455	0	0
0.84905661	1	1
0.96629213	1	1
0.98675023	1	1
1.0	1	1

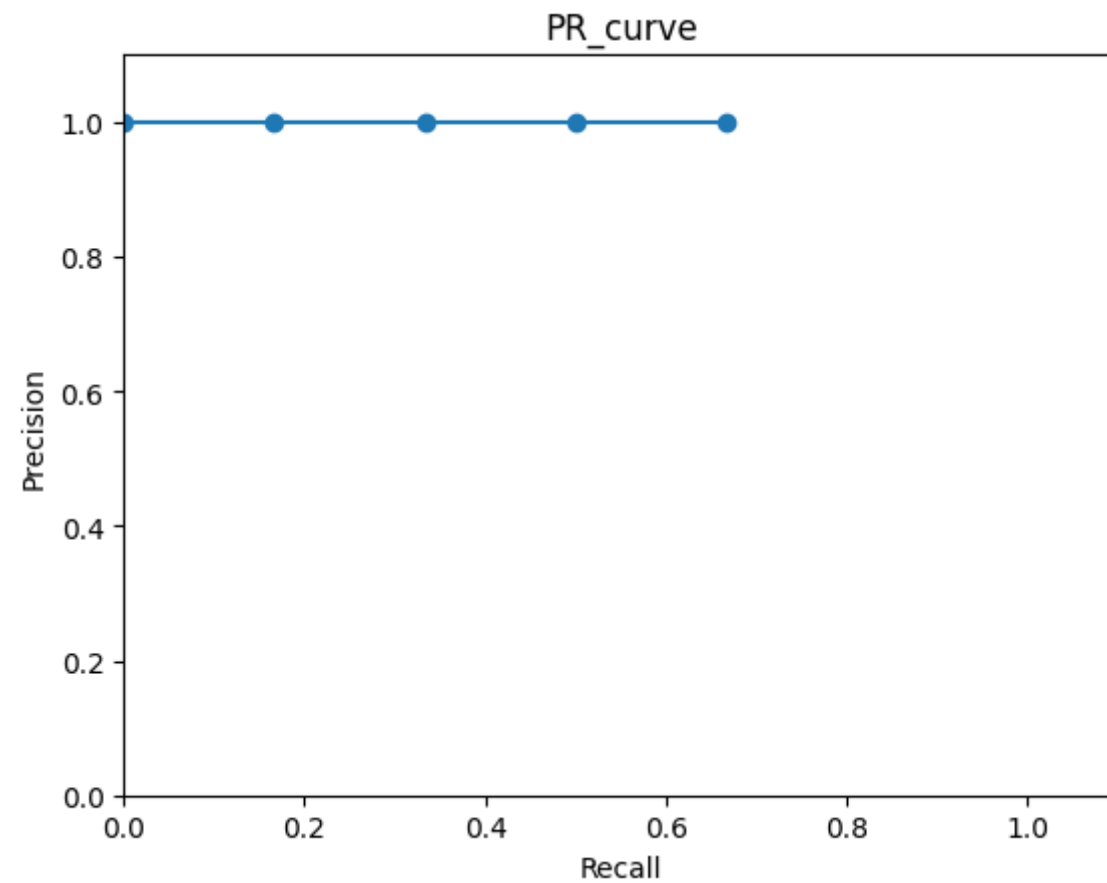
Порог = 0.84



$$\text{Precision} = \frac{TP}{TP+FP} = \frac{4}{4+0} = 1$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{4}{4+2} = 0.667$$

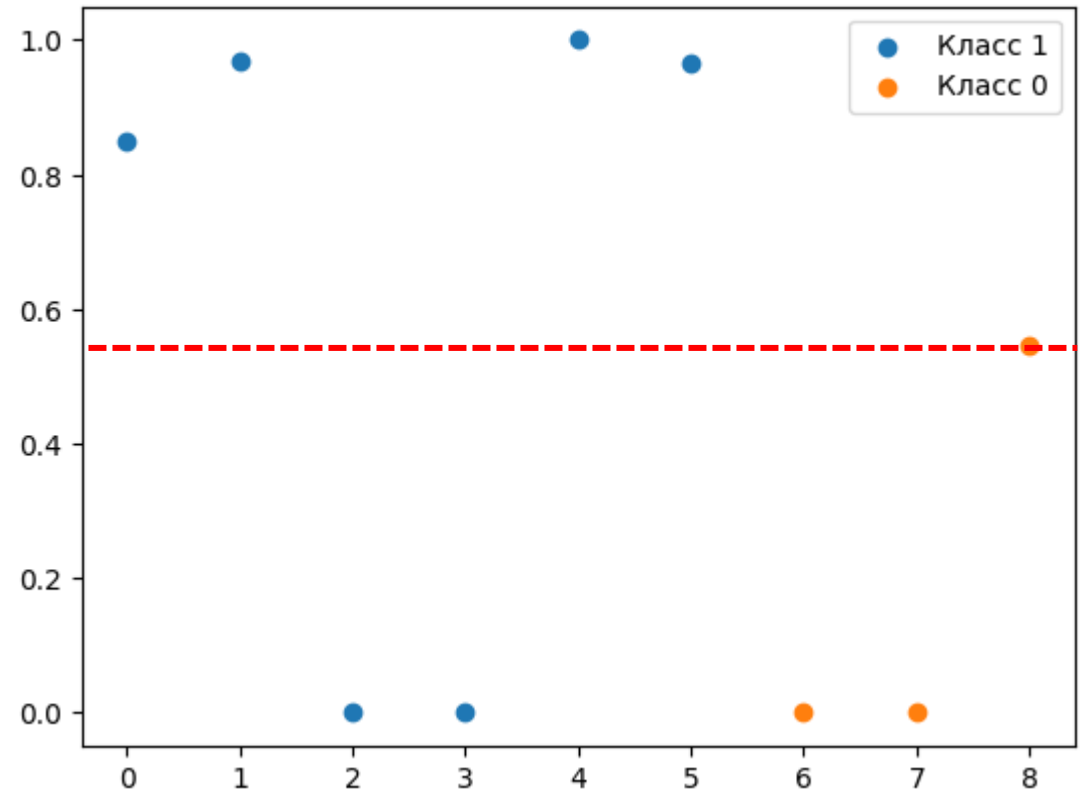
PR-кривая (PR-curve)



PR-кривая (PR-curve)

Proba	Класс	
0.0	0	0
0.0	0	0
0.0	1	0
0.0	1	0
0.54545455	0	1
0.84905661	1	1
0.96629213	1	1
0.98675023	1	1
1.0	1	1

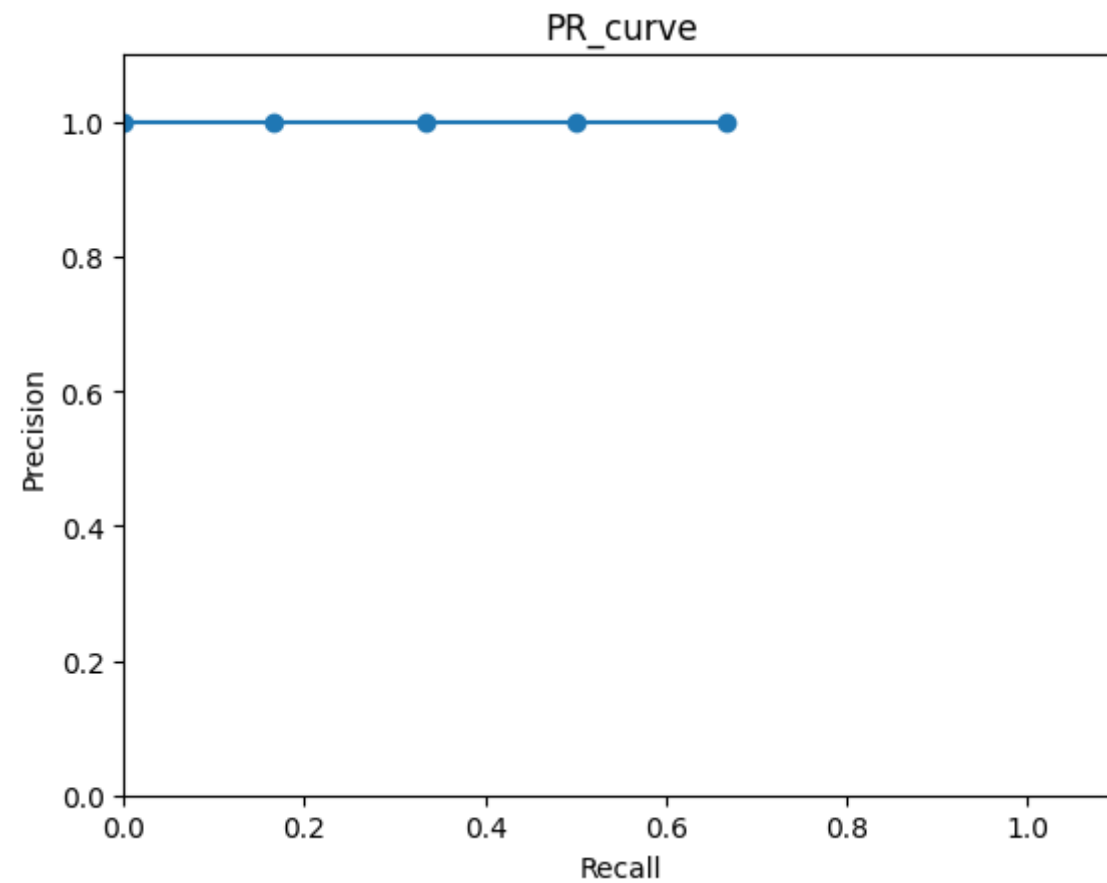
Порог = 0.54



$$\text{Precision} = \frac{TP}{TP+FP} = \frac{4}{4+1} = 0.8$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{4}{4+2} = 0.667$$

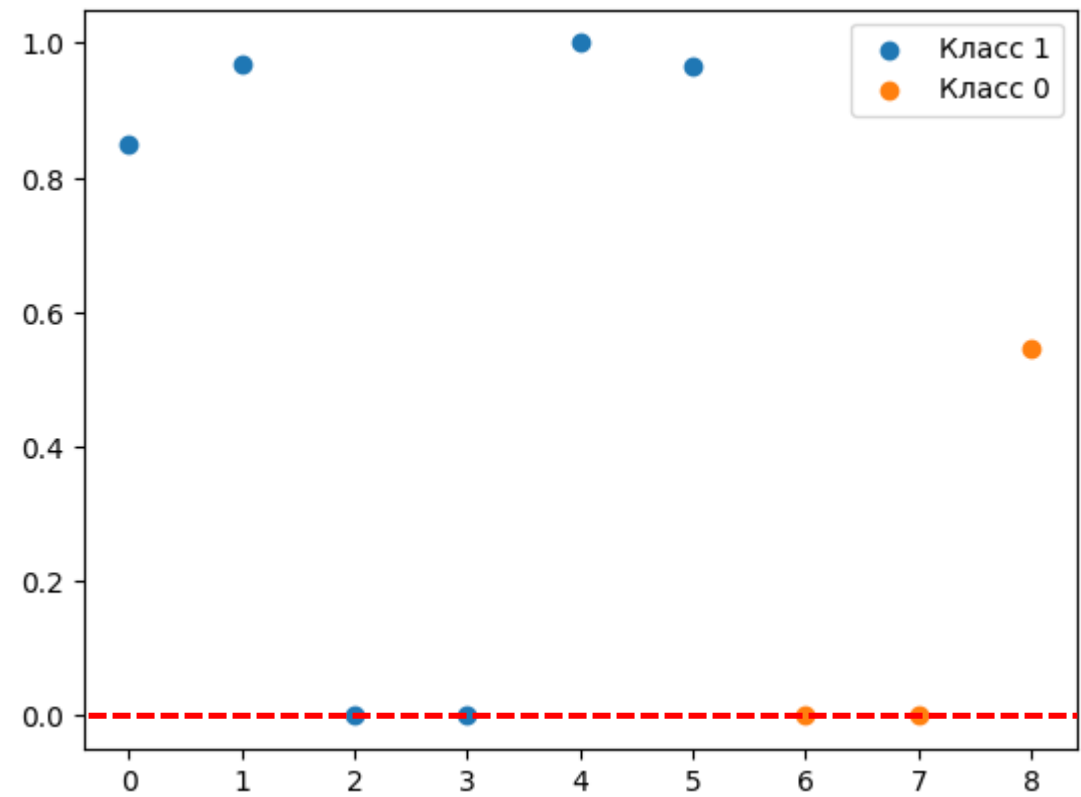
PR-кривая (PR-curve)



PR-кривая (PR-curve)

Proba	Класс	
0.0	0	1
0.0	0	1
0.0	1	1
0.0	1	1
0.54545455	0	1
0.84905661	1	1
0.96629213	1	1
0.98675023	1	1
1.0	1	1

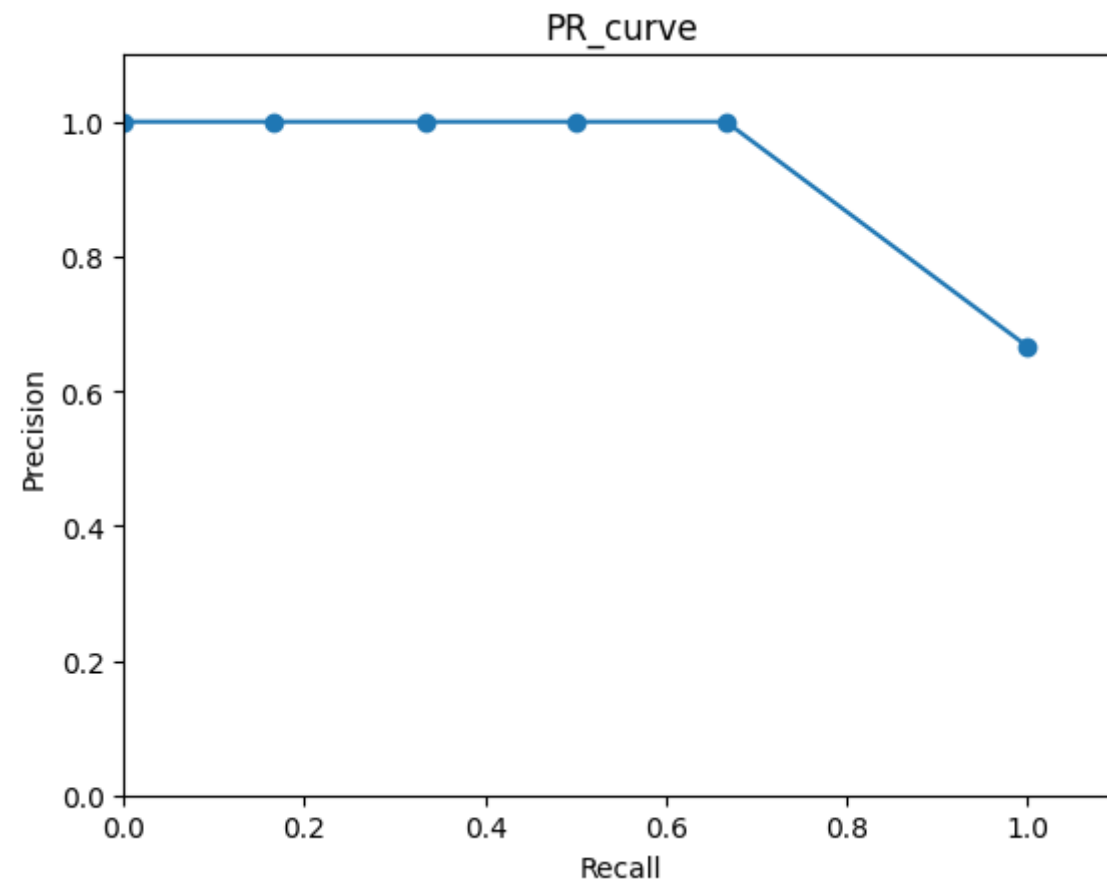
Порог = 0.0



$$\text{Precision} = \frac{TP}{TP+FP} = \frac{6}{6+3} = 0.667$$

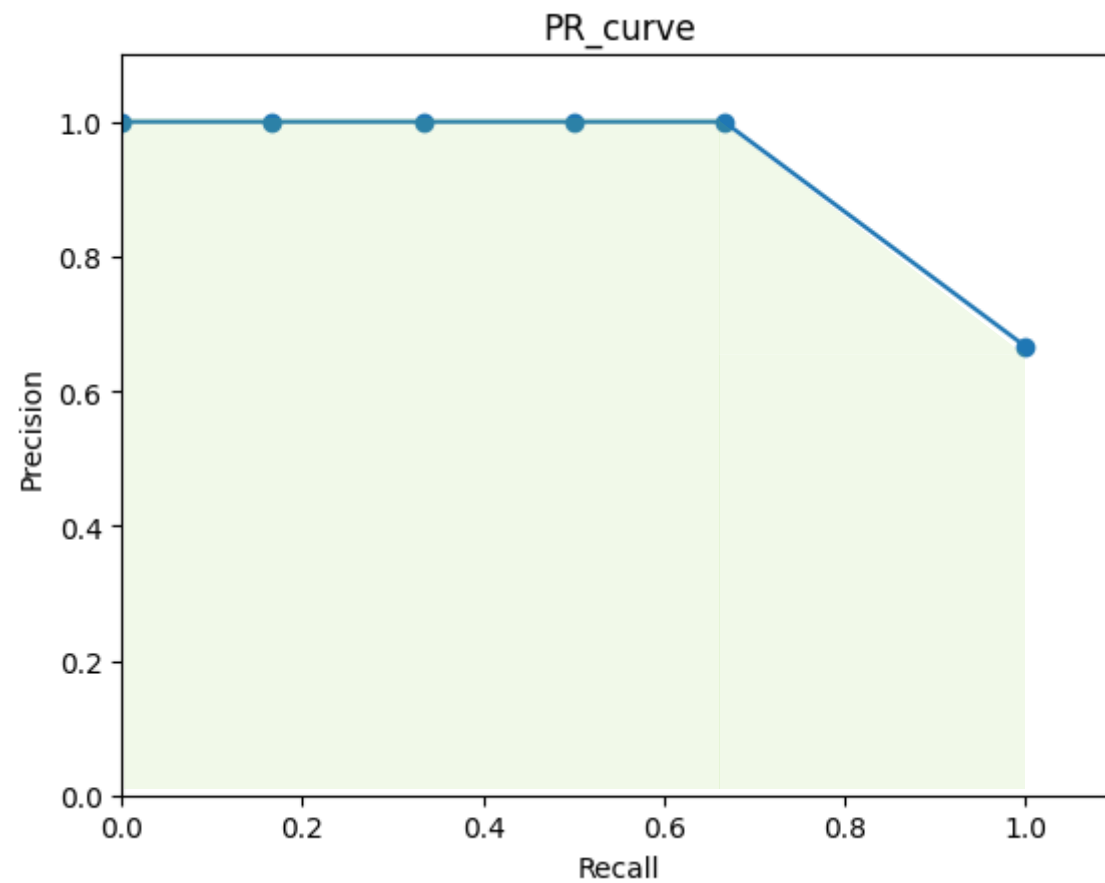
$$\text{Recall} = \frac{TP}{TP+FN} = \frac{6}{6+0} = 1$$

PR-кривая (PR-curve)



PR-AUC

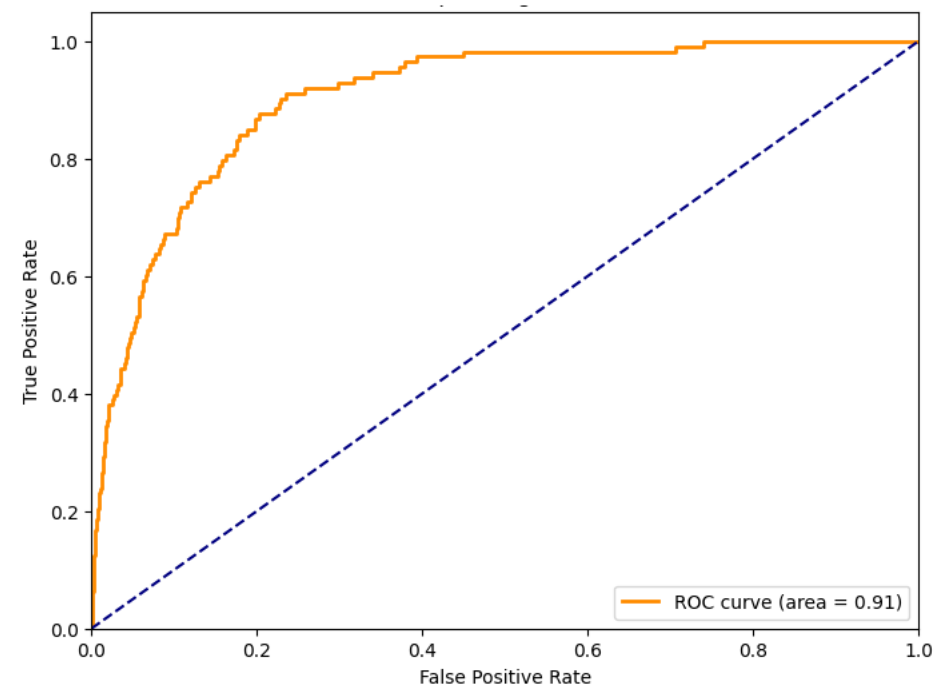
PR-AUC площадь под PR-кривую



PR-AUC=0.98

ROC-кривая (Receiver Operating Characteristic)

ROC-кривая (Receiver Operating Characteristic)

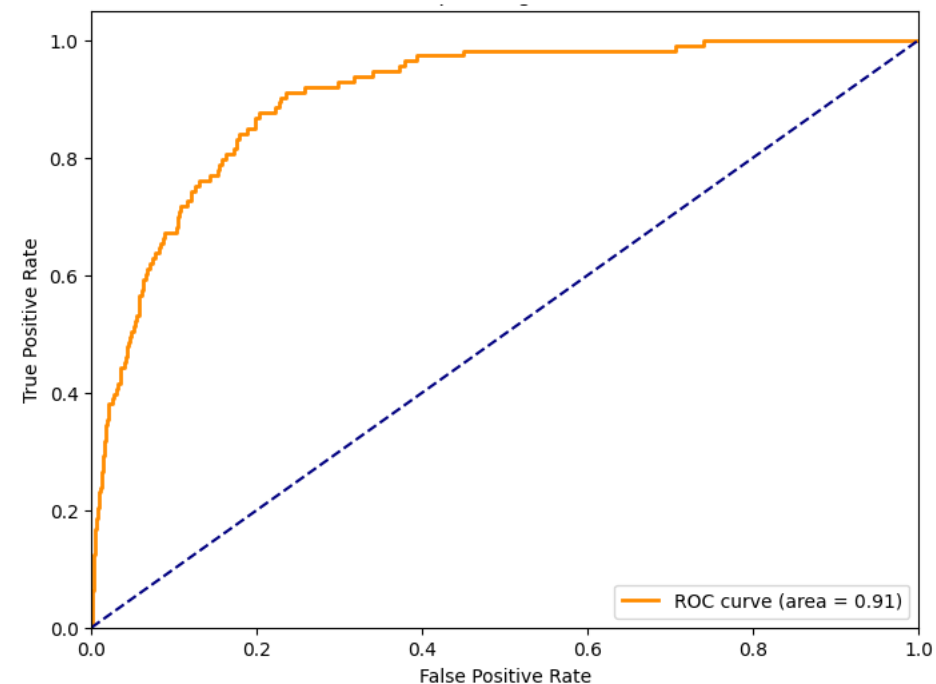


ROC-кривая (Receiver Operating Characteristic)

Ось X – False Positive Rate

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

доля неверно принятых объектов отрицательного класса



ROC-кривая (Receiver Operating Characteristic)

Ось X – False Positive Rate

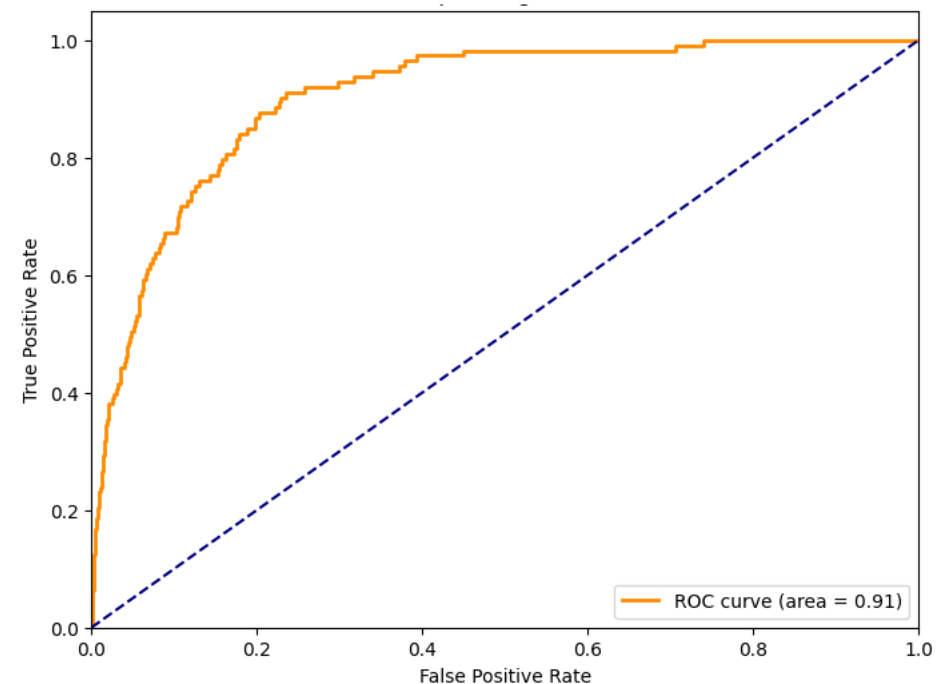
$$FPR = \frac{TP}{FP+TN}$$

доля неверно принятых объектов отрицательного класса

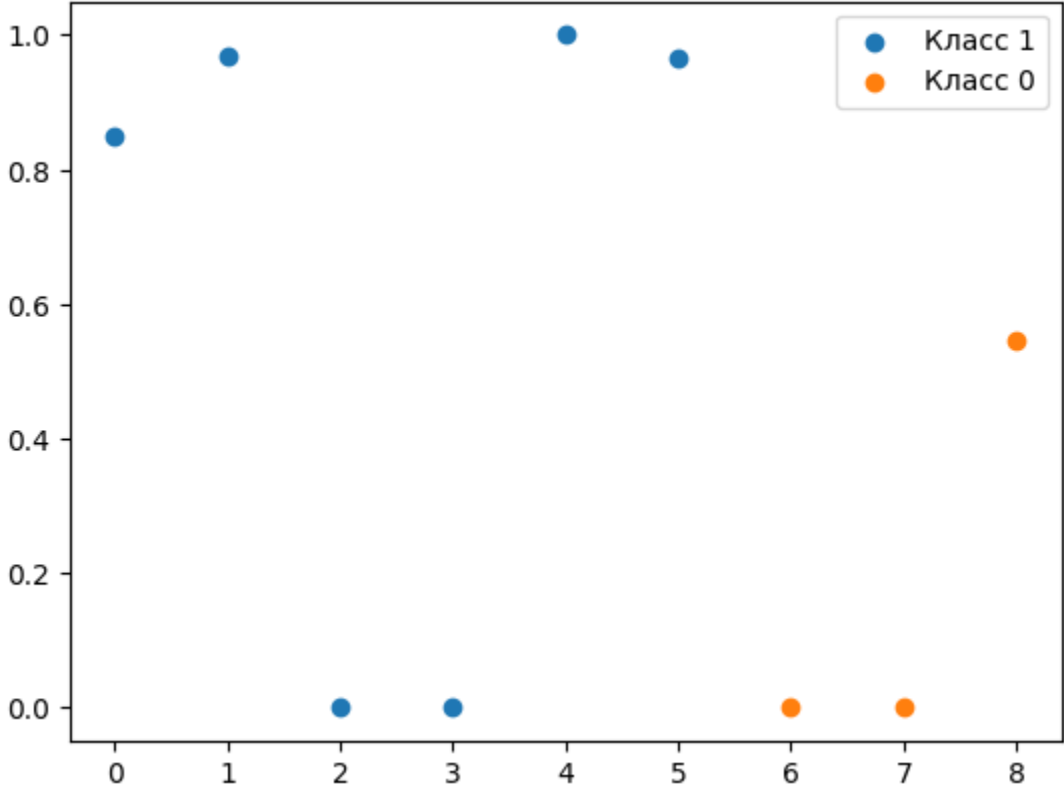
Ось Y – True Positive Rate

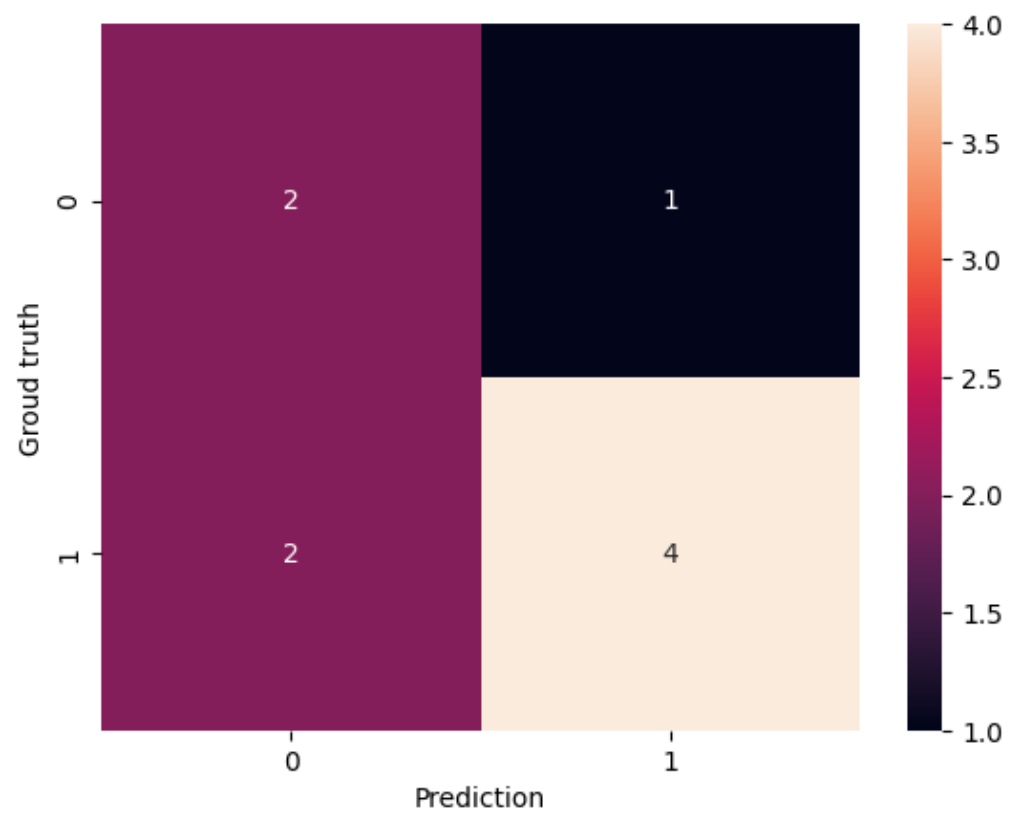
$$TPR = \frac{TP}{TP+FN}$$

доля верно принятых объектов положительного класса



Класс 0	Класс 1
0.1509434	0.8490566
0.03125	0.96875
1.	0.
1.	0.
1.	0.
1	0
0.	1.
0.03370787	0.96629213
0.45454545	0.54545455

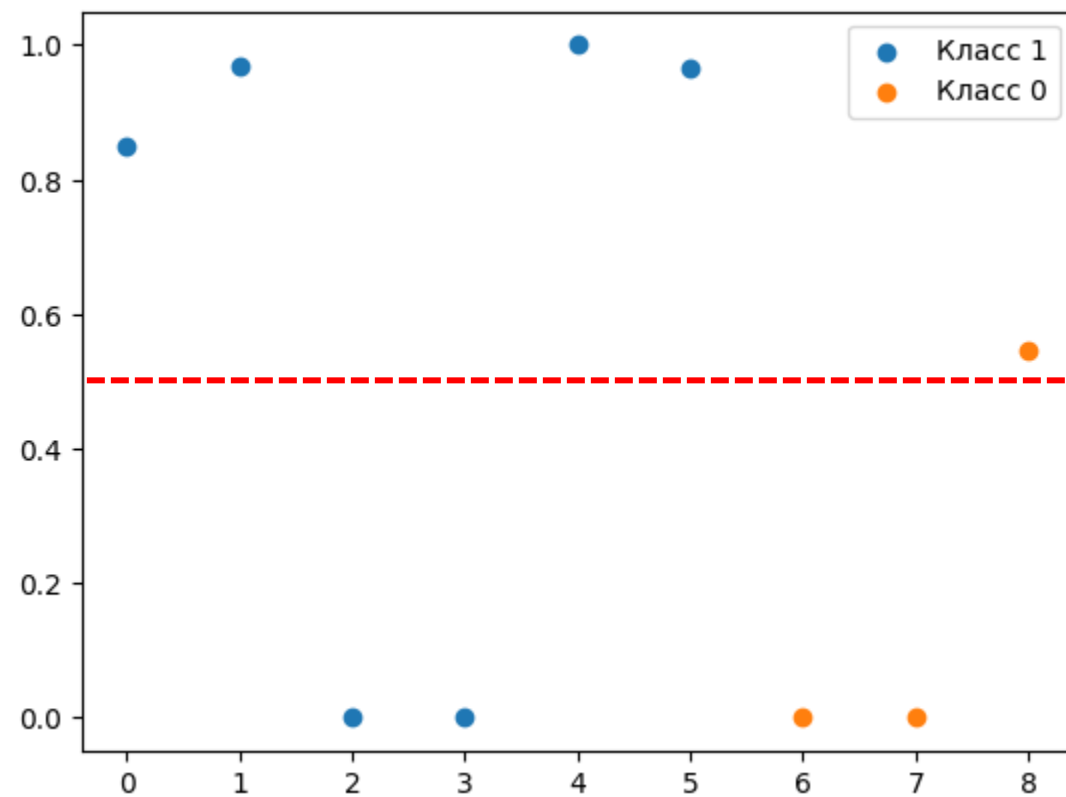


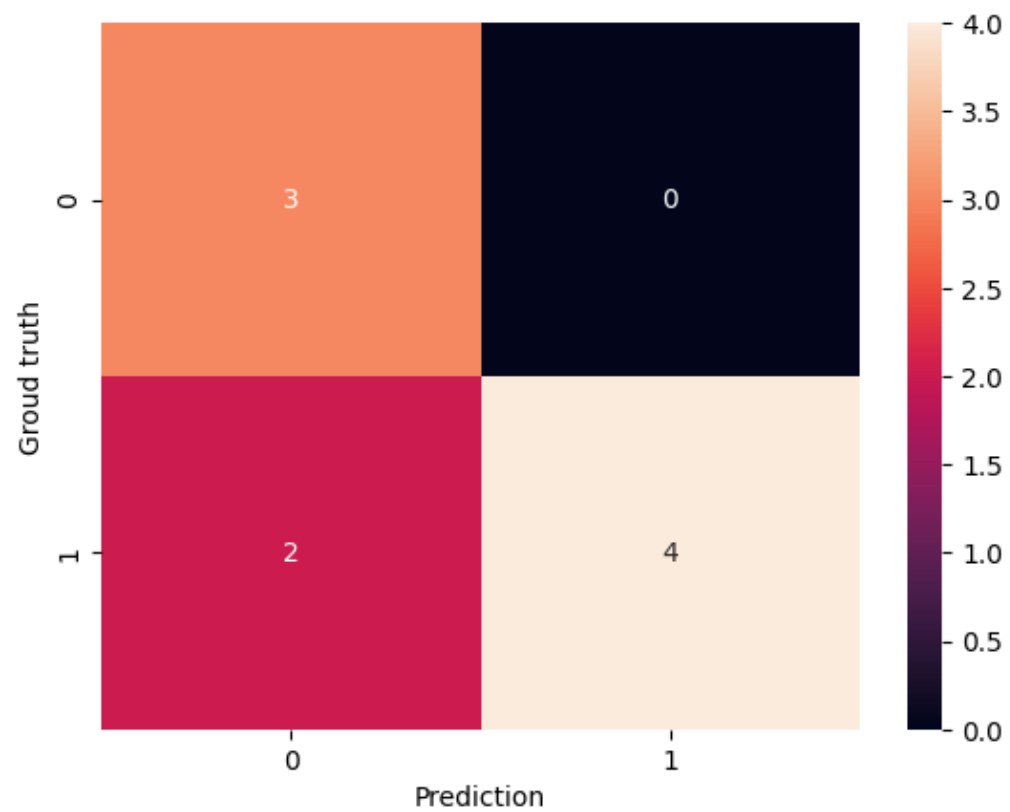


Порог = 0.5

TPR = 0.667

FPR = 0.334

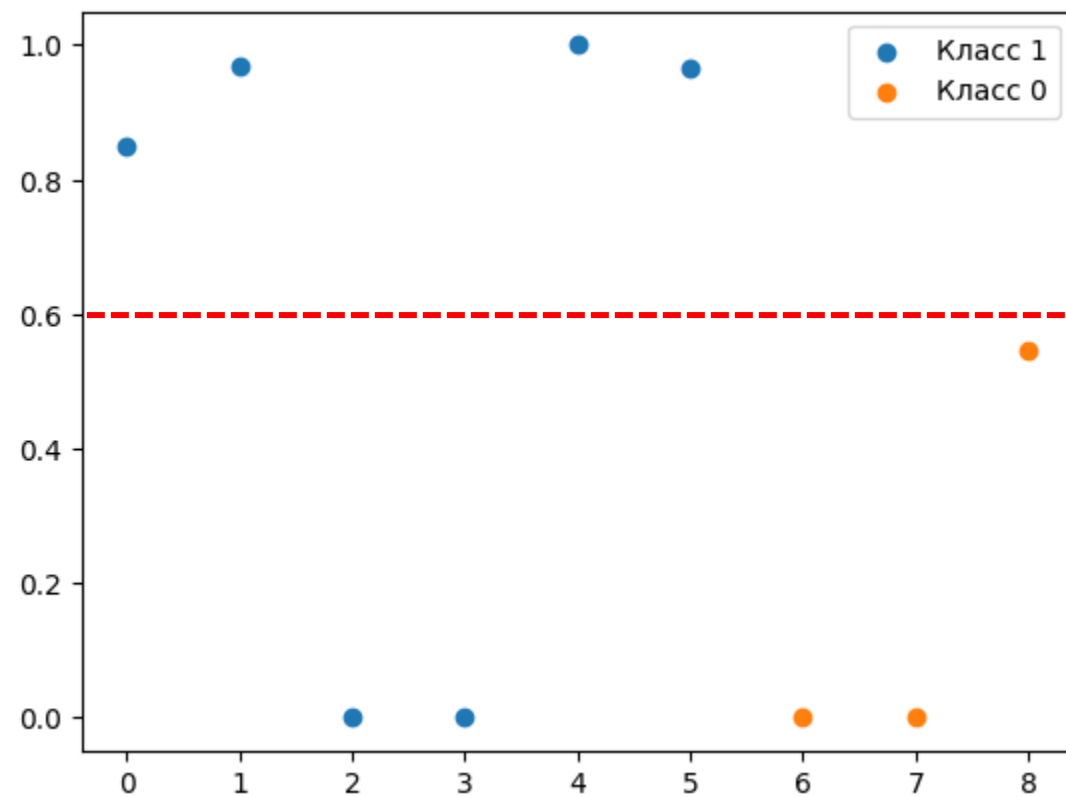


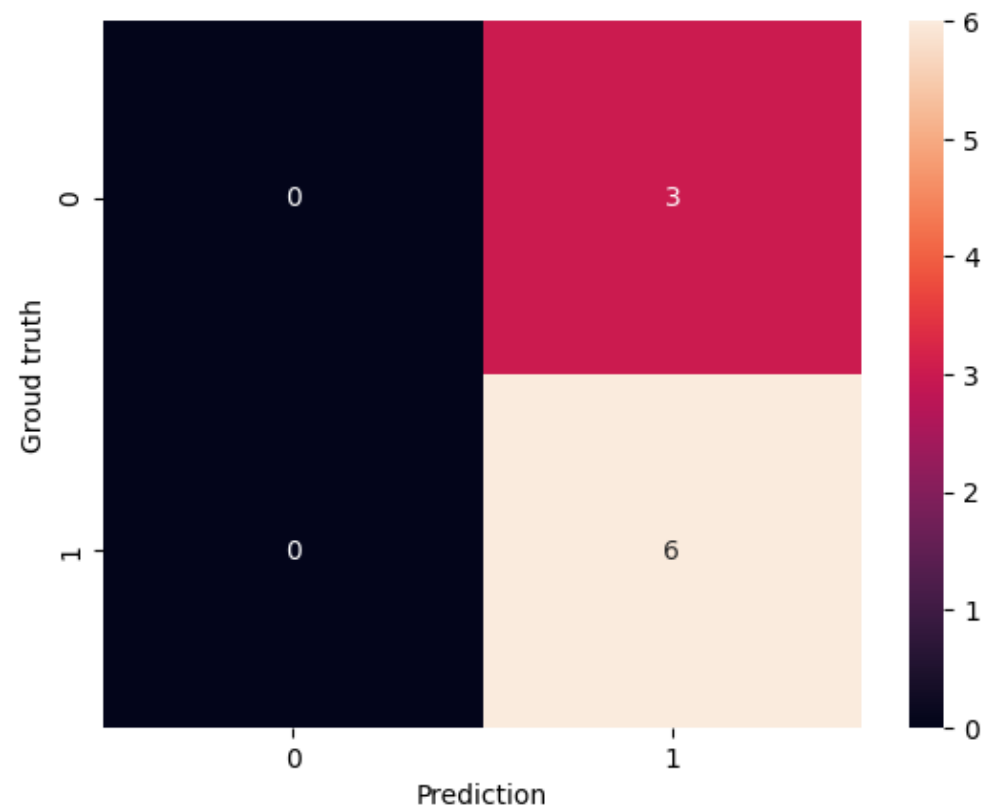


Порог = 0.6

TPR = 0.666

FPR = 0.0

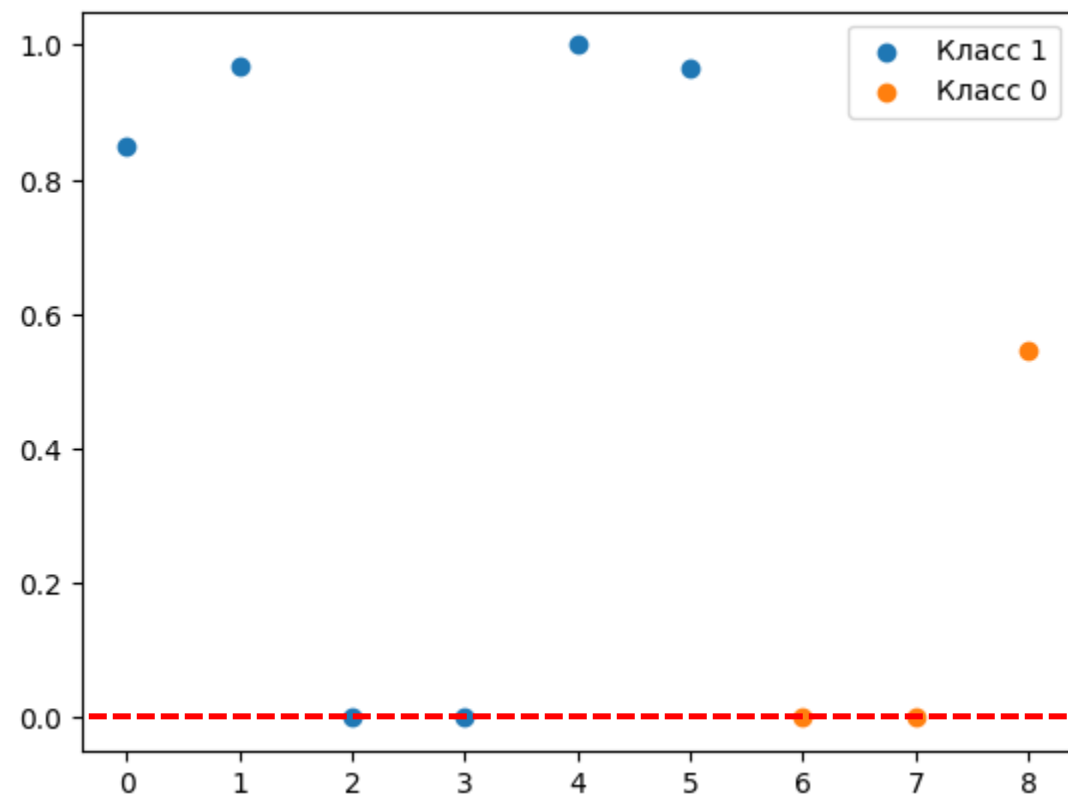




Порог = 0.0

TPR = 1

FPR = 1

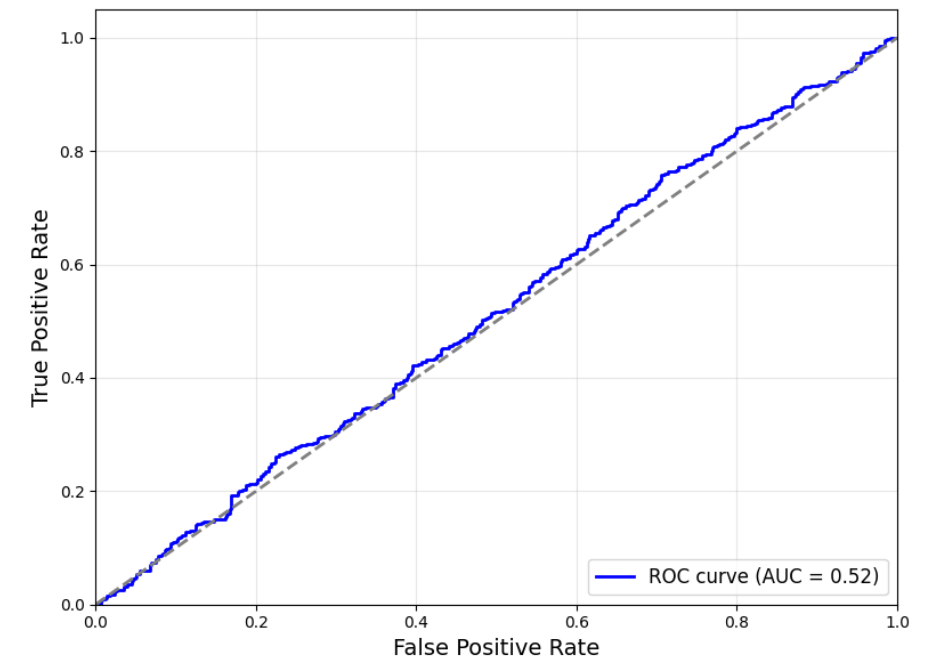
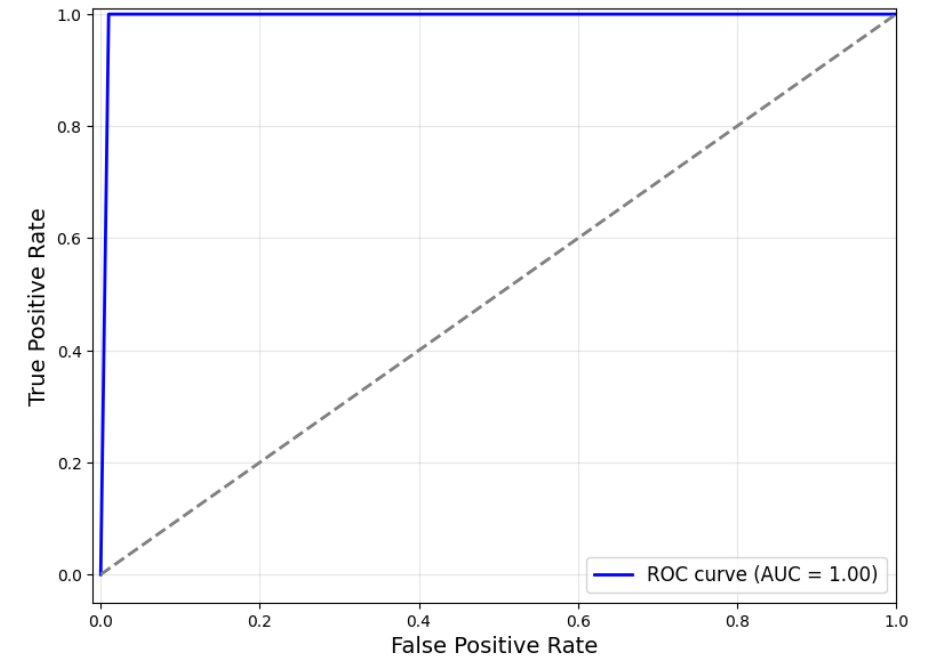


ROC-кривая

Левая точка: (0,0)

Правая точка: (1,1)

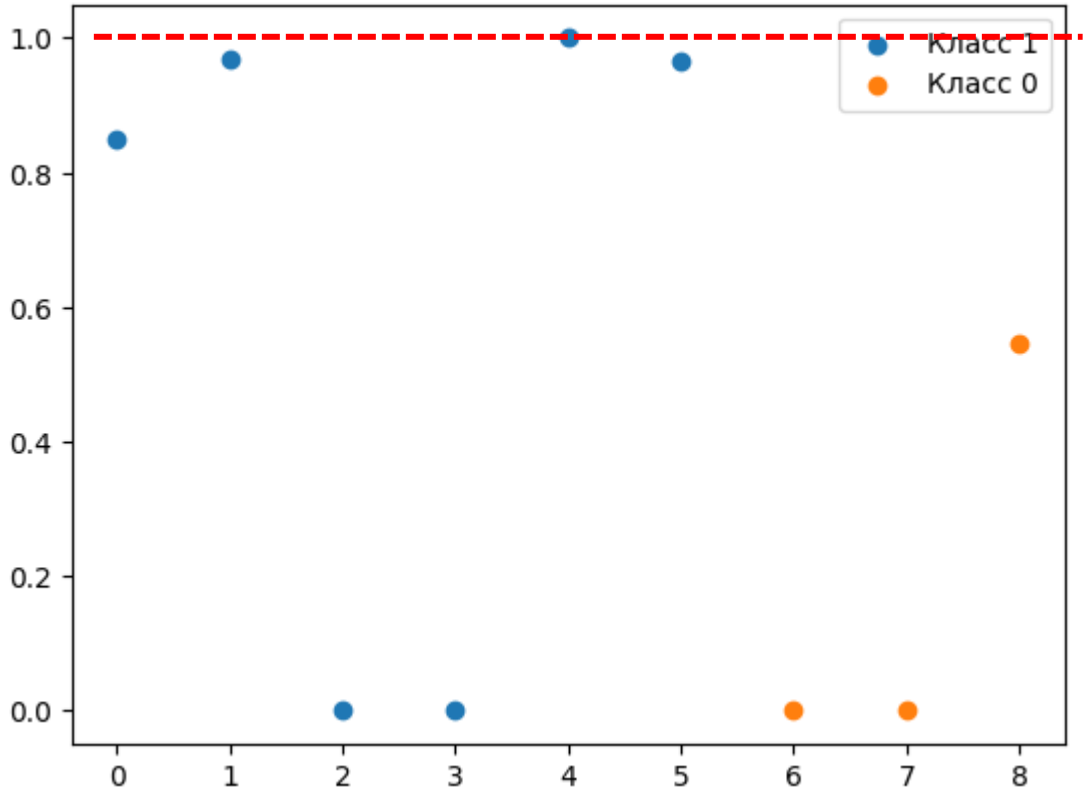
Для идеального классификатора проходит через (0,1)



ROC-кривая

Proba	Класс	
0.0	0	0
0.0	0	0
0.0	1	0
0.0	1	0
0.54545455	0	0
0.84905661	1	0
0.96629213	1	0
0.98675023	1	0
1.0	1	1

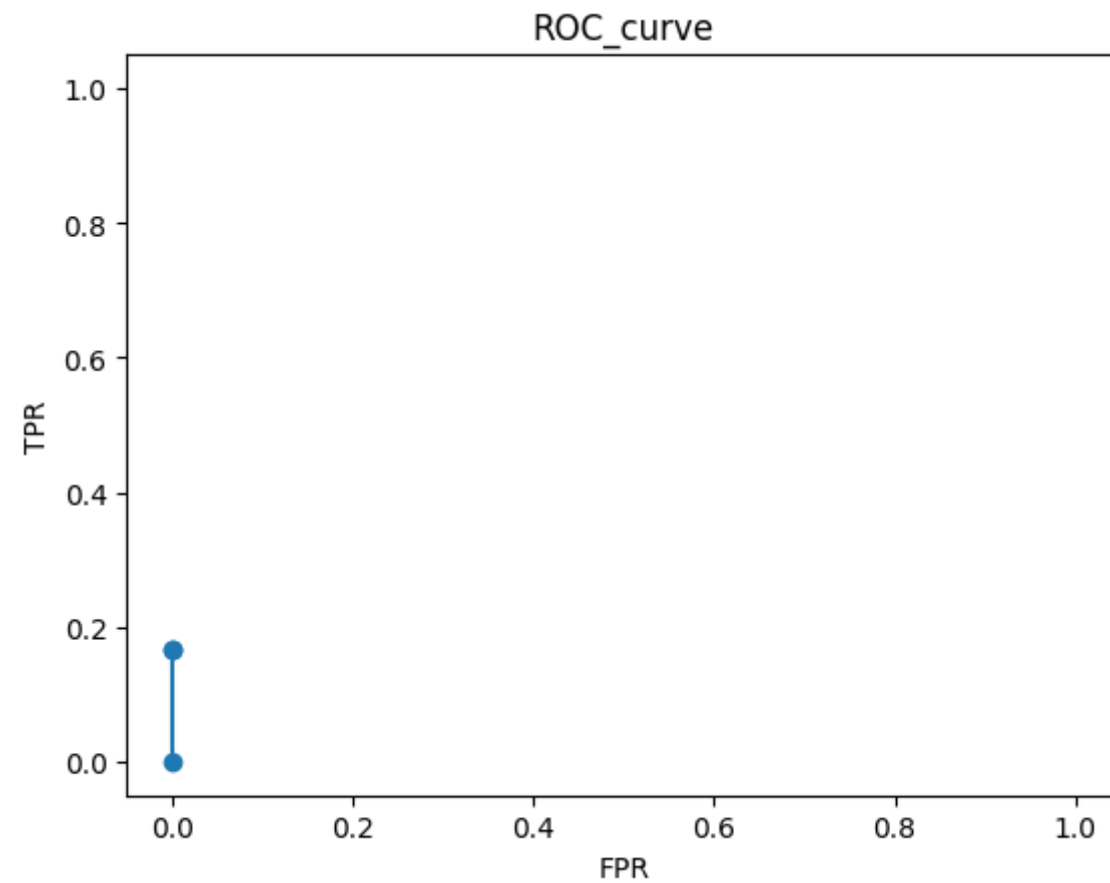
Порог = 1.0



$$TPR = \frac{TP}{TP+FN} = \frac{1}{1+5} = 0.167$$

$$FPR = \frac{FP}{FP+TN} = \frac{0}{0+3} = 0$$

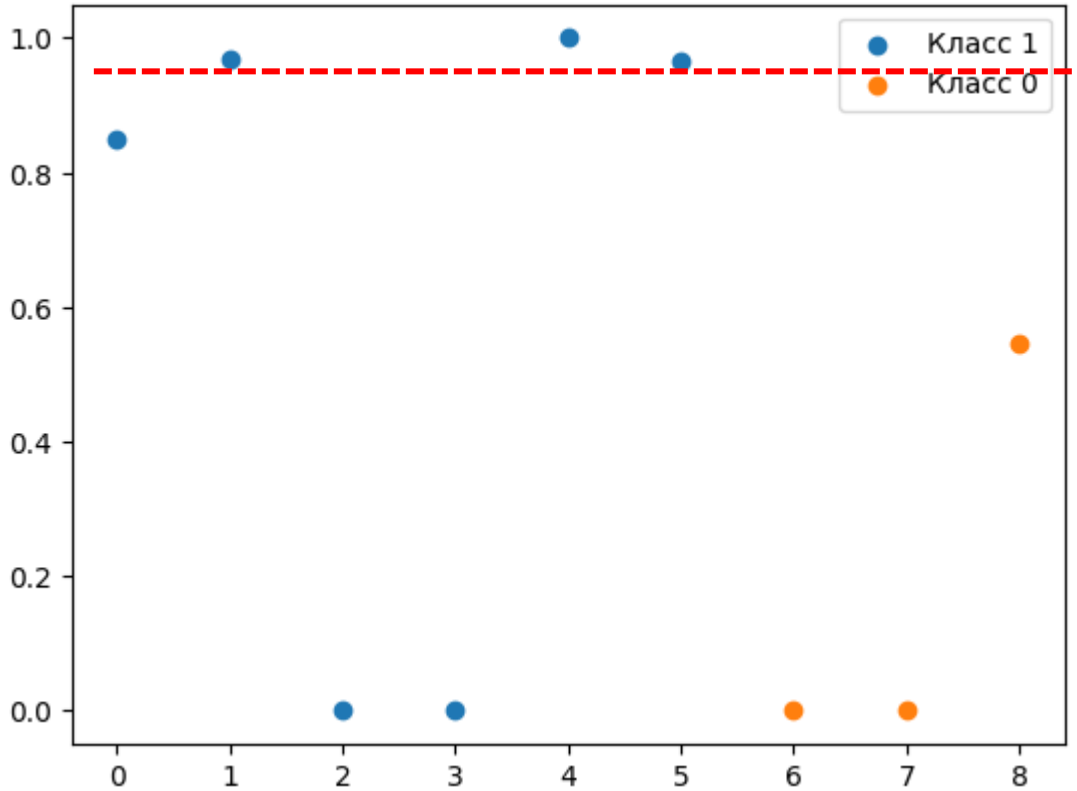
ROC-кривая



ROC-кривая

Proba	Класс	
0.0	0	0
0.0	0	0
0.0	1	0
0.0	1	0
0.54545455	0	0
0.84905661	1	0
0.96629213	1	1
0.98675023	1	1
1.0	1	1

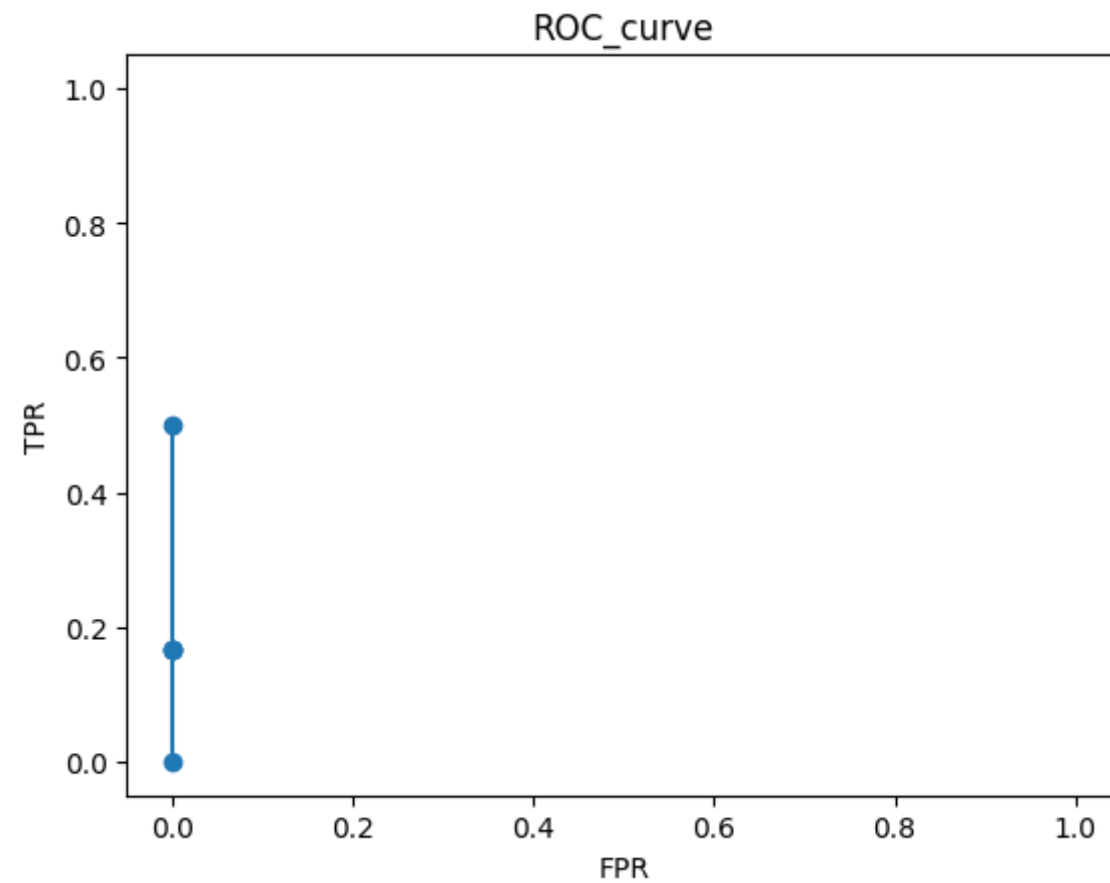
Порог = 0.96



$$TPR = \frac{TP}{TP+FN} = 0.5$$

$$FPR = \frac{FP}{FP+TN} = 0$$

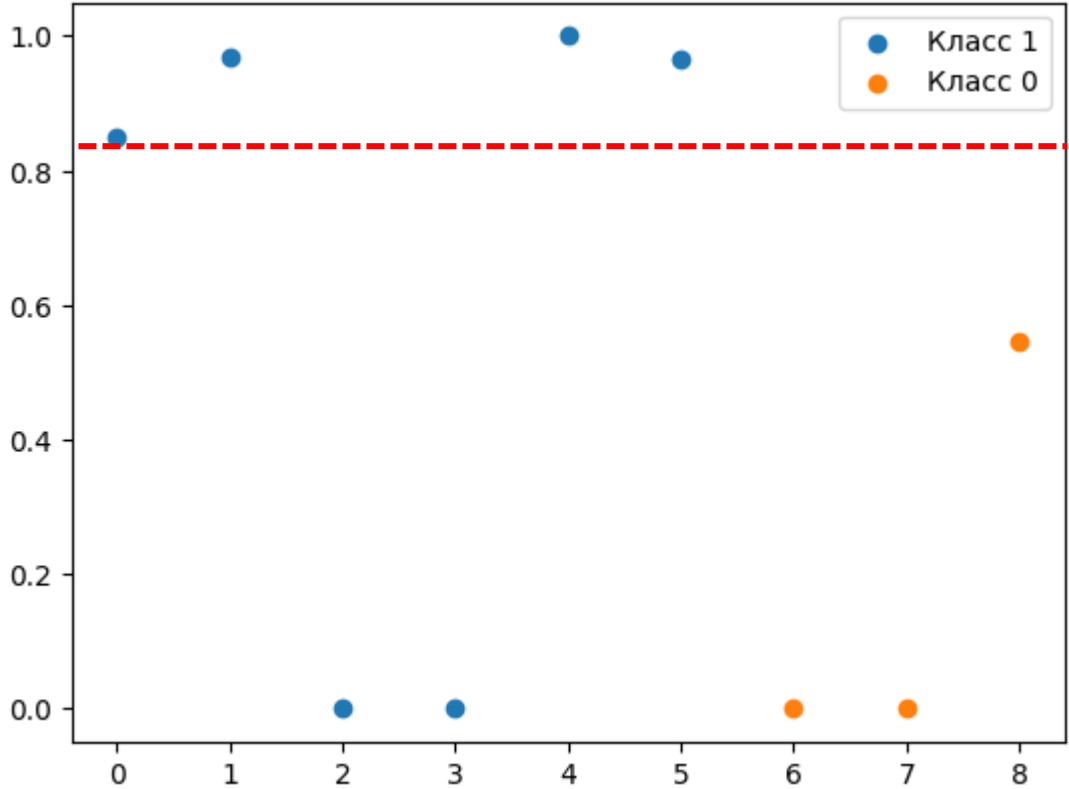
ROC-кривая



ROC-кривая

Proba	Класс	
0.0	0	0
0.0	0	0
0.0	1	0
0.0	1	0
0.54545455	0	0
0.84905661	1	1
0.96629213	1	1
0.98675023	1	1
1.0	1	1

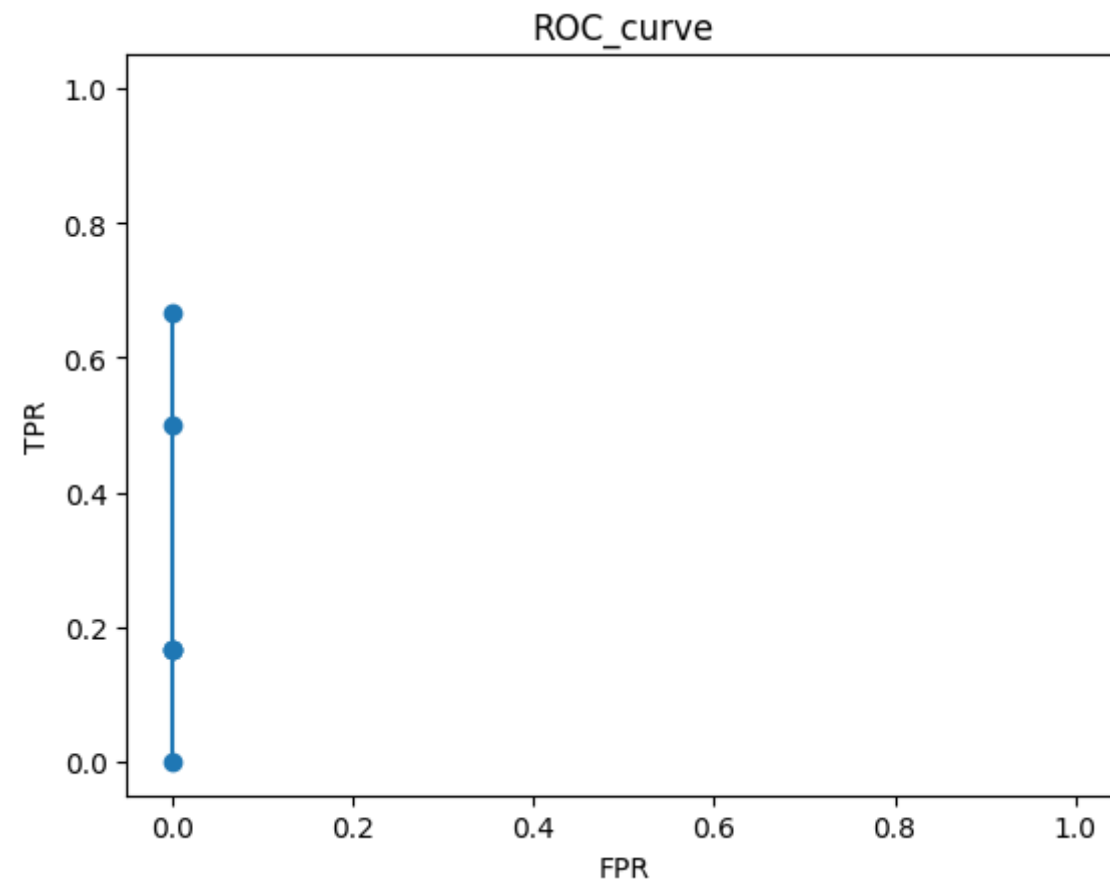
Порог = 0.84



$$TPR = \frac{TP}{TP+FN} = 0.66$$

$$FPR = \frac{FP}{FP+TN} = 0$$

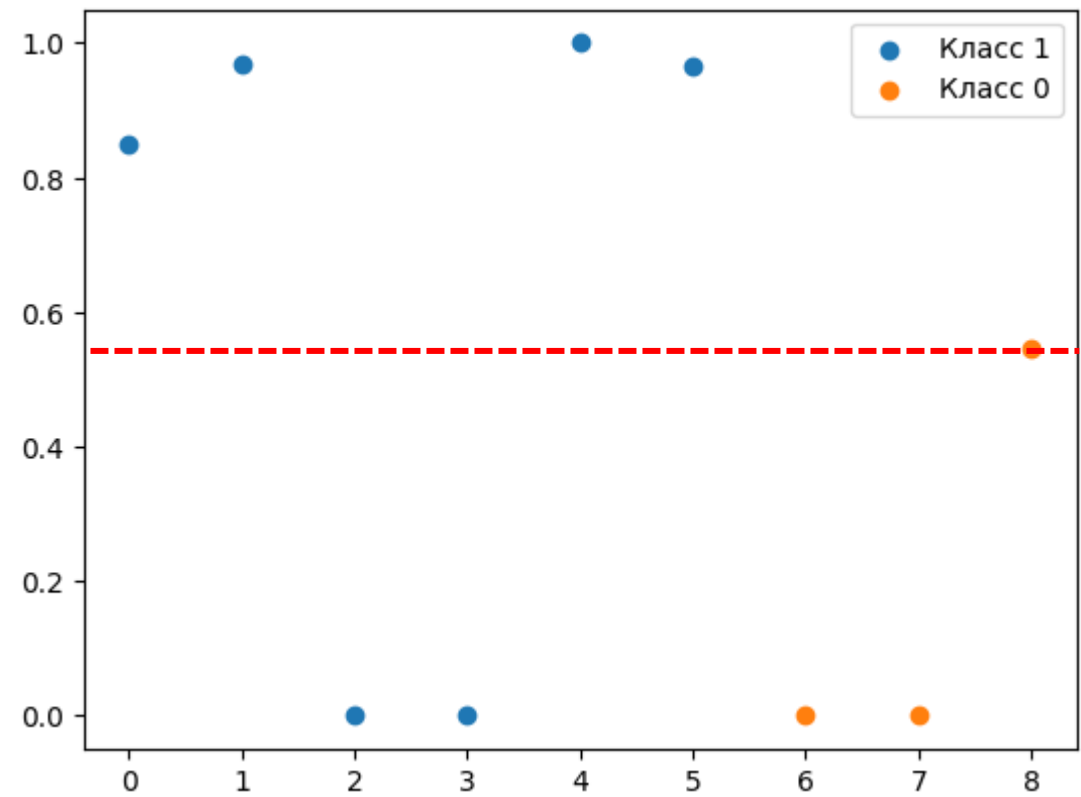
ROC-кривая



ROC-кривая

Proba	Класс	
0.0	0	0
0.0	0	0
0.0	1	0
0.0	1	0
0.54545455	0	1
0.84905661	1	1
0.96629213	1	1
0.98675023	1	1
1.0	1	1

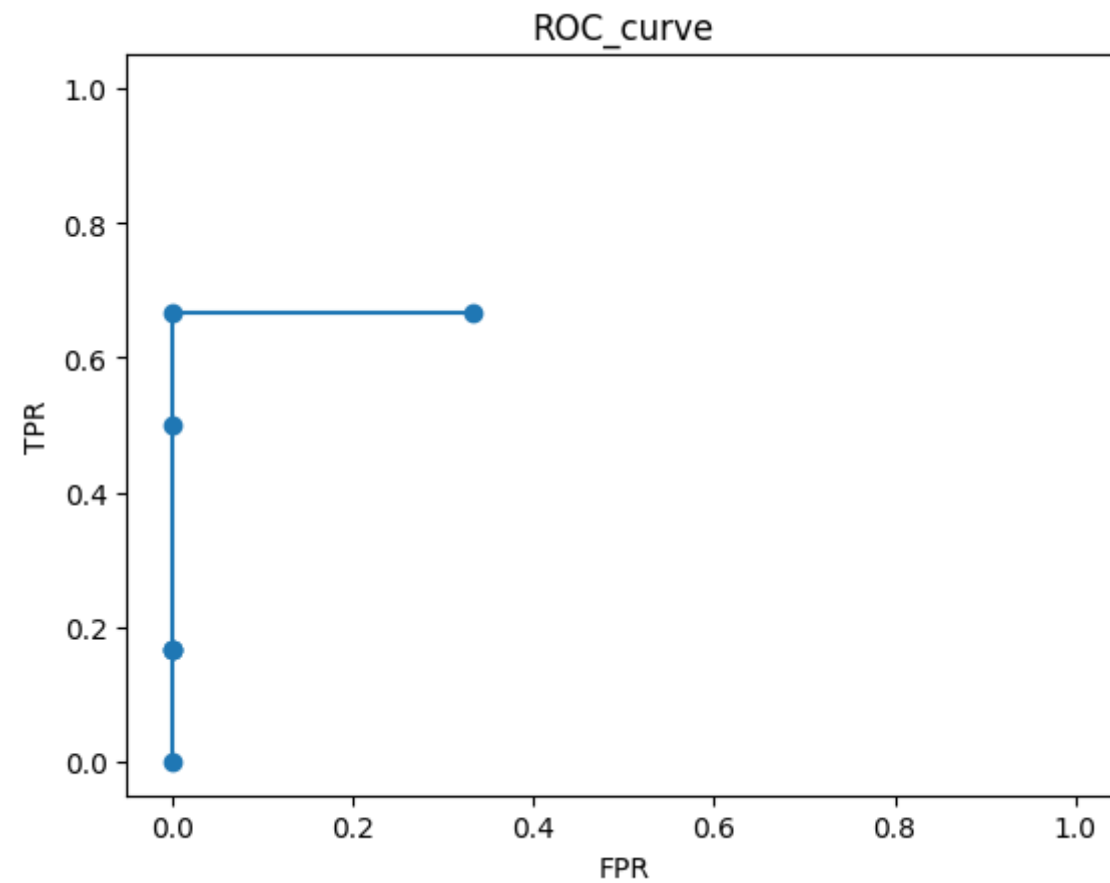
Порог = 0.54



$$TPR = \frac{TP}{TP+FN} = \frac{4}{4+2} = 0.667$$

$$FPR = \frac{FP}{FP+TN} = \frac{1}{1+2} = 0.334$$

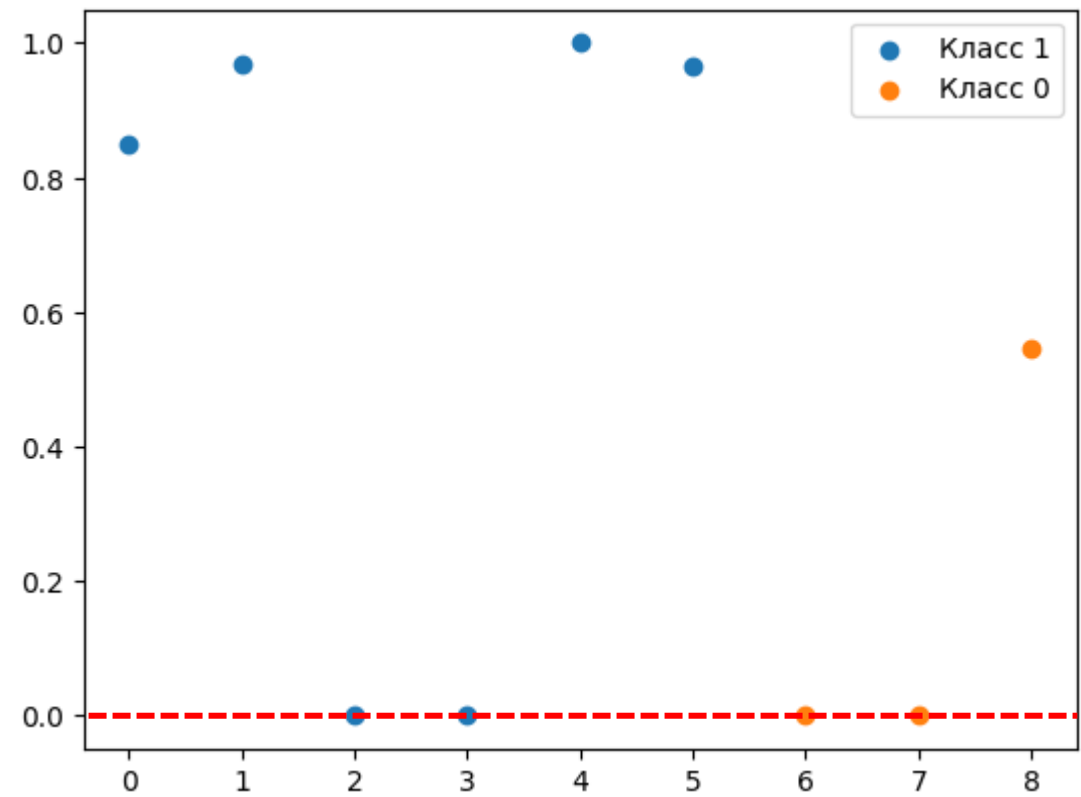
ROC-кривая



ROC-кривая

Proba	Класс	
0.0	0	1
0.0	0	1
0.0	1	1
0.0	1	1
0.54545455	0	1
0.84905661	1	1
0.96629213	1	1
0.98675023	1	1
1.0	1	1

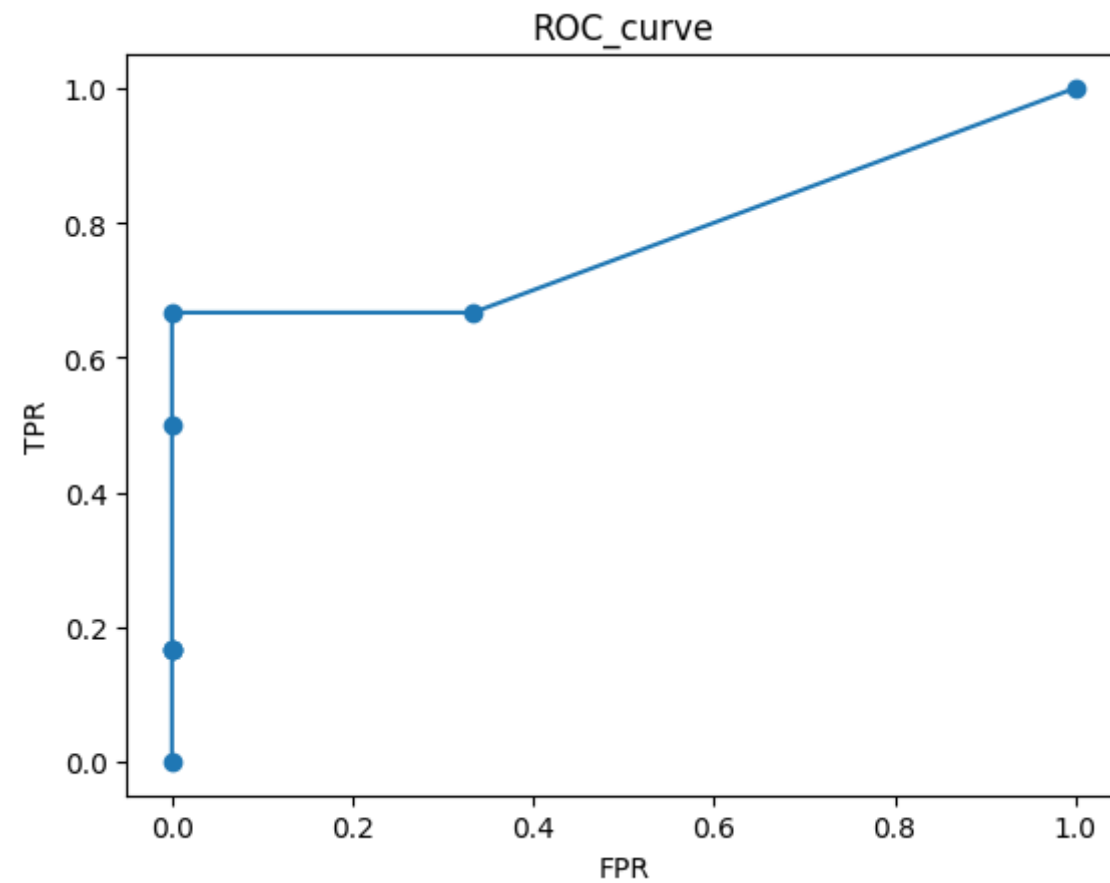
Порог = 0.0



$$TPR = \frac{TP}{TP+FN} = \frac{6}{6+0} = 1$$

$$FPR = \frac{FP}{FP+TN} = \frac{3}{3+0} = 1$$

ROC-кривая



ROC-AUC

Свойства ROC-AUC:

- ROC-AUC принимает значение от 0 до 1.

ROC-AUC идеальной модели равен 1

- ROC-AUC модели, которая случайным образом предсказывает классы, равен 0.5.
- Чем больше ROC-AUC, тем лучше наш алгоритм сортирует объекты по вероятностям - то есть тем лучше алгоритм решает задачу классификации.

