



# Data learning

Курс “Машинное обучение”  
Лабораторная работа



## Binary confusion matrix based measures

Сорокин С.В., М16-524  
Вариант 1-08

# Исходные данные

## Дано:

Два класса данных:

50 – Положительные примеры

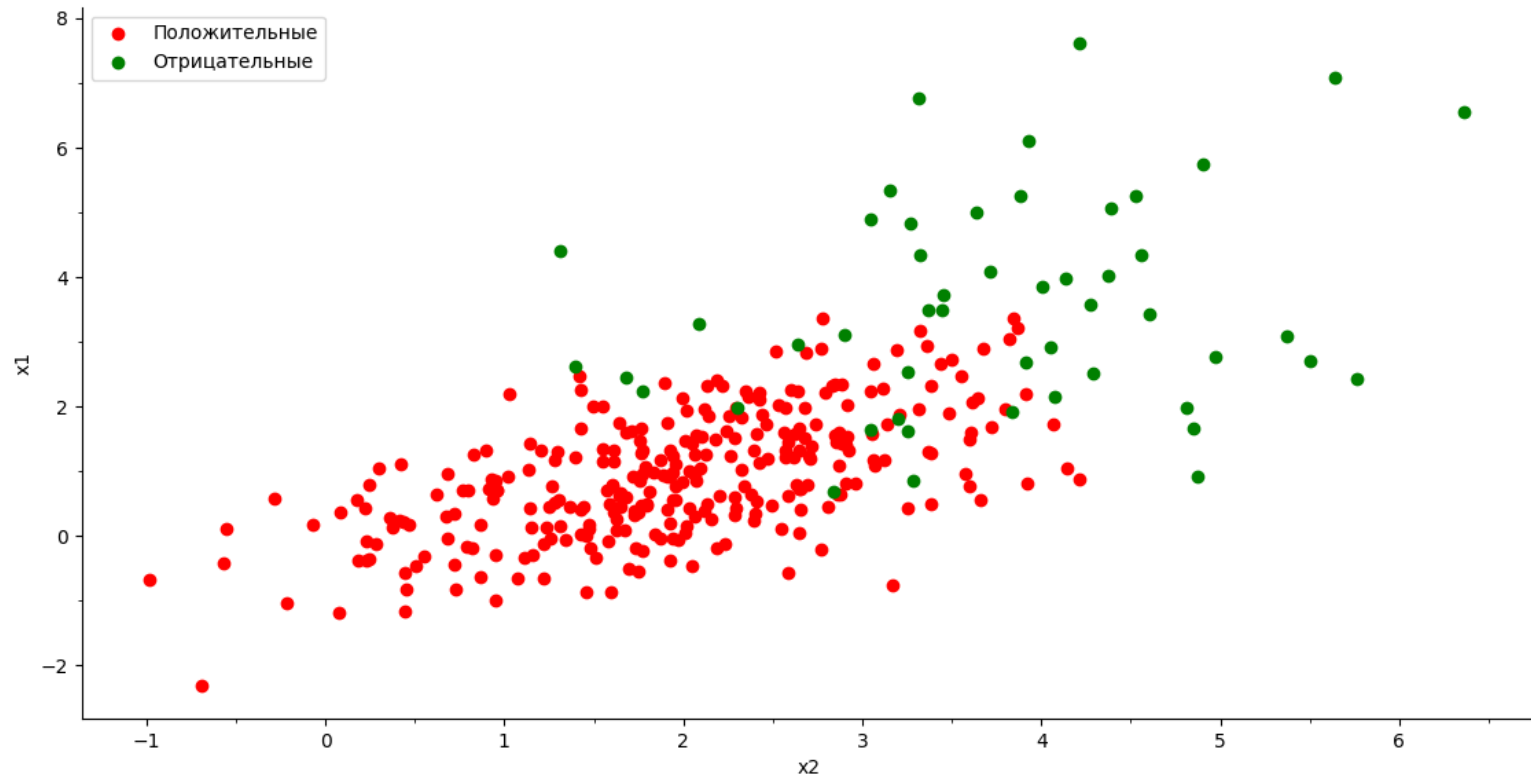
300 – Отрицательные примеры

Результат работы классификатора на каждом примере в виде числа `score`

## Задача:

Оценить качество классификации

# Исходные данные



. Рисунок 1. Диаграмма рассеяния для исходных данных

# Используемые методы и формулы

## Таблица сопряженности:

		Ответ классификатора	
		Отрицательный	Положительный
Исходные значения	Отрицательный	TN	FP
	Положительный	FN	TP

# Используемые методы и формулы

$$ERROR RATE = \frac{FP+FN}{P+N}$$



$$ACCURACY = \frac{TP+TN}{P+N}$$



# Используемые методы и формулы

$$SENSITIVITY = \frac{TP}{P}$$



$$SPECIFICITY = \frac{TN}{N}$$



# Используемые методы и формулы

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



$$FALL\ OUT = \frac{FP}{N}$$



# Используемые методы и формулы

$$F_{\beta} = \frac{(1 + \beta^2) * PREC * SENS}{\beta^2 * PREC + SENS}$$

$$Kappa = \frac{ACC - ACC_0}{1 - ACC_0}$$

$$ACC_0 = \frac{TN + FN}{P + N} * \frac{TN + FP}{P + N} + \frac{TP + FP}{P + N} * \frac{TP + FN}{P + N}$$



# Результаты исследований

## Задание 1. Таблица сопряженности

		Ответ классификатора	
		Отрицательный	Положительный
Исходные значения	Отрицательный	TN = 295	FP = 5
	Положительный	FN = 18	TP = 32

# Результаты исследований

## Задание 2. Расчет основных показателей

Показатель	Значение
ERR	0.066
ACC	0.934
SENS	0.640
SPEC	0.983
PRE	0.865
FALL	0.017
$F_1$	0.736
KAPPA	0.737

# Результаты исследований

## Задание 3.

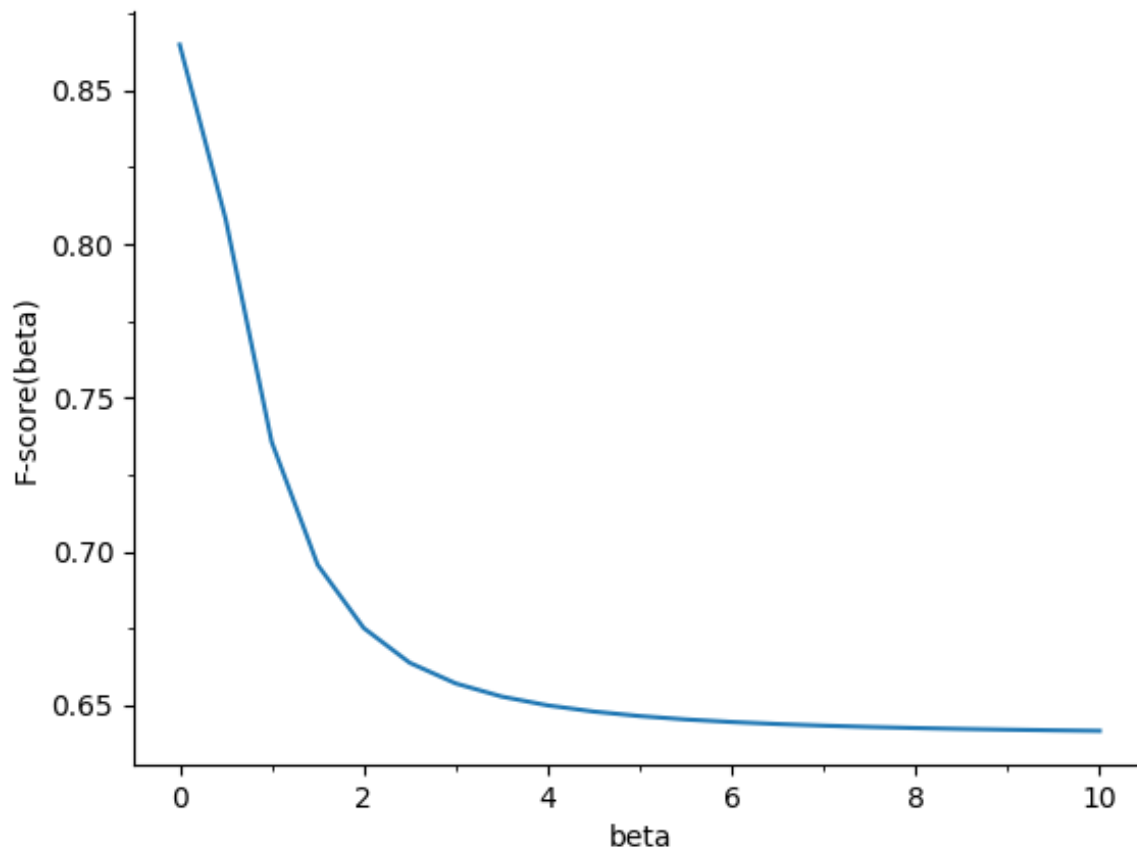


Рисунок 2. График зависимости  $F_{\beta}(\beta)$

# Результаты исследований

## Задание 4.

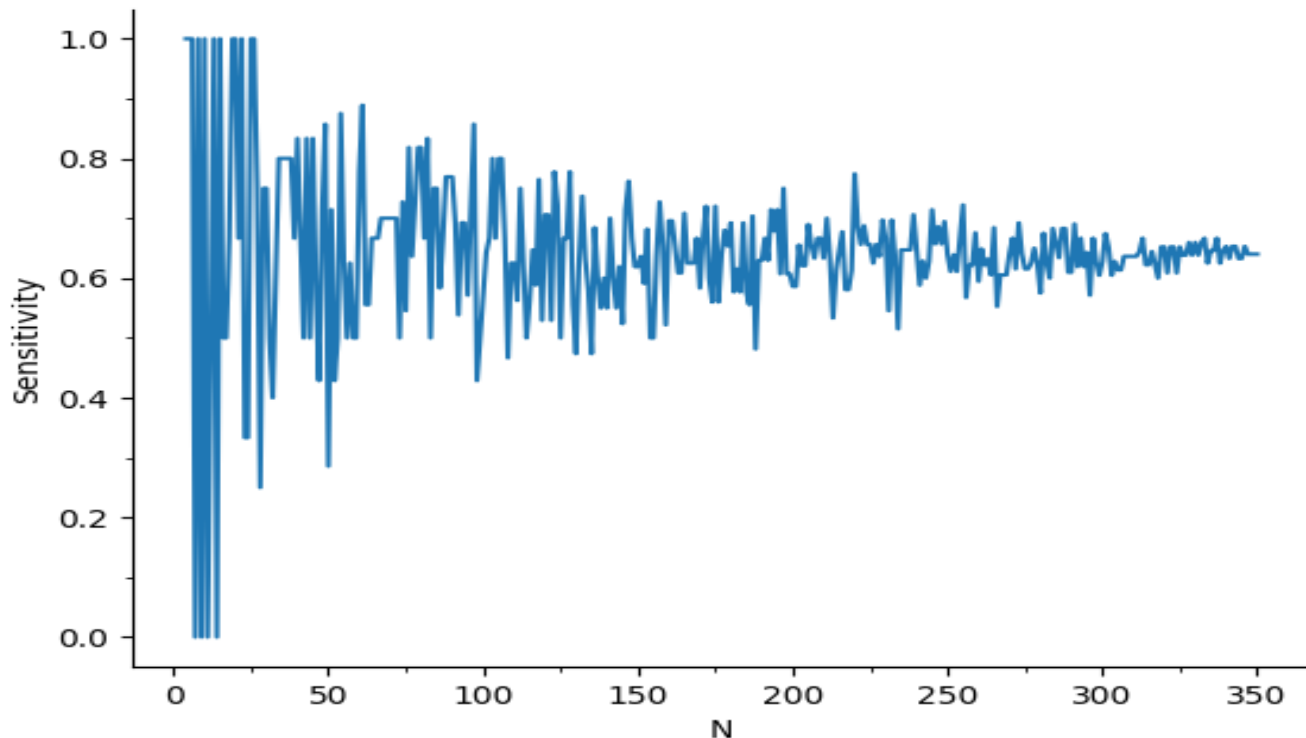


Рисунок 3. График зависимости чувствительности от  
объема выборки  
 $N \in (4; 350]$

# Результаты исследований

## Задание 4.

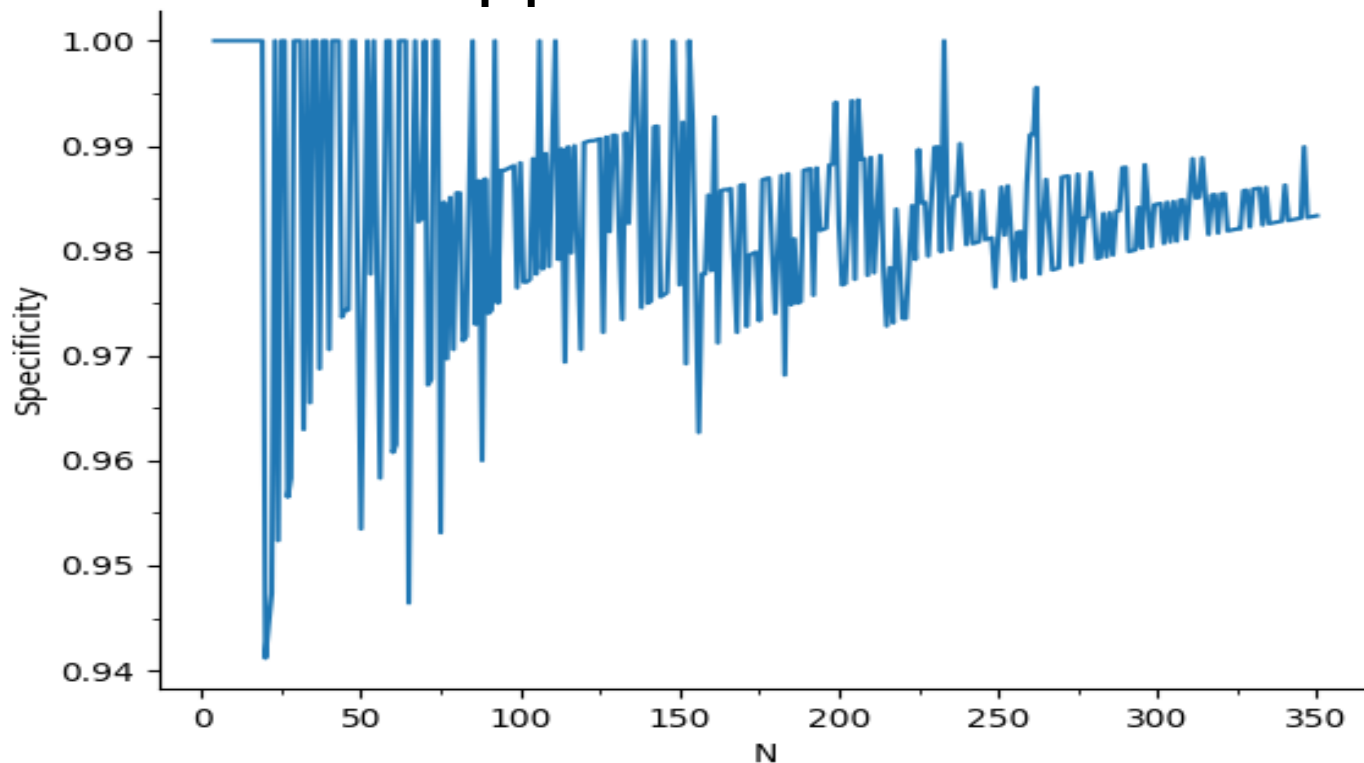


Рисунок 4. График зависимости специфичности от объема выборки

$N \in (4; 350]$

# Результаты исследований

## Задание 5. Инвертирование классов

Показатель	Исходное	Инвертированное
ERR	0.066	0.934
ACC	0.934	0.066
SENS	0.640	0.017
SPEC	0.983	0.360
PRE	0.865	0.135
FALL	0.017	0.640
$F_1$	0.736	0.030
KAPPA	0.737	-0.051

# Выводы

1. Предложенный классификатор имеет неплохую точность и низкую степень ошибки
2. Классификатор обладает высокими значениями сбалансированных оценок F-score и Карра
3. Графики зависимости чувствительности и специфичности показывают, что размер выборки не следует брать меньше 275
4. При инвертировании классов классификатор оказывается полностью неработоспособным