



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

Разработка управленческих решений в маркетинге

# ОТТОК КЛИЕНТОВ

Метрики классификации. Рекомендательные системы

Элен Теванян

Москва, 2018

# ПРИМЕР ЗАДАЧИ КЛАССИФИКАЦИИ

---

- Изготовлена тестовая партия нового продукта, которая была предложена части клиентской базы.
- Хотим понять, стоит ли выпускать его в массовое производство - будут ли все клиенты заказывать новый продукт?
- **Задача:** клиент  $\rightarrow$  [купит продукт (1)/ не купит продукт (0)]
- $x_i$  — объект, для которого строим предсказания (клиент)
- $y_i$  — целевая переменная: 1 или 0
- $(x_i, y_i)$  — прецедент
- **Обучающая выборка** — все клиенты, кому предложили тестовую партию продукта.
- Обучили классификатор  $a(x)$ . Хороший? Плохой? Поможет принять решение?

# ПРИМЕР ЗАДАЧИ КЛАССИФИКАЦИИ

---

- Пусть в эксперименте участвовало 100 клиентов
- В таблице приведены результаты обучения модели машинного обучения

	Клиент купил $y = 1$	Клиент не купил $y = 0$
Модель спрогнозировала, что клиент купил $a(x) = 1$	60 True Positive (TP)	5 False Positive (FP)
Модель спрогнозировала, что клиент не купил $a(x) = 0$	25 False Negative (FN)	10 True Negative (TN)

# ДОЛЯ ПРАВИЛЬНЫХ ОТВЕТОВ

---

- Доля правильных ответов (accuracy) показывает долю объектов в выборке, которым классификатор присвоил их истинный класс.

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

	$y = 1$	$y = 0$
$a(x) = 1$	60	5
$a(x) = 0$	25	10

# ДОЛЯ ПРАВИЛЬНЫХ ОТВЕТОВ

---

- Доля правильных ответов (accuracy) показывает долю объектов в выборке, которым классификатор присвоил их истинный класс.

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

	$y = 1$	$y = 0$
$a(x) = 1$	60	5
$a(x) = 0$	25	10

$$\text{Accuracy} = \frac{60+10}{100} = 0.7$$

# ТОЧНОСТЬ

---

- Точность (precision) показывает уровень доверия к классификатору при  $a(x) = 1$

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

	$y = 1$	$y = 0$
$a(x) = 1$	60	5
$a(x) = 0$	25	10

# ТОЧНОСТЬ

---

- Точность (precision) показывает уровень доверия к классификатору при  $a(x) = 1$

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

	$y = 1$	$y = 0$
$a(x) = 1$	60	5
$a(x) = 0$	25	10

$$P = \frac{60}{60+5} = 0.92$$

- Отвечаем на вопрос: можно ли доверять классификатору при  $a(x) = 1$

# ПОЛНОТА

---

- Полнота (recall) показывает, какую долю правильно-положительных ответов находит классификатор

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

	$y = 1$	$y = 0$
$a(x) = 1$	60	5
$a(x) = 0$	25	10



# ПОЛНОТА

---

- Полнота (recall) показывает, какую долю правильно-положительных ответов находит классификатор

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

	$y = 1$	$y = 0$
$a(x) = 1$	60	5
$a(x) = 0$	25	10

$$P = \frac{60}{60+25} = 0.71$$

- Отвечаем на вопрос: много ли положительных ответов находит классификатор?

# ОЦЕНКА ПРИНАДЛЕЖНОСТИ К КЛАССУ

---

- Будем жить в мире, где два класса: +1 и -1
- $t$  – порог
- $b(x)$  – оценка принадлежности к классу +1

- Классифицируем так:

$$a(x) = [b(x) > t]$$

# ОЦЕНКА ПРИНАДЛЕЖНОСТИ К КЛАССУ

---

- Как оценить качество  $b(x)$ ?
- Как выбрать порог?

# ОЦЕНКА ПРИНАДЛЕЖНОСТИ К КЛАССУ


---

- Как оценить качество  $b(x)$ ?
- Как выбрать порог?
- Порог выбирается позже
- Порог высокий:
  - Мало объектов к классу +1
  - Точность выше
  - Полнота ниже
- Порог низкий:
  - Много объектов относим к +1
  - Точность ниже
  - Полнота выше

# ОЦЕНКА ПРИНАДЛЕЖНОСТИ К КЛАССУ


---

-1	-1	+1	-1	-1	-1	+1	+1	-1	+1
0.01	0.09	0.12	0.15	0.29	0.4	0.48	0.6	0.83	0.9



# ОЦЕНКА ПРИНАДЛЕЖНОСТИ К КЛАССУ

---



-1	-1	+1	-1	-1	-1	+1	+1	-1	+1
0.01	0.09	0.12	0.15	0.29	0.4	0.48	0.6	0.83	0.9

# ОЦЕНКА ПРИНАДЛЕЖНОСТИ К КЛАССУ

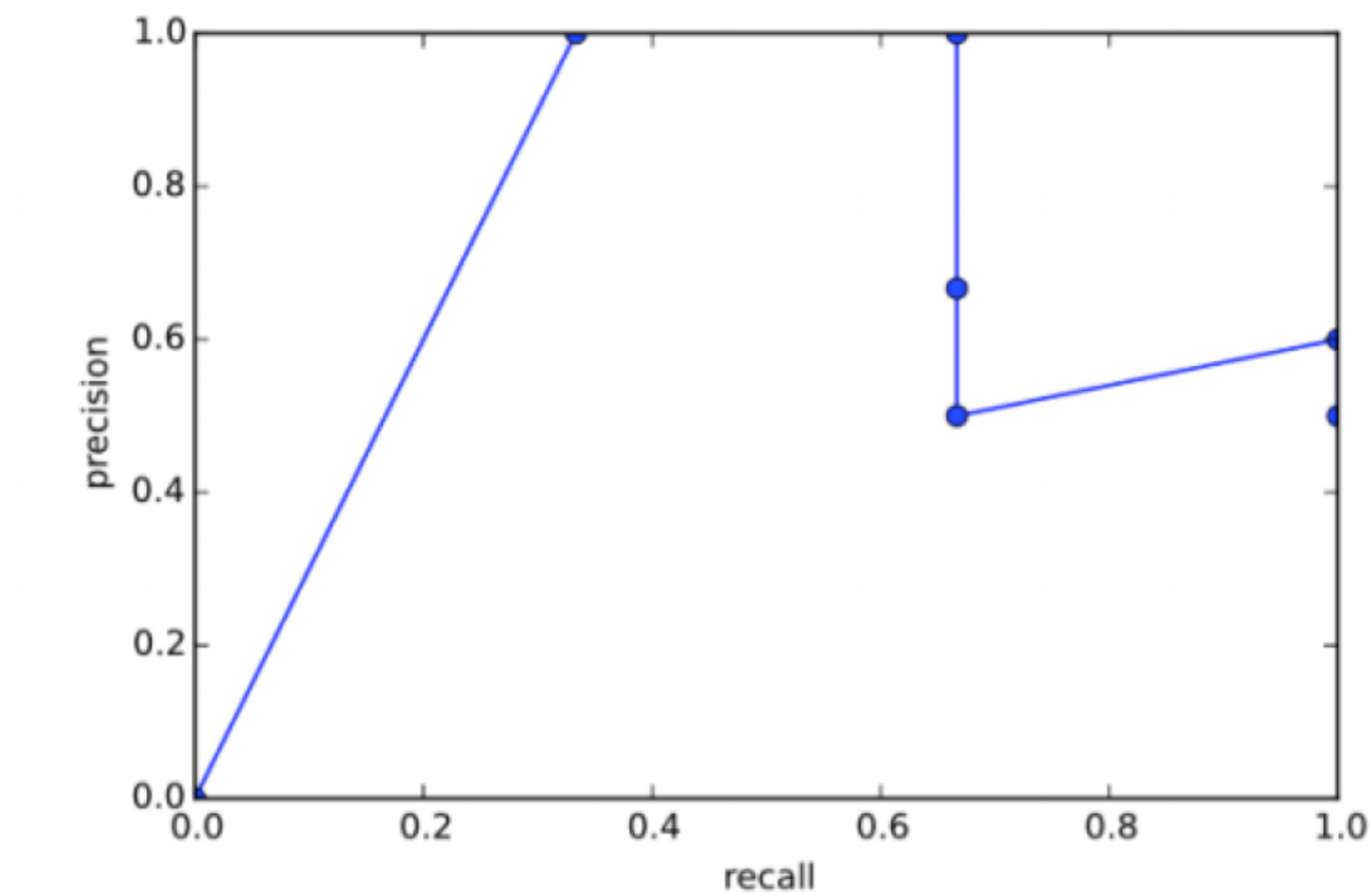
---

- Пример: кредитный скоринг
- $b(x)$  – оценка вероятности возврата кредита
- $a(x) = [b(x) > 0.5]$
- $\text{precision} = 0.1$
- $\text{recall} = 0.7$
- ??????

# PR-КРИВАЯ

---

- Кривая точности-полноты
- Ось X – полнота
- Ось Y – точность
- Точки значения полноты и точности при последовательных порогах





# ROC- КРИВАЯ

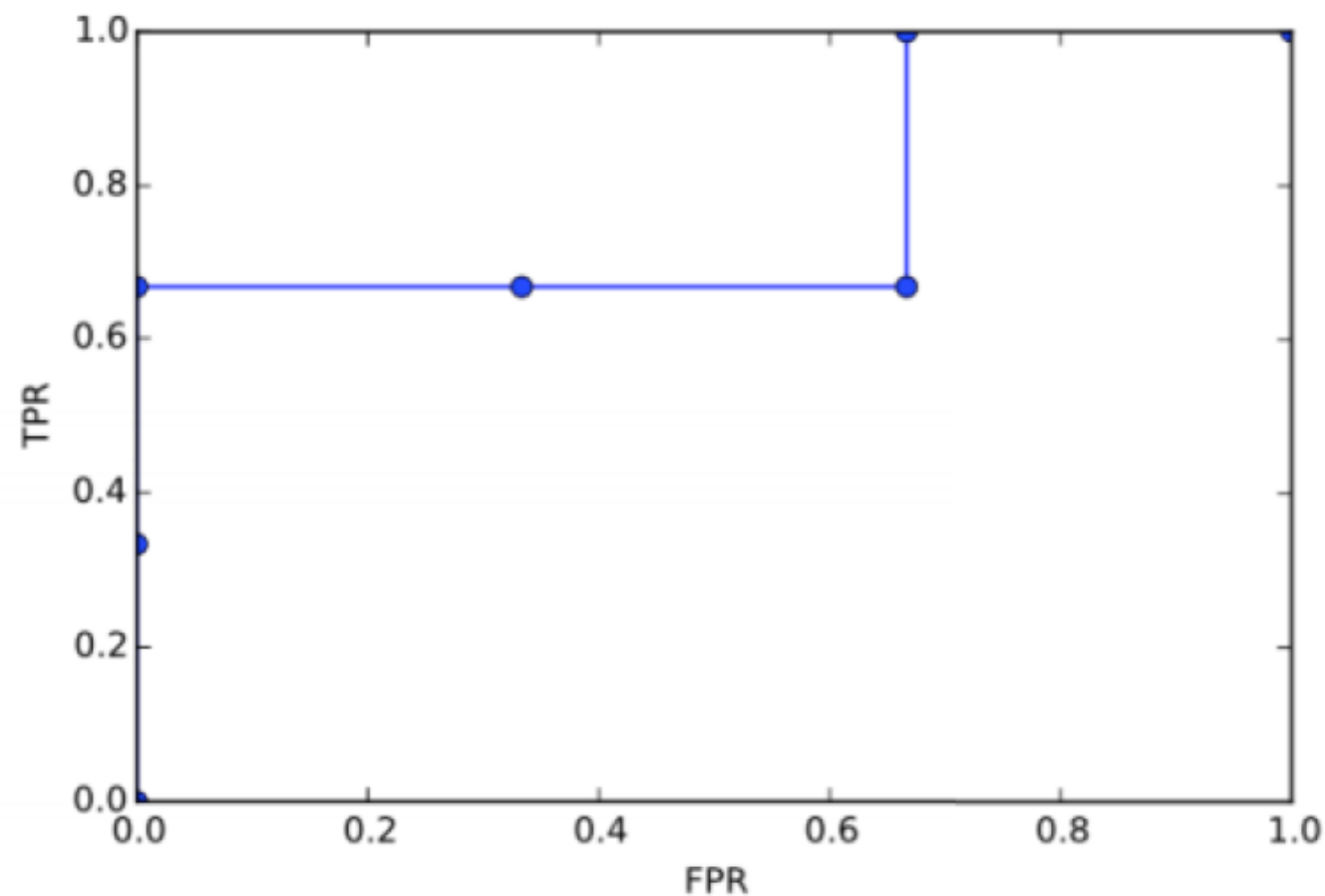
---

- Receiver Operating Characteristic
- Ось X – False Positive Rate

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

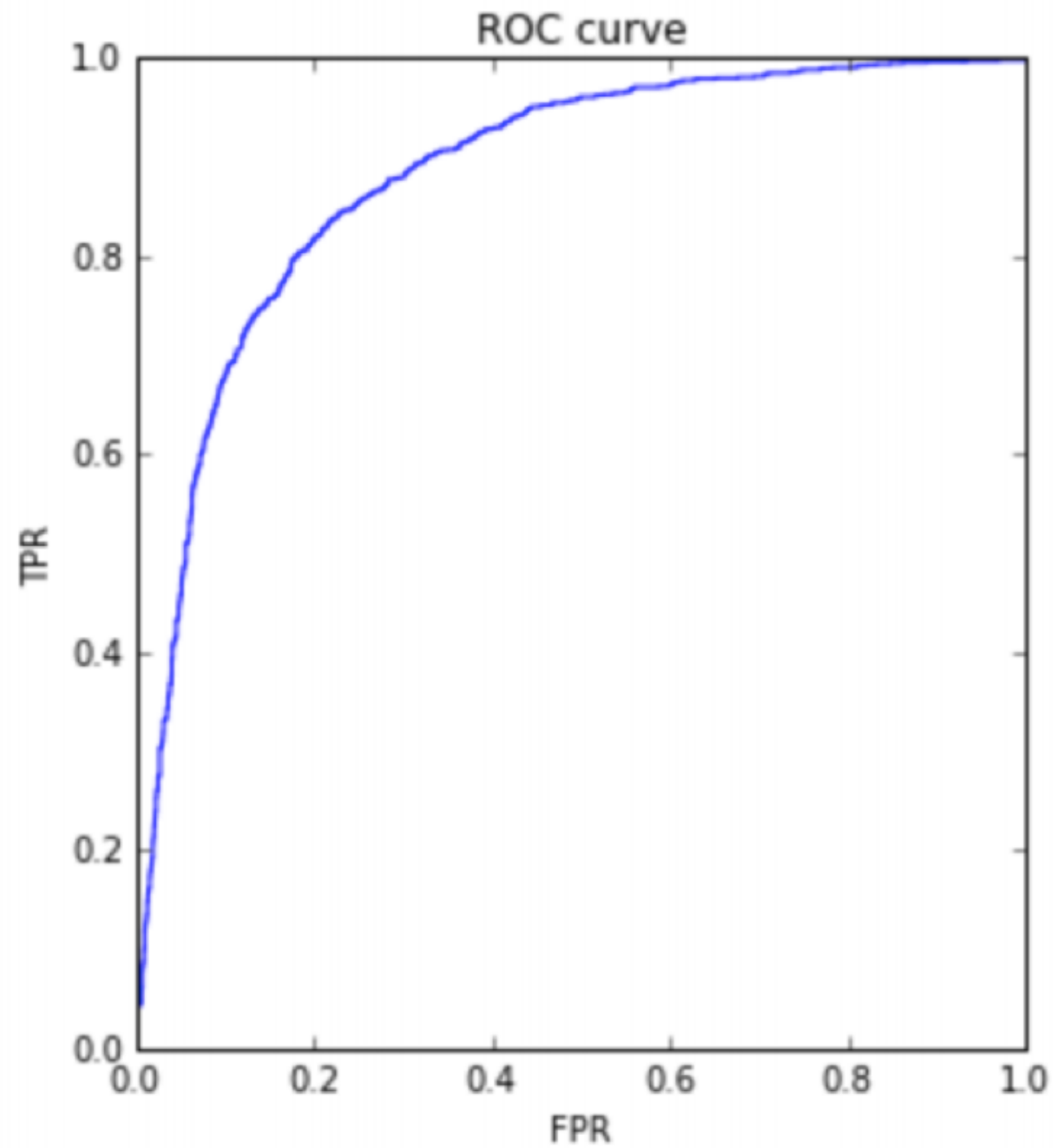
- Ось Y – True Positive Rate

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



# ROC- КРИВАЯ В РЕАЛЬНОСТИ

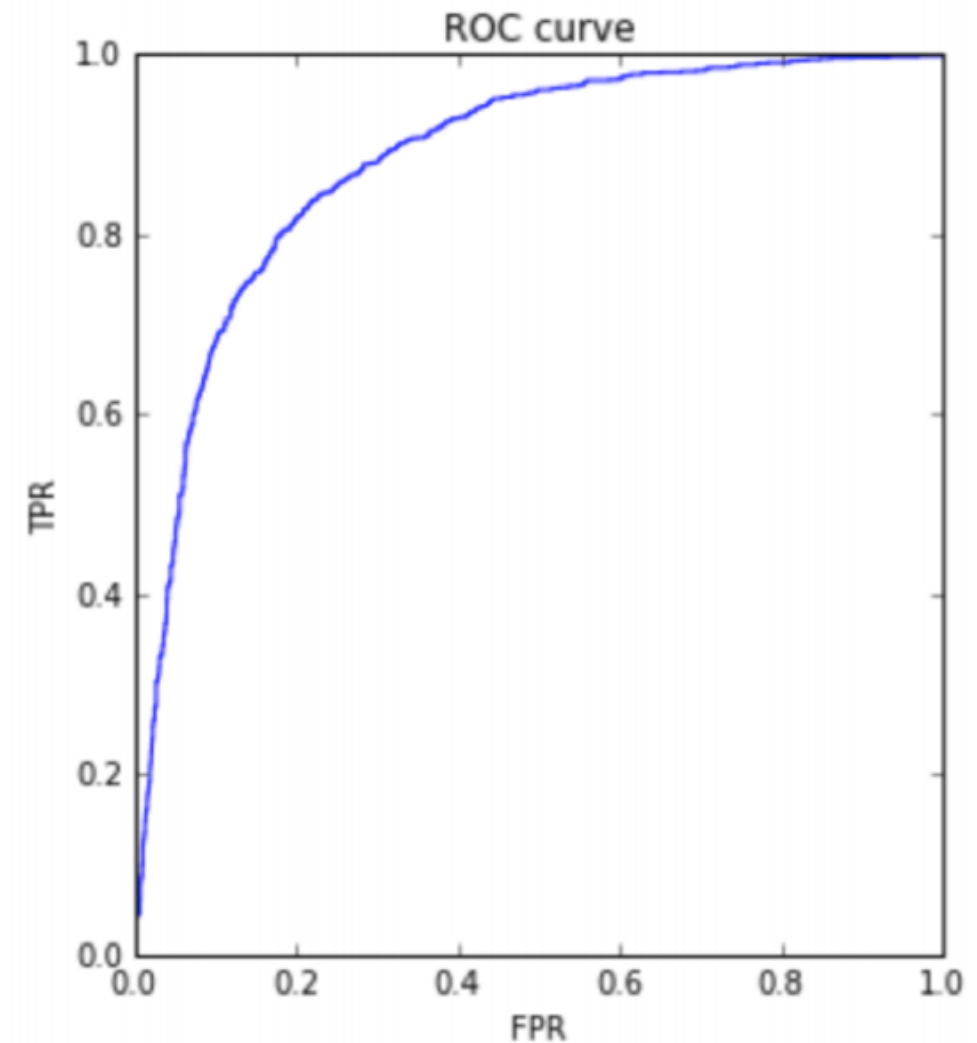
---



# ROC-КРИВАЯ

---

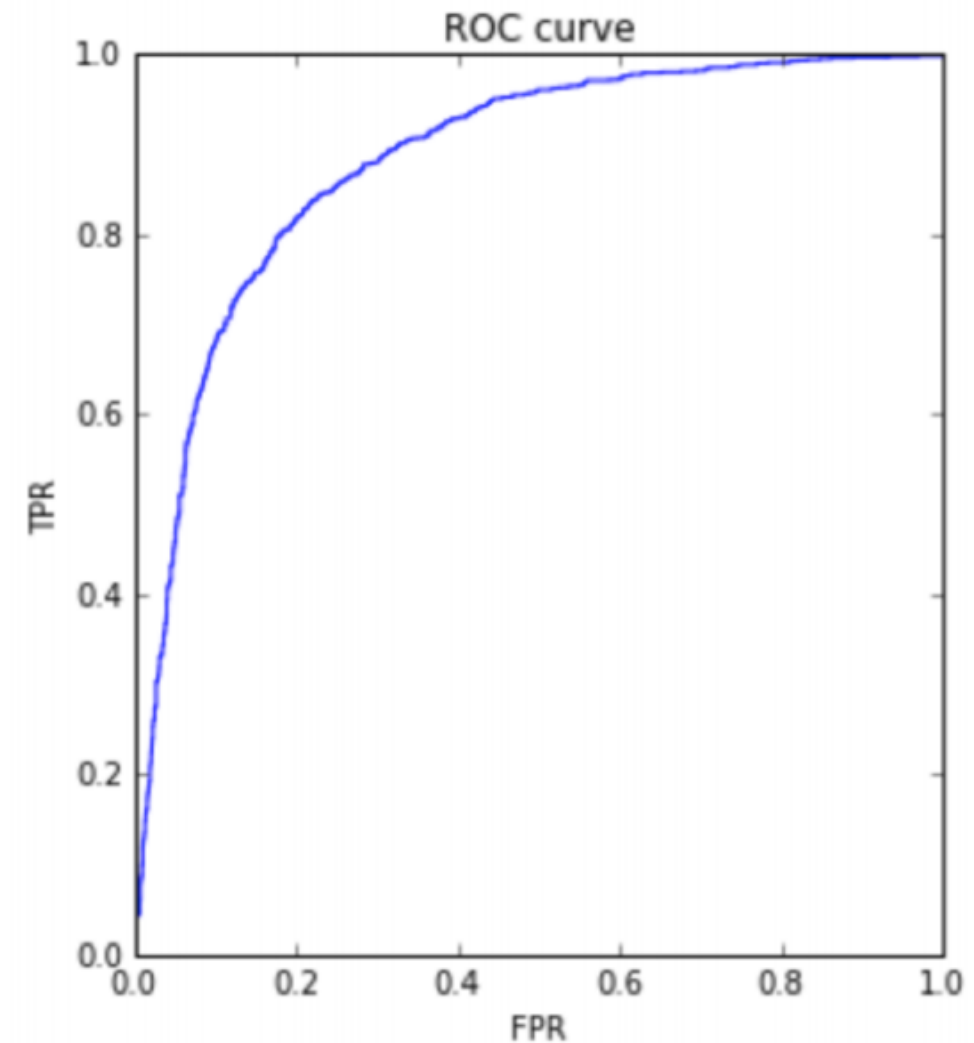
- Левая точка:  $(0, 0)$
- Правая:  $(1, 1)$
- Идеально:  
захватили точку  $(0, 1)$
- Считаем еще AUC-ROC –  
площадь под кривой



# ROC-КРИВАЯ

---

- Левая точка: (0, 0)
- Правая: (1,1)
- Идеально:  
захватили точку (0, 1)
- Считаем еще AUC-ROC –  
площадь под кривой
- Идеально: AUC-ROC = 1
- Плохо: AUC-ROC = 0.5





НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ