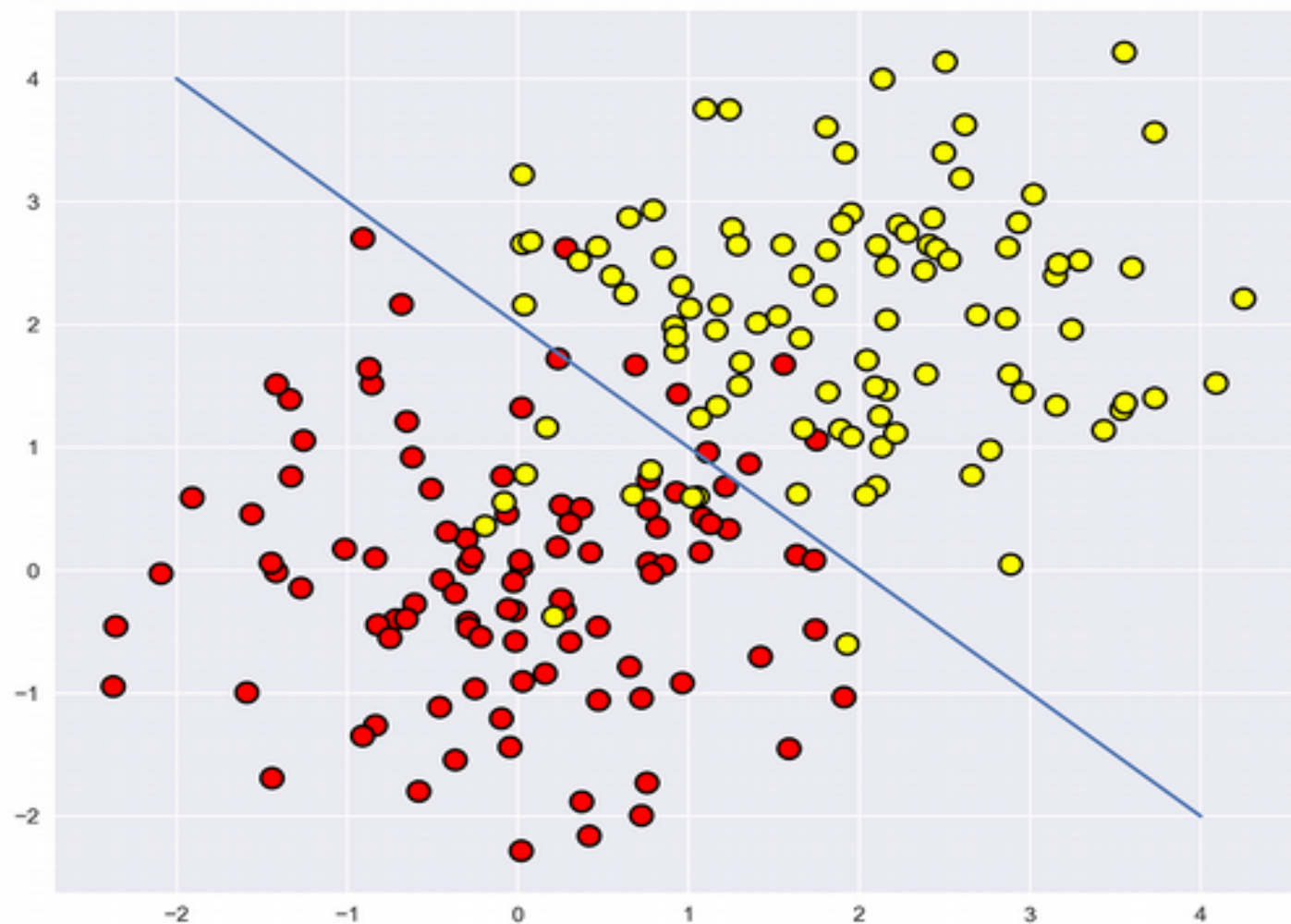


# **Введение в задачу классификации**

# Бинарная классификация

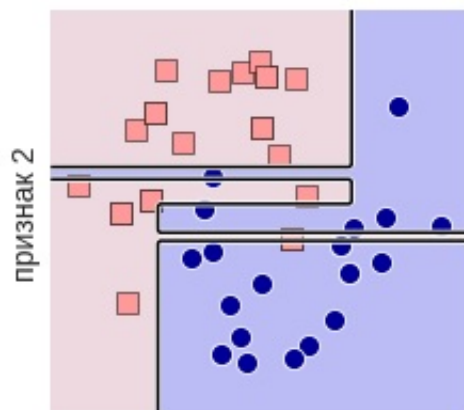


$$T = \{(X, d)\}$$

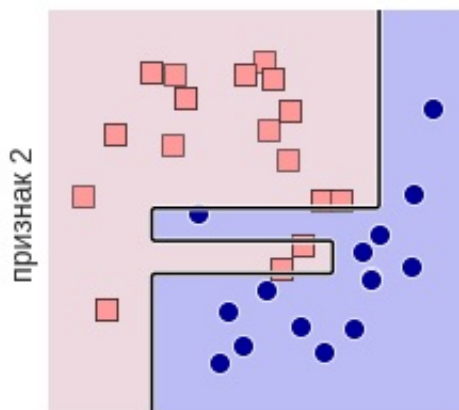
$$D = \{0, 1\}$$

$$\sum [y(x) \neq d] \rightarrow \min$$

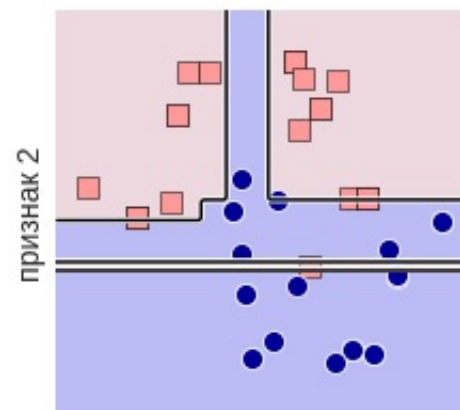
# Деревья решений – Случайный лес



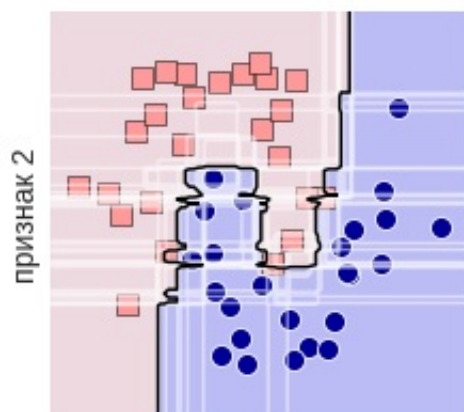
признак 1  
дерево № 1



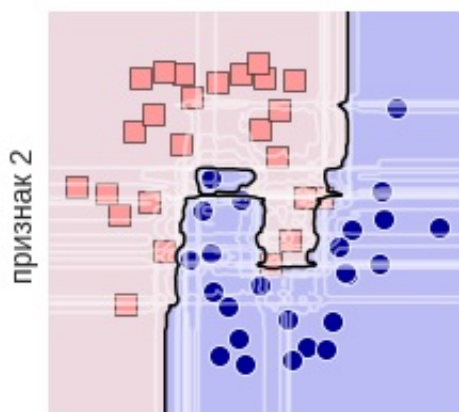
признак 1  
дерево № 2



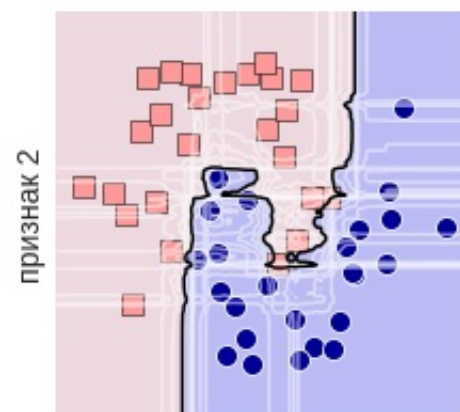
признак 1  
дерево № 3



признак 1  
RF, число деревьев=10



признак 1  
RF, число деревьев=100



признак 1  
RF, число деревьев=1000

# Дерево решений

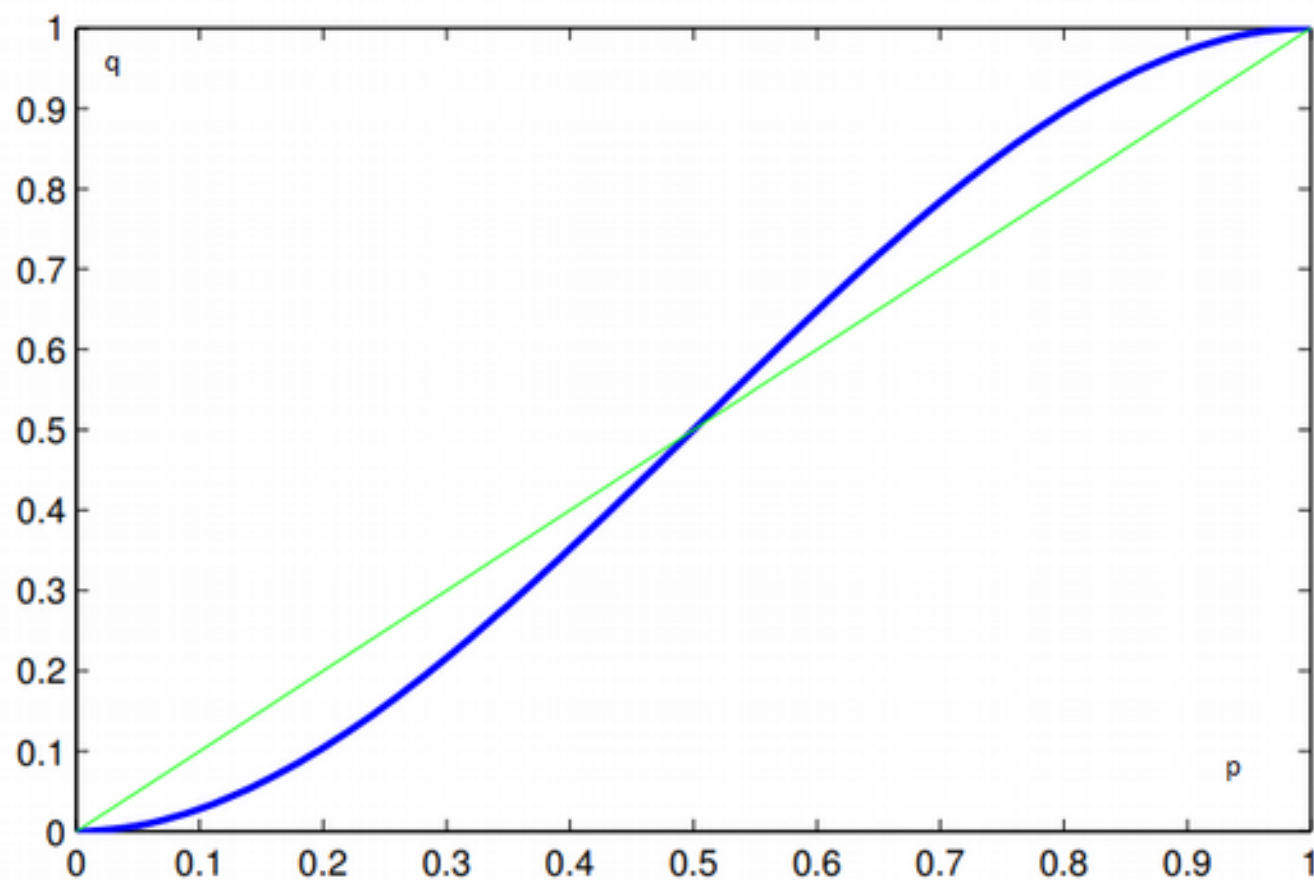
Давать ли кредит?



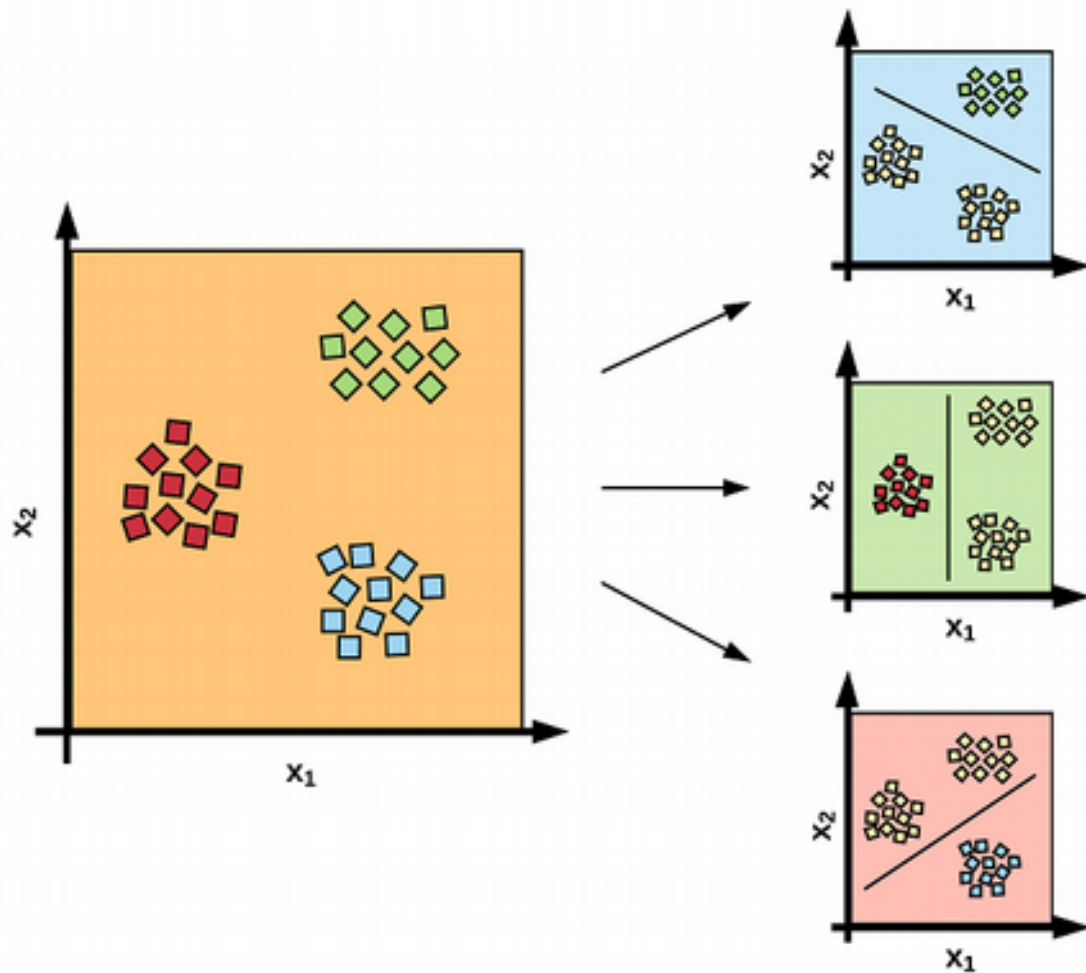
Дер

# Ансамбли

$$q = p^3 + 3p^2(1 - p) = p^2(3 - 2p).$$



# One-vs-All



$D=\{1,2,3\}$

$D = \{d1, d2, d3\}$

$D_i = \{0,1\}$

One-hot\_encoding

$1 = [1,0,0]$

$2 = [0,1,0]$

$3 = [0,0,1]$

# Метрики

## Confusion matrix (матрица ошибок)

		Actual class	
		Cat	Non-cat
Predicted class	Cat	5 True Positives	2 False Positives
	Non-cat	3 False Negatives	3 True Negatives

8 — кошки

5 — не кошки

# Метрики

## Accuracy

		Actual class	
		Cat	Non-cat
Predicted class	Cat	5 True Positives	2 False Positives
	Non-cat	3 False Negatives	3 True Negatives

$$accuracy = \frac{T}{N} = \frac{TP + TN}{TP + TN + FP + FN}$$

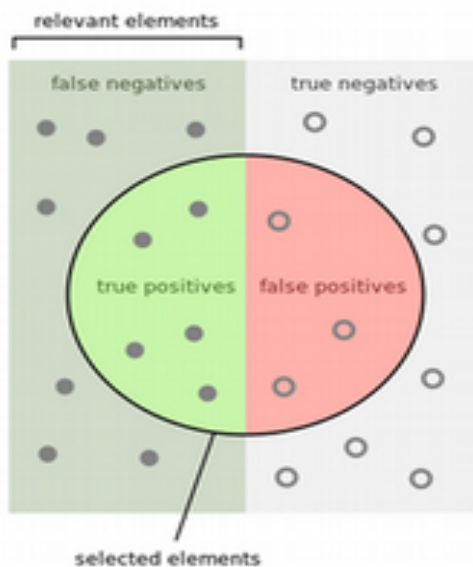


# Метрики

## Precision, Recall

		Actual class	
		Cat	Non-cat
Predicted class	Cat	5 True Positives	2 False Positives
	Non-cat	3 False Negatives	3 True Negatives

$$precision = \frac{TP}{TP + FP} \quad recall = \frac{TP}{TP + FN}$$



How many selected items are relevant?

$$Precision = \frac{\text{Green}}{\text{Green} + \text{Red}}$$

How many relevant items are selected?

$$Recall = \frac{\text{Green}}{\text{Green} + \text{Grey}}$$

# Метрики

## Гармоническая мера

		Actual class	
		Cat	Non-cat
Predicted class	Cat	5 True Positives	2 False Positives
	Non-cat	3 False Negatives	3 True Negatives

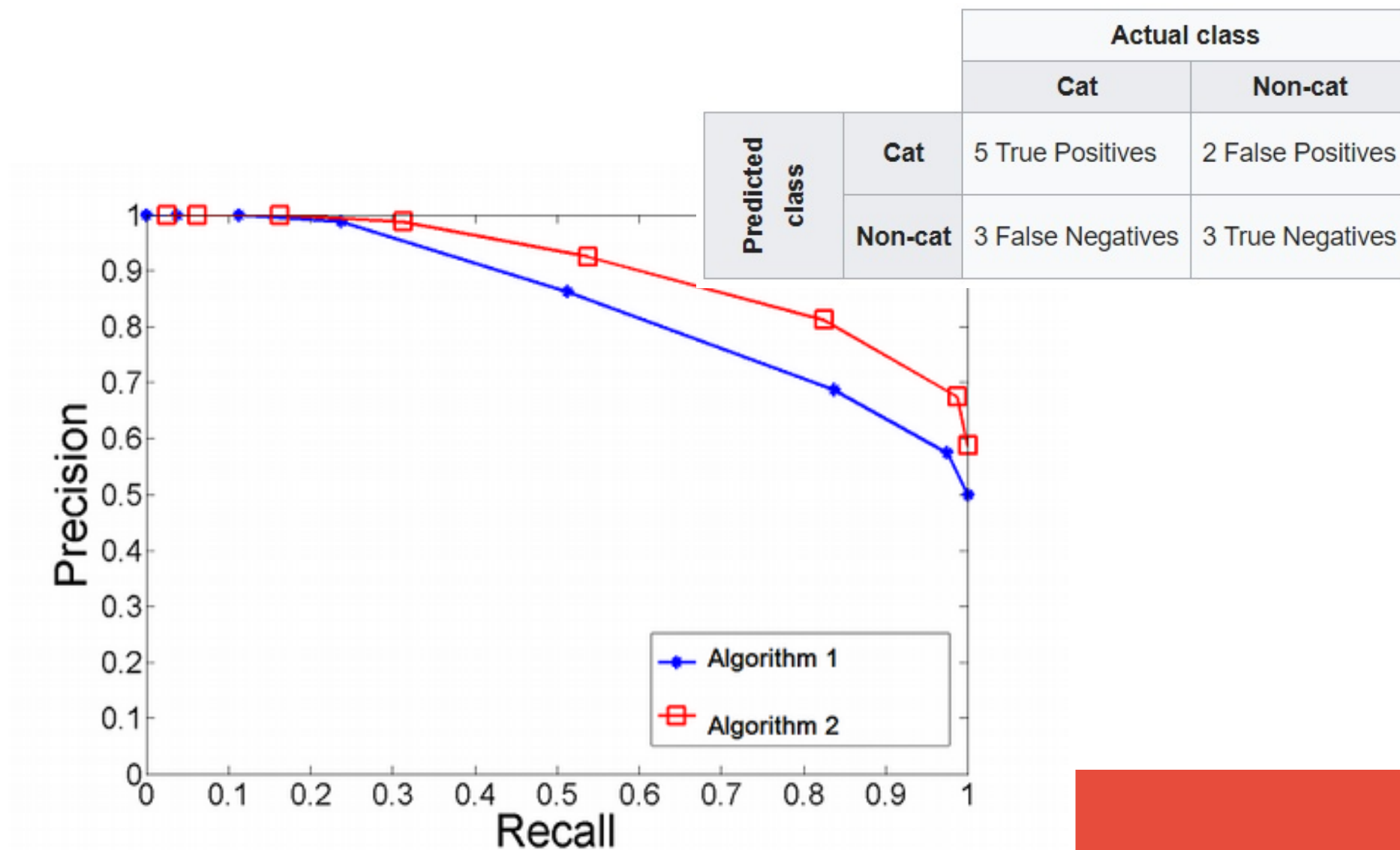
$$F_{\beta} = (1 + \beta^2) \cdot \frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall}$$

$$\beta=1$$

$$\beta>1$$

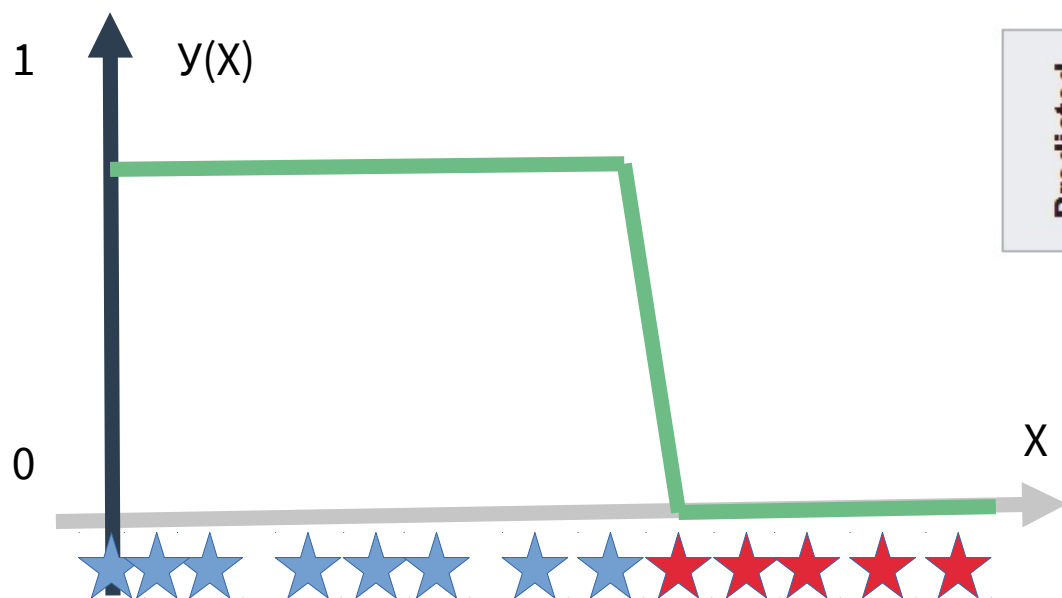
$$0<\beta<1$$

# Метрики



# Метрики

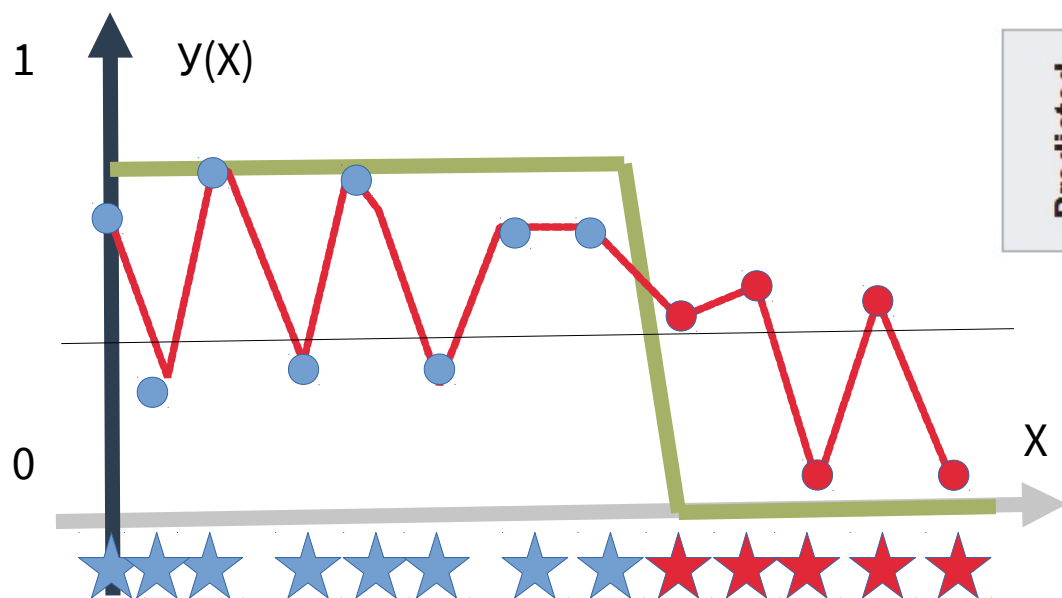
Классификатор



		Actual class	
		Cat	Non-cat
Predicted class	Cat	5 True Positives	2 False Positives
	Non-cat	3 False Negatives	3 True Negatives

# Метрики

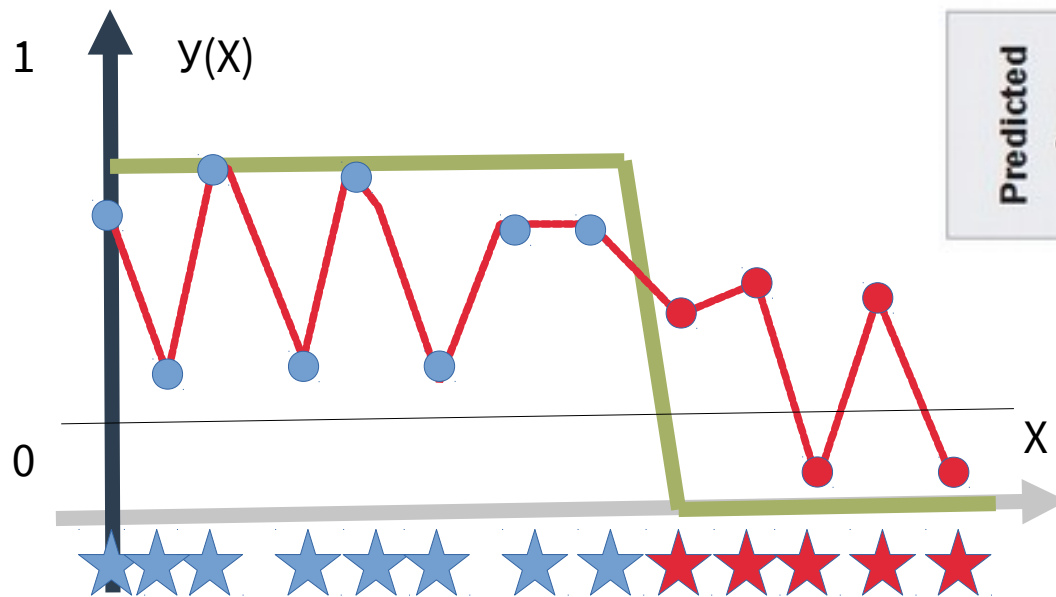
Классификатор



		Actual class	
		Cat	Non-cat
Predicted class	Cat	5 True Positives	2 False Positives
	Non-cat	3 False Negatives	3 True Negatives

# Метрики

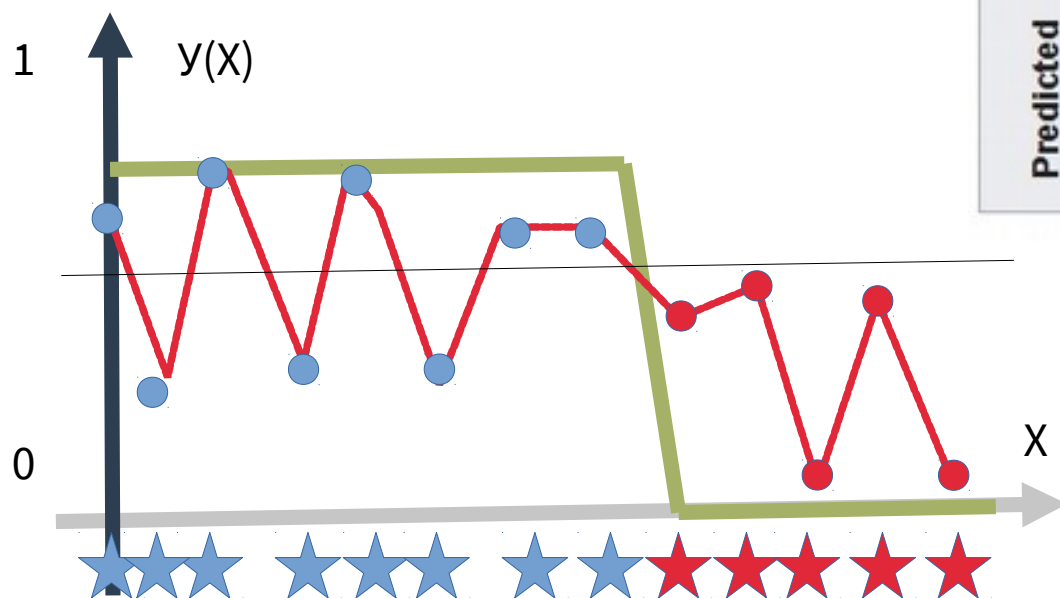
Классификатор



		Actual class	
		Cat	Non-cat
Predicted class	Cat	8 True Positives	3 False Positives
	Non-cat	0 False Negatives	2 True Negatives

# Метрики

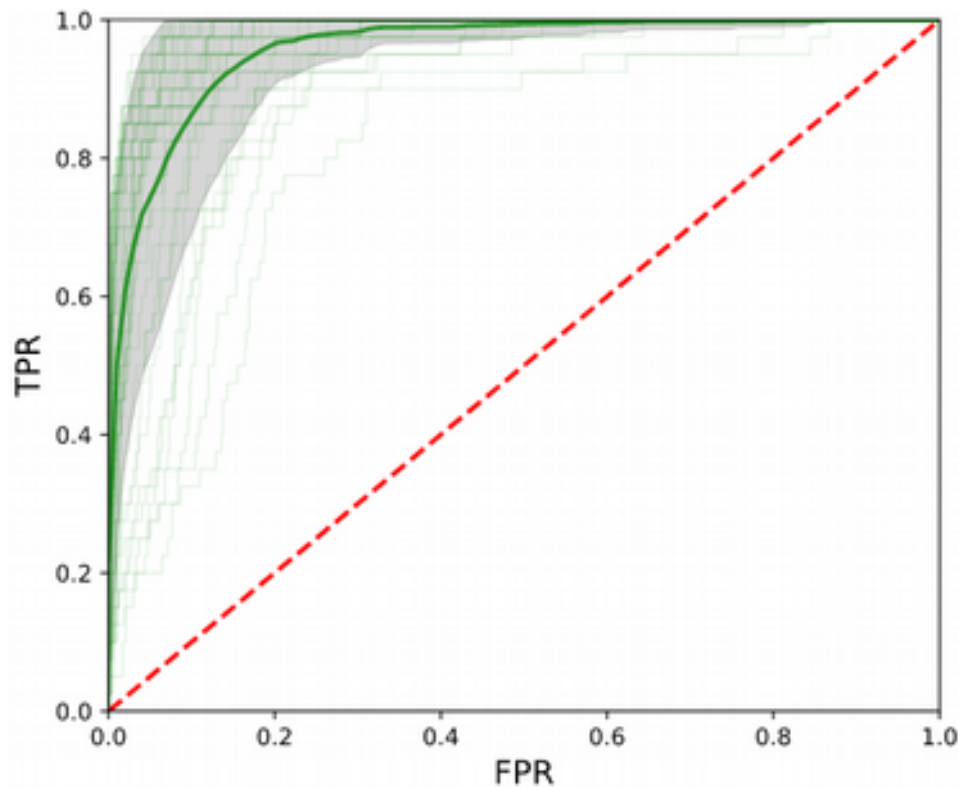
Классификатор



		Actual class	
		Cat	Non-cat
Predicted class	Cat	5 True Positives	0 False Positives
	Non-cat	3 False Negatives	5 True Negatives

# Метрики

ROC - AUC



$$TPR = \frac{TP}{TP + FN} \quad FPR = \frac{FP}{FP + TN}$$

		Actual class	
		Cat	Non-cat
Predicted class	Cat	5 True Positives	2 False Positives
	Non-cat	3 False Negatives	3 True Negatives

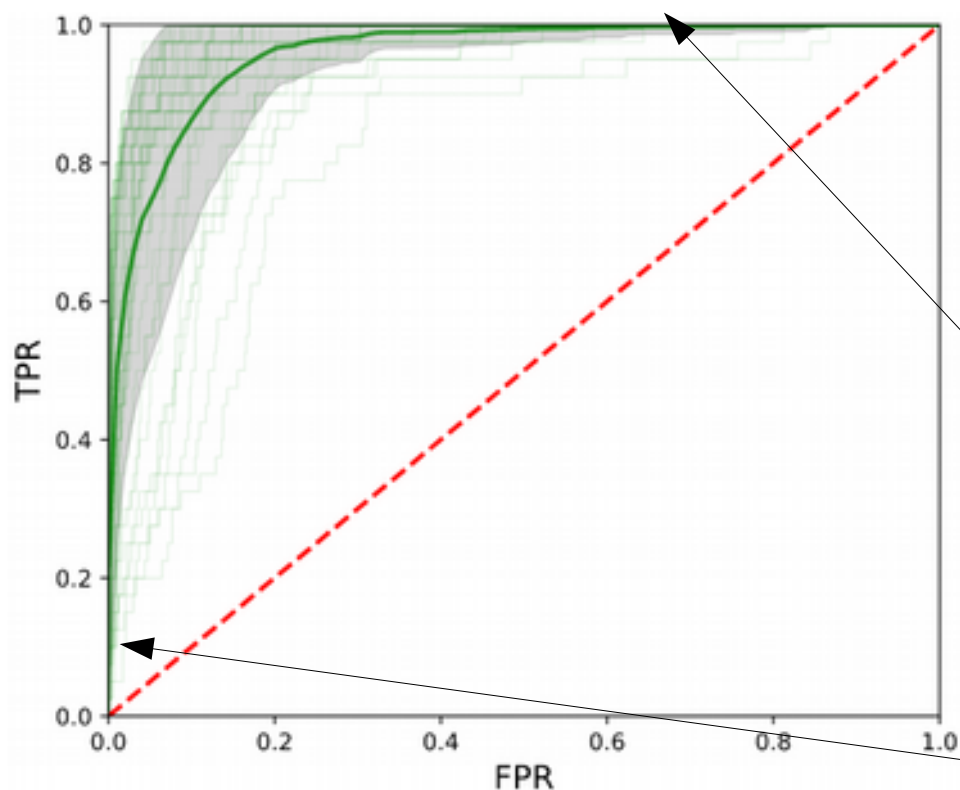
		Actual class	
		Cat	Non-cat
Predicted class	Cat	8 True Positives	3 False Positives
	Non-cat	0 False Negatives	2 True Negatives

		Actual class	
		Cat	Non-cat
Predicted class	Cat	5 True Positives	0 False Positives
	Non-cat	3 False Negatives	5 True Negatives



# Метрики

ROC - AUC



$$TPR = \frac{TP}{TP + FN} \quad FPR = \frac{FP}{FP + TN}$$

		Actual class	
		Cat	Non-cat
Predicted class	Cat	5 True Positives	2 False Positives
	Non-cat	3 False Negatives	3 True Negatives

		Actual class	
		Cat	Non-cat
Predicted class	Cat	8 True Positives	3 False Positives
	Non-cat	0 False Negatives	2 True Negatives

		Actual class	
		Cat	Non-cat
Predicted class	Cat	5 True Positives	0 False Positives
	Non-cat	3 False Negatives	5 True Negatives