

# ТЕХНОЛОГИИ В ОБРАЗОВАНИИ УНИВЕРСИТЕТ

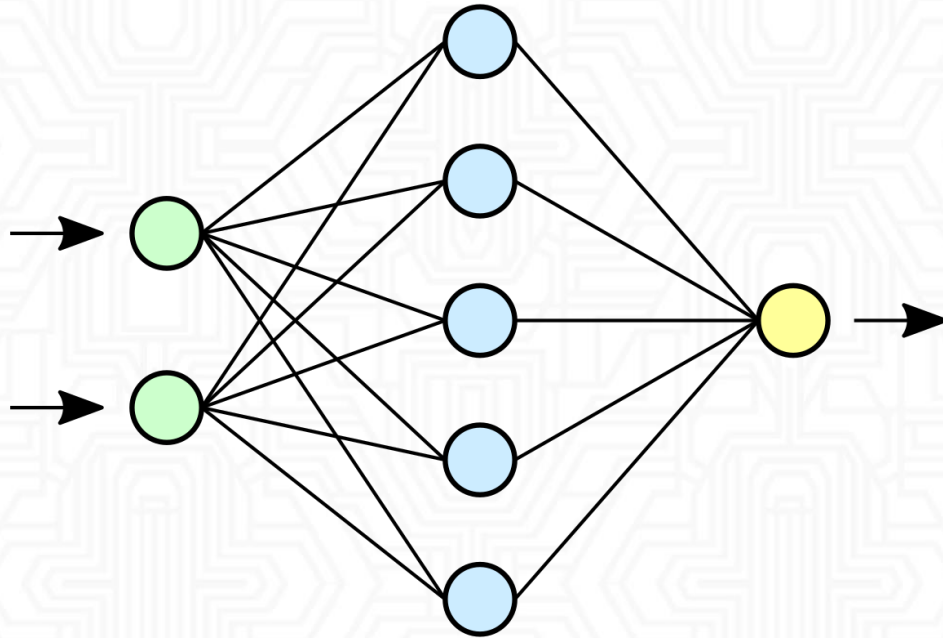
МИКРОЭЛЕКТРОНИКА  
ИННОВАЦИИ  
КАТАЛИТИЧЕСКИЕ  
МАТЕРИАЛЫ  
ДИЗАЙН  
ЛЕКАРСТВ  
ТОЧКА  
СБОРКИ  
НАУЧНАЯ  
ЛАБОРАТОРИЯ  
ГЕОХИМИЯ  
ИНЖИНИРИНГ  
ГЕОФИЗИКА  
ГИБРИДНЫЕ  
МАТЕРИАЛЫ  
ЭНЕРГОСБЕРЕЖЕНИЕ  
ВЫСОКИЕ  
ЭНЕРГИИ  
БИОТЕХНОЛОГИИ  
МОДЕЛИРОВАНИЕ  
НАНОТЕХНОЛОГИИ  
СЕМИОТИКА  
НАУКА  
МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ  
ЭЛЕМЕНТАРНЫЕ  
ЧАСТИЦЫ  
ГЕОЛОГИЯ  
КВАНТОВЫЕ  
ТЕХНОЛОГИИ  
БИОЛОГИЯ  
ТЕМНАЯ  
МАТЕРИЯ  
ФОТОНИКА  
БИОМЕДИЦИНА  
ПРИКЛАДНЫЕ  
ИССЛЕДОВАНИЯ  
РАЗВИТИЕ  
АСТРОНОМИЯ  
ГЛОБАЛЬНЫЕ ПРИОРИТЕТЫ  
АСТРОФИЗИКА  
БИОИНФОРМАТИКА  
ЛАЗЕРНАЯ  
ФИЗИКА  
АРХЕОЛОГИЯ  
ЭКОНОМИКА  
ЗНАНИЙ  
СОТРУДНИЧЕСТВО  
ИТ  
DEEP  
LEARNING  
ИЗУЧЕНИЕ  
МОЗГА  
АРКТИКА  
КОГНИТИВНЫЕ ТЕХНОЛОГИИ

**N\*** Новосибирский  
государственный  
университет  
**\*НАСТОЯЩАЯ НАУКА**



## Нейронные сети прямого распространения

## \* НЕЙРОННАЯ СЕТЬ



## \* ОБУЧАЕМОСТЬ

- \* Способность делать прогнозы относительно большого набора данных путем выборки небольшого числа точек данных
- \* Однако, существует бесконечно много способов выбора меньшего множества, но размер этой бесконечности неизвестен
- \* Обобщение - отражение основной взаимосвязи между признаками и классами
- \* Алгоритм должен хорошо работать на новых данных, которых он раньше не видел, а не только на тех, что использовались для обучения модели

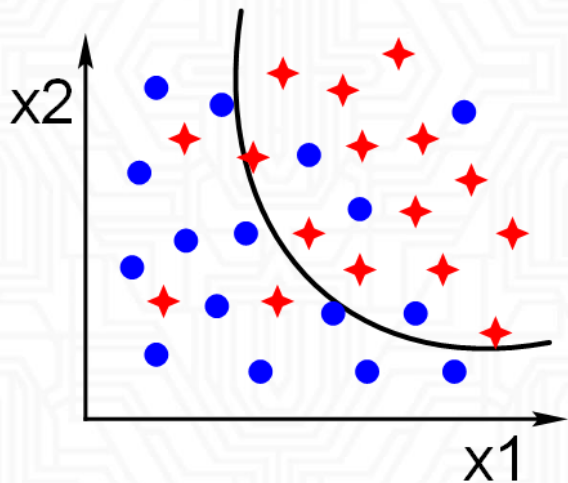
## \* ТЕОРЕТИЧЕСКАЯ ОШИБКА КЛАССИФИКАТОРА

- \*  $D$  – распределение вероятностей образцов
- \* Есть функция пометки, неизвестная алгоритму обучения,  $f: X \rightarrow Y$  такая, что  $y_i = f(x_i)$  для любых  $i$ .
- \* Ошибка классификатора (ошибка обобщения, риск, потеря) – вероятность предсказания неправильной метки для случайно точки, выбранной из  $D$  ( $h(x) \neq f(x)$ ).

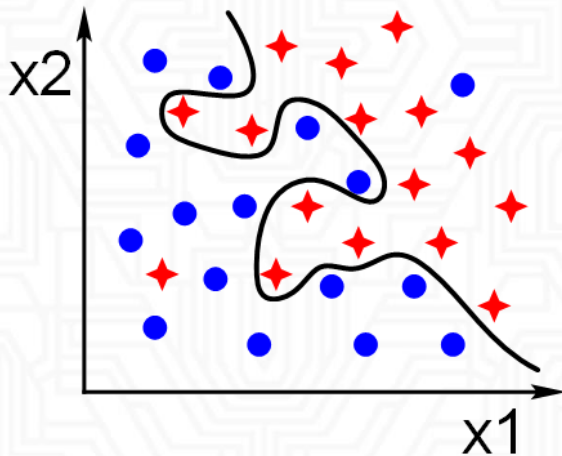
$$L_{D,f}(h) \stackrel{\text{def}}{=} P_{x \sim D}[h(x) \neq f(x)] \stackrel{\text{def}}{=} D(\{x: h(x) \neq f(x)\})$$

## \* НЕПРАВИЛЬНОЕ ОБУЧЕНИЕ

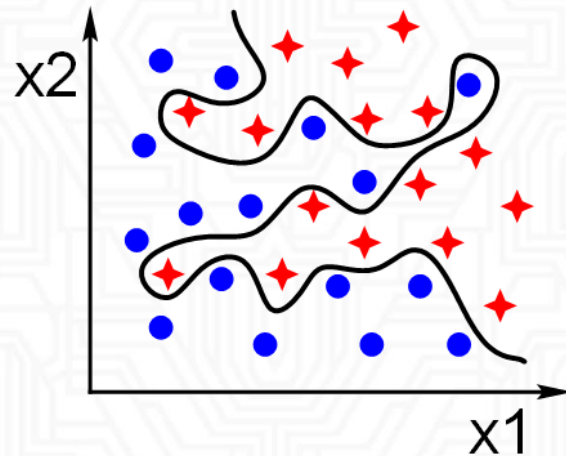
Недообучение



Оптимум



Переобучение



## \* НЕДООБУЧЕНИЕ

- \* Недообучение (underfitting) – когда модель, построенная с помощью алгоритма, является слишком упрощенной, чтобы представлять базовую взаимосвязь между признаками и классом в обучающей выборке.
- \* Это явление можно заметить по большой ошибке на обучающей выборке (еще говорят, что «не удаётся настроиться на выборку»). Помимо простоты модели, недообучение может возникать еще и из-за малого количества эпох обучения.

## \* ПЕРЕОБУЧЕНИЕ

- \* Переобучение (overfitting) – когда модель, построенная с помощью алгоритма, настолько сложна, что модель слишком точно приближает обучающую выборку и становится чувствительной к шуму.
- \* Это явление можно заметить по увеличивающейся разнице между ошибкой на обучающей выборке и тестовой выборке с каждой эпохой обучения. Поэтому при обучении строится график изменения ошибки на обучающей и тестовой выборках. Переобученная модель обладает низкой обобщающей способностью, в эксплуатации она будет часто ошибаться.

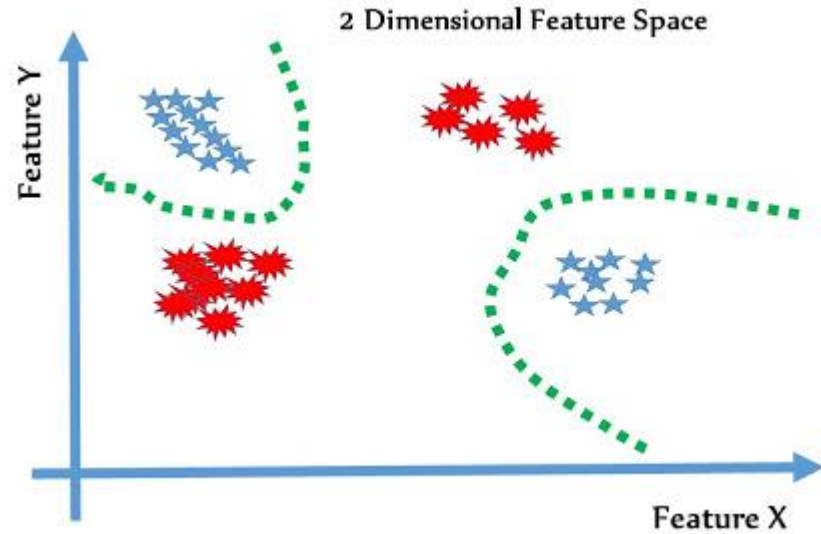
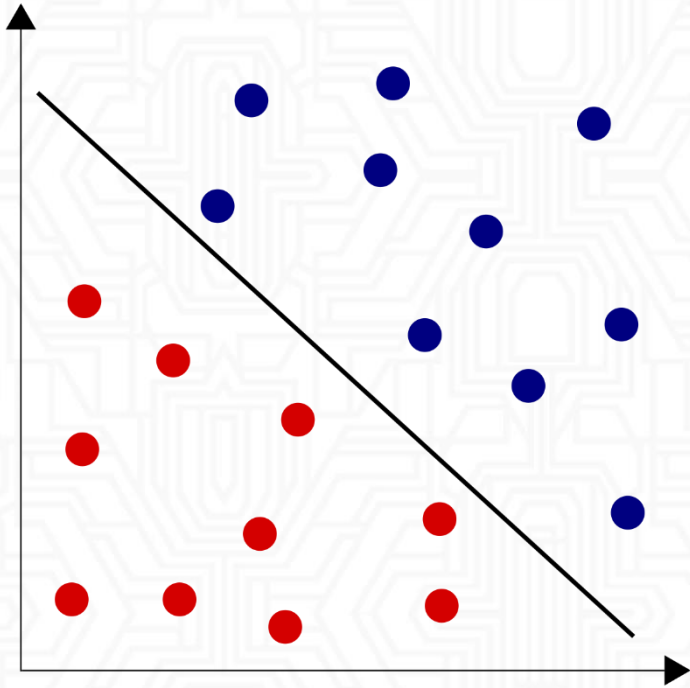


## \* ПРИНЦИП РАЗДЕЛИМОСТИ

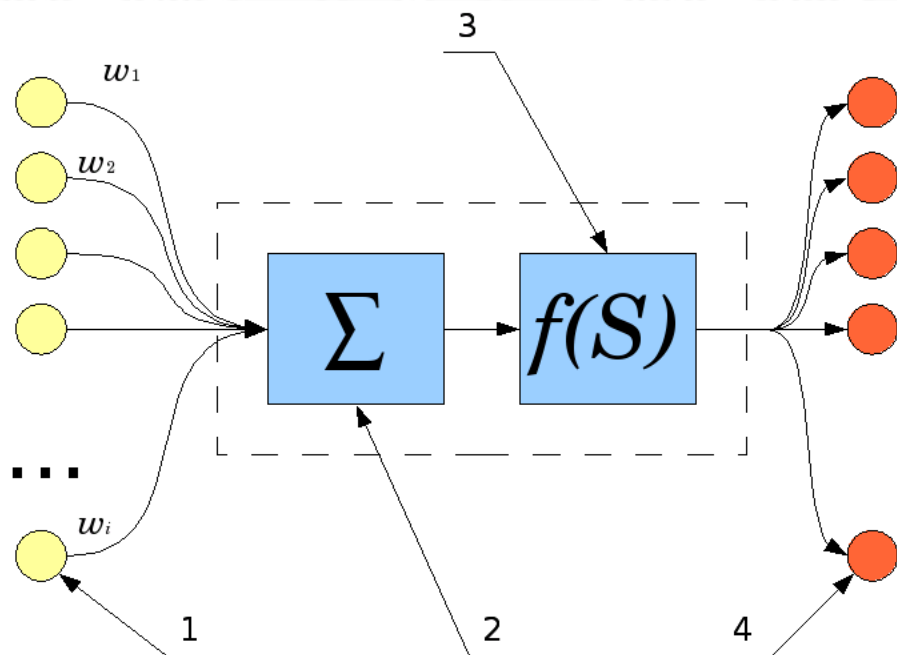
- \* некоторые не пересекающиеся множества могут быть некоторым образом разделены в пространстве
- \* принцип разделимости требует доказательства обоснованности применения в каждом конкретном случае.
- \* Виды:
  - \* Линейная
  - \* Нелинейная



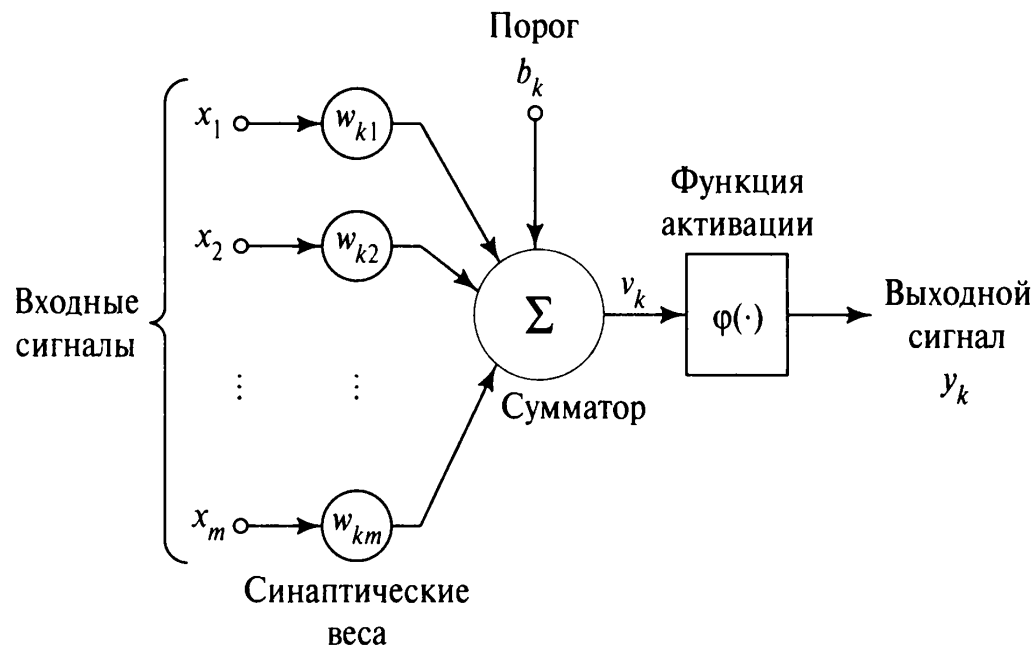
## \* РАЗДЕЛИМОСТЬ



## \* ФОРМАЛЬНЫЙ НЕЙРОН



## \* ФОРМАЛЬНЫЙ НЕЙРОН



$$u_k = \sum_{j=1}^m w_{kj} x_j$$

$$y_k = \varphi(u_k + b_k)$$

Сигмоида  $f(S) = (1 + e^{-aS})^{-1}$

Гиперболический тангенс

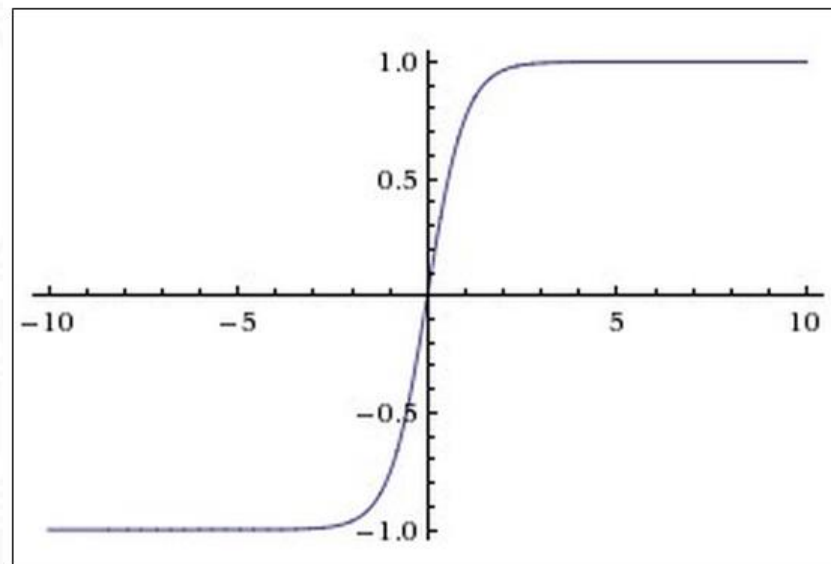
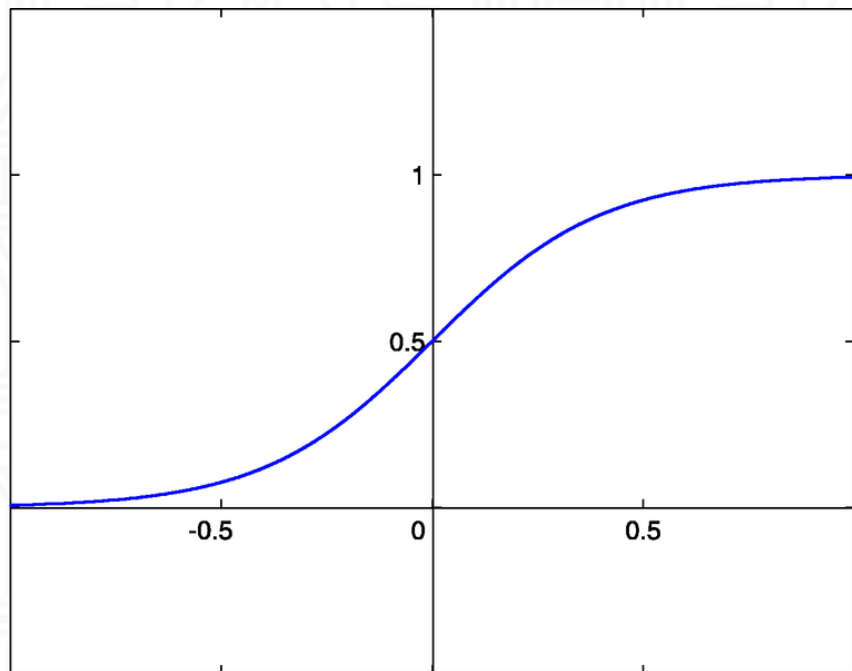
$$f(S) = \tanh(S) = \frac{e^S - e^{-S}}{e^S + e^{-S}}$$

ReLU  $f(S) = \max(0, S)$

Пороговая (Хевисайда)

$$f(S) = \begin{cases} 1, & S > T \\ 0, & \text{else} \end{cases}$$

## \* СИГМОИДА



## \* УНИВЕРСАЛЬНАЯ ТЕОРЕМА АППРОКСИМАЦИИ

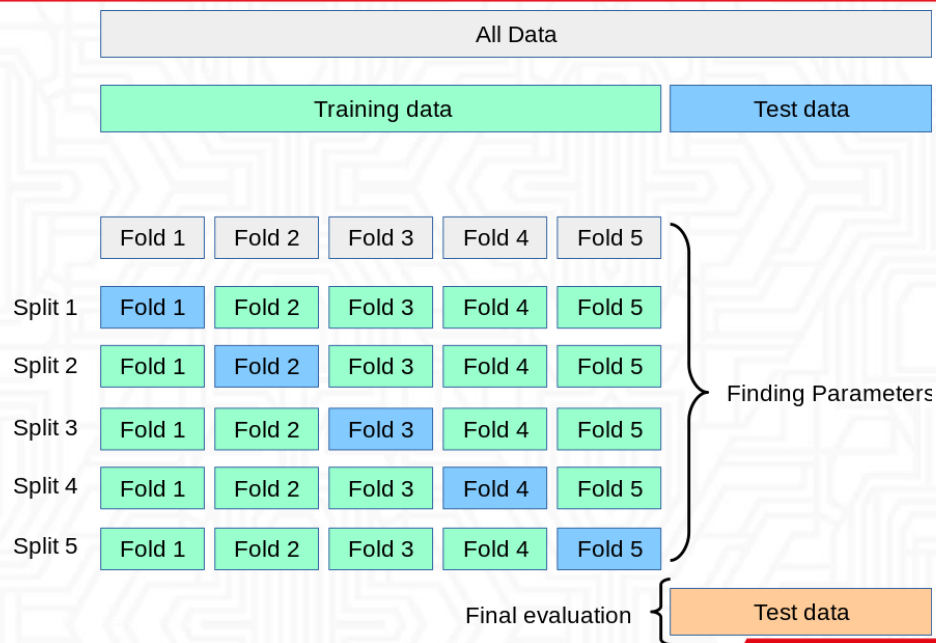
- \* Нейронная сеть прямого распространения с одним скрытым слоем может аппроксимировать любую непрерывную функцию многих переменных с любой точностью.

## \* КРОСС-ВАЛИДАЦИЯ (СКОЛЬЗЯЩИЙ КОНТРОЛЬ)

- \* Метод отложенных данных (holdout method) – разделение 70-30 или 60-40 или 80-20.
- \* Оценка ошибки близка к ошибке модели на новых данных, но сильно зашумлена.
- \* Для борьбы с шумом многократно случайно разделяют обучающую и тестовую выборку, параметр ошибки при этом усредняют.
- \* Но в процессе итераций каждая точка данных будет попадать в тестовое подмножество разное число раз, что может привести к смещению оценки.

## \* КРОСС-ВАЛИДАЦИЯ (СКОЛЬЗЯЩИЙ КОНТРОЛЬ)

- \* Контроль по k-блокам (k-fold cross-validation) - данные случайным образом делятся на k непересекающихся подмножеств (5, 10 или 20). После циклического перебора всех k подмножеств полученная оценка усредняется.





## \* ОЦЕНКА НЕЙРОСЕТЕЙ

Y=1	Y=-1	
True positive	False positive	A(x)=1
False negative	True negative	A(x)=-1

$$N = TP + FP + FN + TN$$

Полнота (способность обнаруживать данный класс, sensitivity)

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

Точность (способность отличать этот класс от других)

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

Достоверность (правильность)

$$accuracy = \frac{TP + TN}{N}$$

## \* ОЦЕНКА НЕЙРОСЕТЕЙ

- \* **Специфичность** (specificity), чаще всего применяется в медицинской статистике. Высокая специфичность позволяет отсеять людей, у которых действительно нет этого заболевания.

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

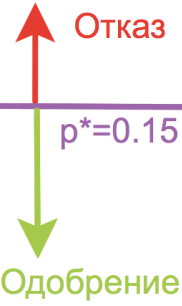
- \* **Коэффициент корреляции Мэтьюса** (в статистике известен как фи-коэффициент) может применяться как мера качества для бинарной классификации при высоком дисбалансе классов. Принимает значение в диапазоне  $[-1,1]$ , где 1 – идеальное предсказание, 0 – случайное предсказание, -1 – полное расхождение между предсказанием и наблюдением.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

## \* ROC-КРИВАЯ

- \* Вектора вероятностей классов, порог позволяет разделить классы, кривая строится для разных значений порога. Выбор порога обусловлен задачей, можно сдвинуть в сторону того или иного класса.
- \* Для каждого класса своя ROC-кривая при многоклассовой классификации

Клиент	Вероятность невозврата
Mike	0.78
Jack	0.45
Larry	0.13
Kate	0.06
William	0.03
Jessica	0.02

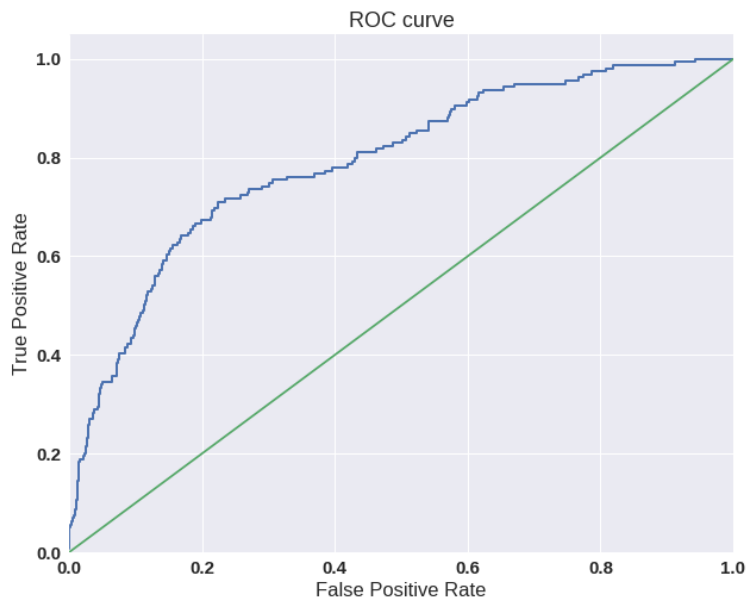


Отказ

$p^*=0.15$

Одобрение

## \* РОС-КРИВАЯ



$TPR = recall$

$FPR = \frac{FP}{FP+TN}$  - доля объектов negative класса, предсказанных неверно

Каждая точка на кривой соответствует значению порога

## \* МЕТРИКИ РЕГРЕССИИ

\* Средняя абсолютная ошибка (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |a(x_i) - y_i|$$

## \* МЕТРИКИ РЕГРЕССИИ

- \* Среднеквадратическая ошибка (MSE) применяется когда надо подчеркнуть большие ошибки, но поэтому она более чувствительна к выбросам, чем MAE.
- \* Чем ошибка меньше, тем лучше, но важно учитывать масштаб данных. Чтобы MSE имел размерность исходных данных, из него извлекают корень и получают RMSE, что затрудняет сравнение на разных наборах.

$$MSE = \frac{1}{n} \sum_{i=1}^n (a(x_i) - y_i)^2$$

$$RMSE = \sqrt{MSE}$$

## \* МЕТРИКИ РЕГРЕССИИ

- \* Коэффициент детерминации ( $R^2$ ) – доля дисперсии, объясненная моделью, в общей дисперсии целевой переменной. Чем ближе к 1, тем модель лучше объясняет данные. Модели с коэффициентов детерминации больше 0,8 можно считать хорошими.  $\bar{y}$  – среднее арифметическое  $y_i$ .

$$R^2 = 1 - \frac{\sum_{i=1}^n (a(x_i) - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$



## \* МЕТРИКИ РЕГРЕССИИ

- \* SMAPE (Symmetric Mean Absolute Percentage Error) рассматривает не абсолютные, а относительные ошибки на объектах

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{2|a(x_i) - y_i|}{y_i + a(x_i)}$$