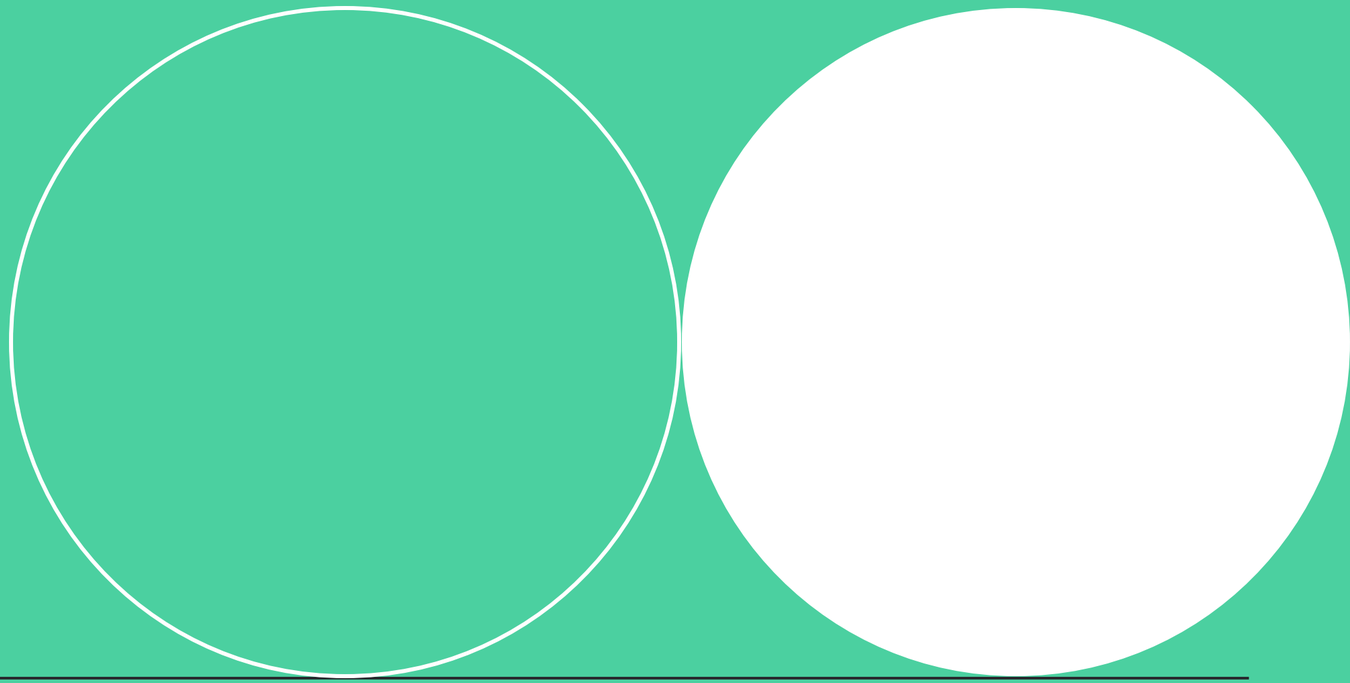
# Линейный классификатор и логистическая регрессия

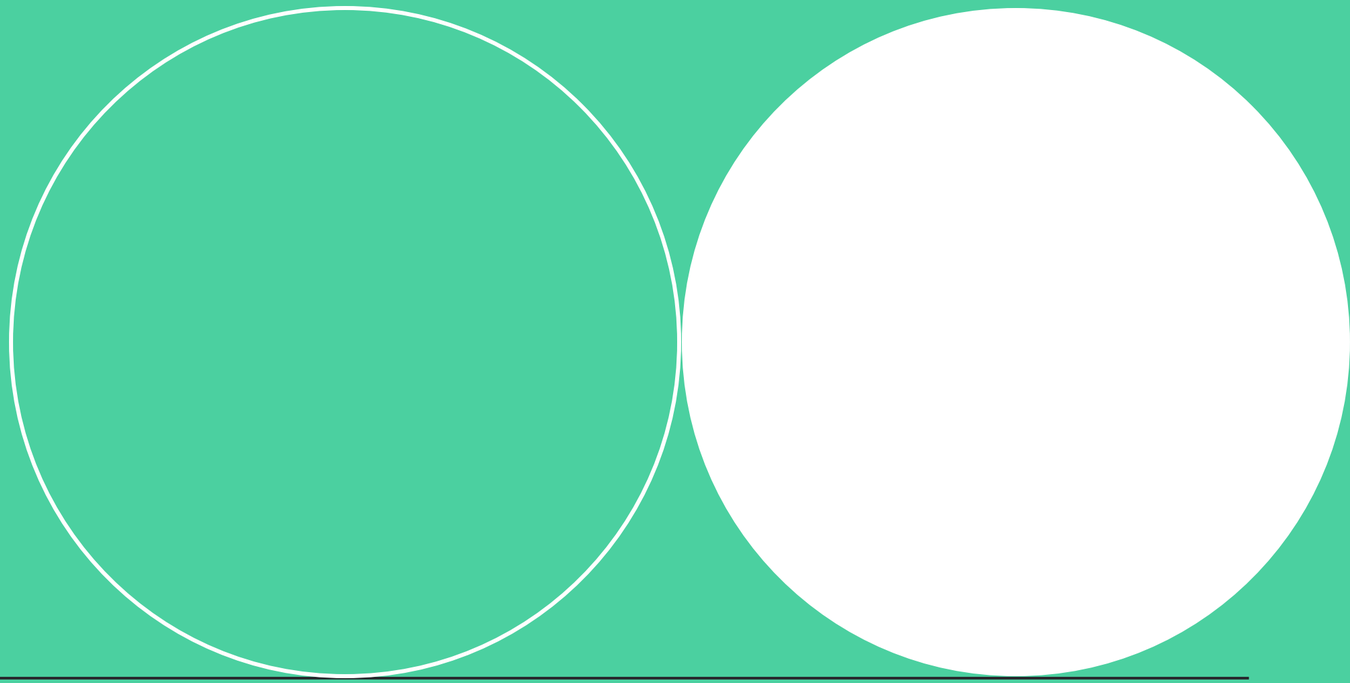
# Цели занятия



нетология

- Рассмотреть задачу классификации
- Рассмотреть линейный классификатор
- Рассмотреть варианты линейного классификатора. Логистическую регрессию и SVM.

# О чем поговорим и что сделаем



нетология

# План занятия

1. Задача классификации
2. Логистическая регрессия: практическое задание
3. SVM
4. Практика

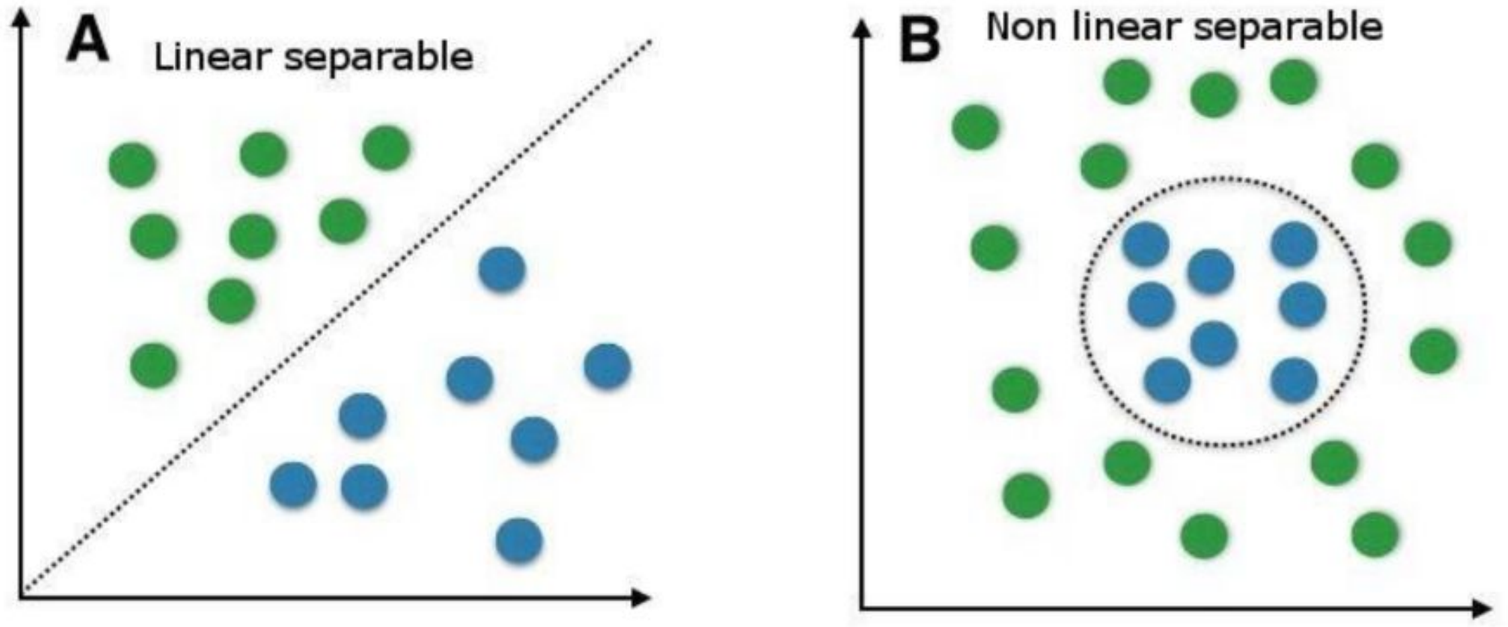
# Классификация

1

нетология

**Классификации** – задача *предсказания ответа* из ***конечного*** множества вариантов.

Линейный классификатор - решает задачу разделения признакового пространства на две части, в каждом из которых находиться свой класс.

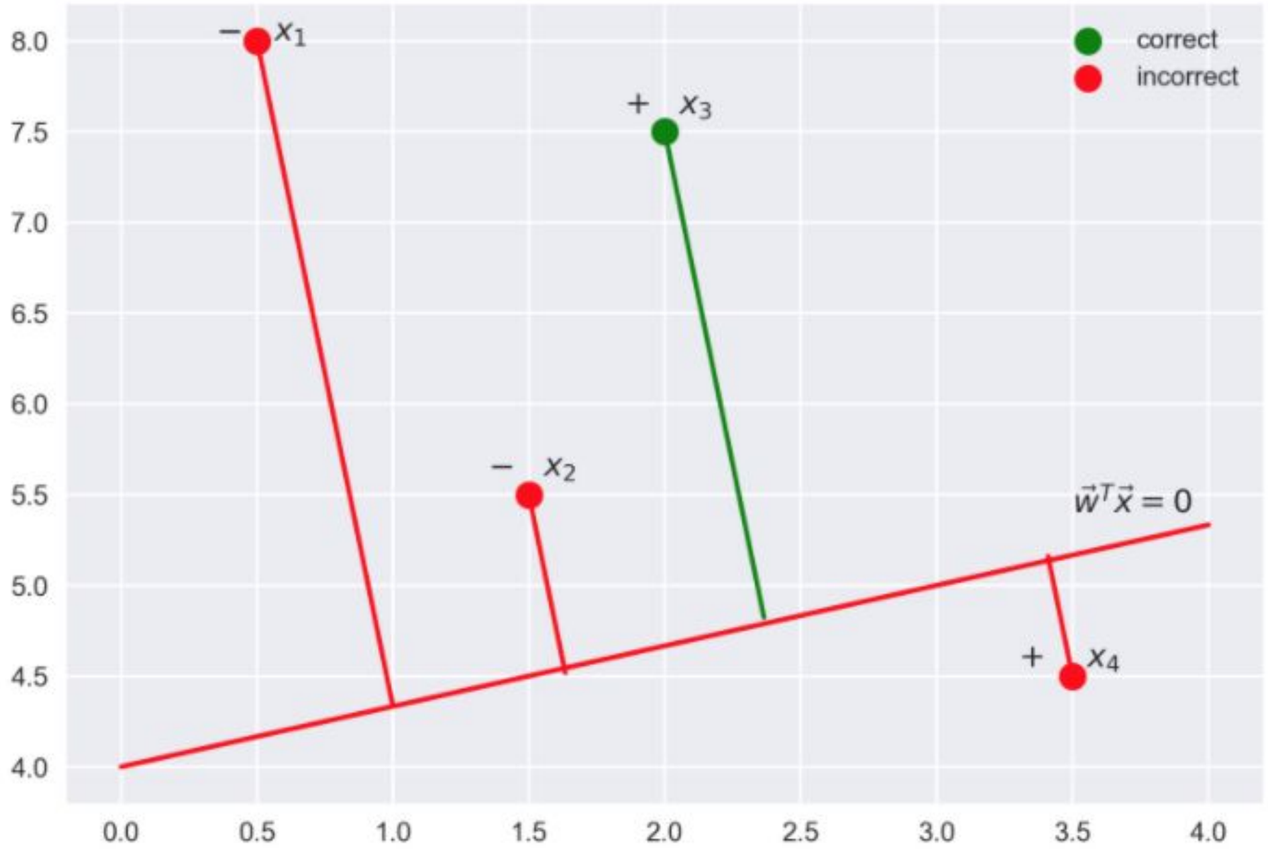



Линейная разделимость данных

Отступ - расстояние от разделяющей гиперповерхности до объекта

$$M(\overrightarrow{x_i}) = y_i \vec{w}^T \overrightarrow{x_i}$$

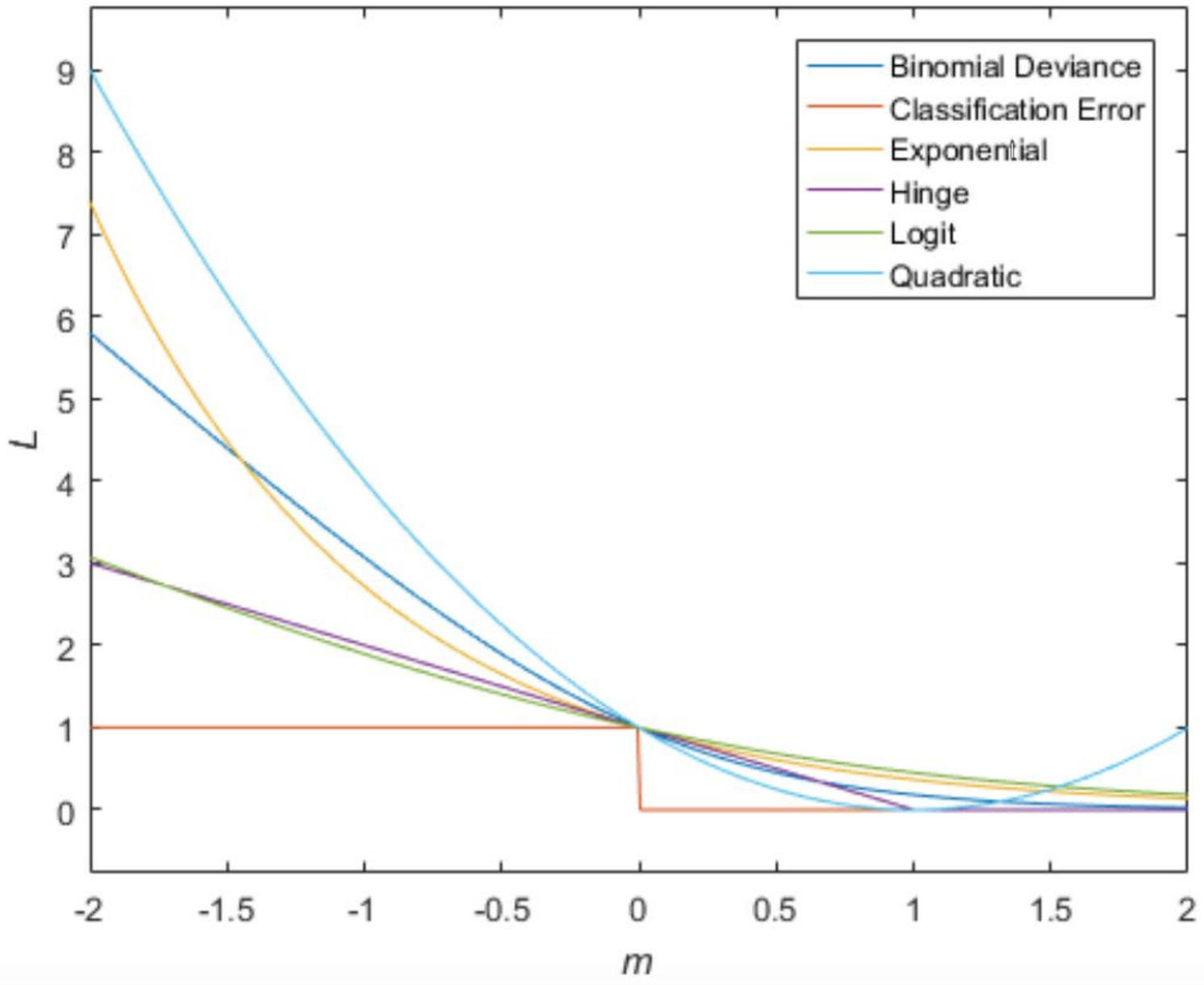
Отступ можно понимать как «степень погруженности» объекта в свой класс.




https://habr.com/ru/company/ods/blog/323890/
https://dyakonov.org/логистическая-функция-ошибки/#more-6139

Функция потерь *L* как функция от отступа *M*

$$L = f(M)$$



# Логистическая регрессия

4

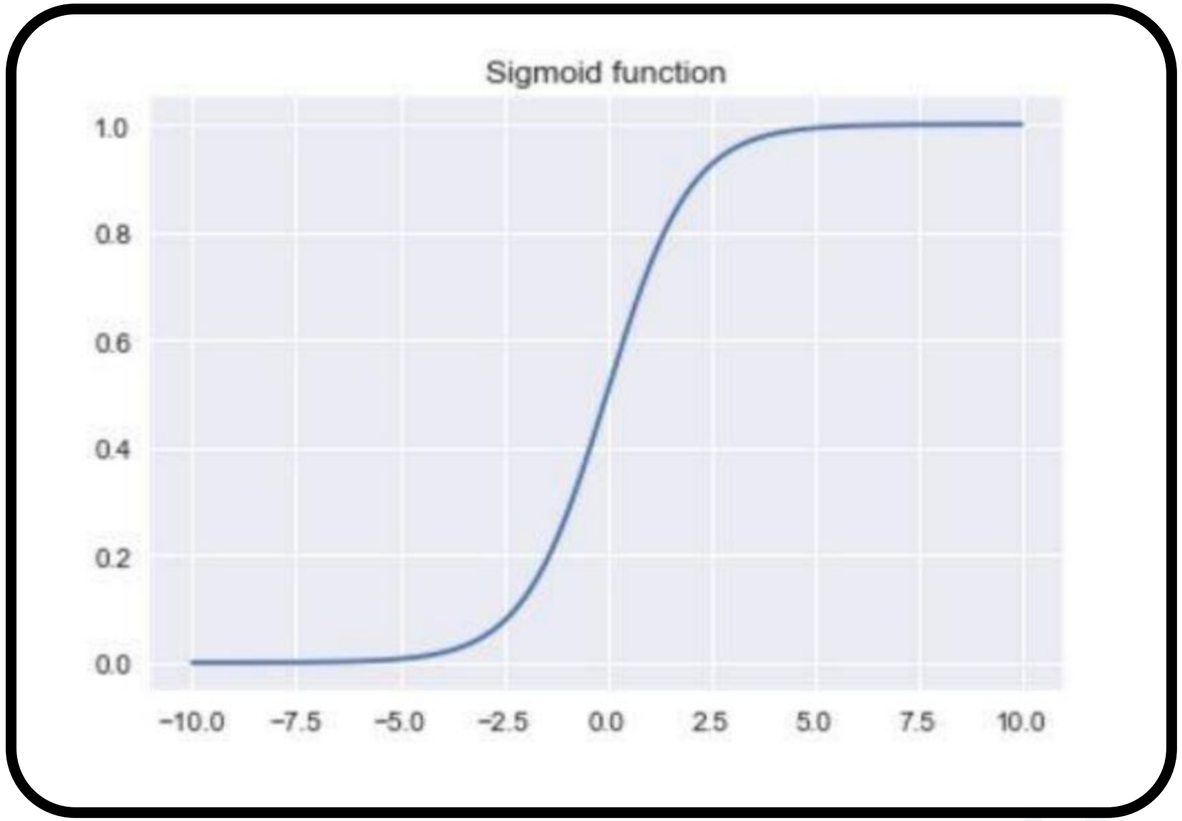
нетология

# Логистическая регрессия

**Логистическая регрессия -** линейный классификатор, позволяющий ***оценивать вероятности*** принадлежности объектов классам.

$$L = a_0 + a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$$

$$p = \frac{1}{1 + e^{-L}}$$




https://habr.com/ru/post/485872/
https://habr.com/ru/company/ods/blog/323890/
https://dyakonov.org/логистическая-функция-ошибки/#more-6139

# Функция потерь

Модель предсказывает вероятность классов {0, +1}

$$p(y_i \mid x_i, w) = a_i^{y_i}(1 - a_i)^{1-y_i}$$

Максимизировать правдоподобие

$$p(y \mid X, w) = \prod_i p(y_i \mid x_i, w)$$

Функция потерь

$$\mathcal{L}_{log}(X, \vec{y}, \vec{w}) = \sum_i (-y_i \log a_i - (1 - y_i)\log(1 - a_i))$$

https://habr.com/ru/post/485872/
https://habr.com/ru/company/ods/blog/323890/
https://dyakonov.org/логистическая-функция-ошибки/#more-6139

# Функция потерь

Модель предсказывает вероятность классов {-1, +1}

$$P\left(y = y_i \mid \overrightarrow{x_i}, \vec{w}\right) = \sigma(y_i \vec{w}^T \overrightarrow{x_i})$$

Максимизировать правдоподобие

$$P\left(\vec{y} \mid X, \vec{w}\right) = \prod_{i=1}^{\ell} P\left(y = y_i \mid \overrightarrow{x_i}, \vec{w}\right)$$

$$\log P\left(\vec{y} \mid X, \vec{w}\right) = -\sum_{i=1}^{\ell} \log(1 + \exp^{-y_i \vec{w}^T \overrightarrow{x_i}})$$

Функция потерь

$$\mathcal{L}_{log}(X, \vec{y}, \vec{w}) = \sum_{i=1}^{\ell} \log(1 + \exp^{-y_i \vec{w}^T \overrightarrow{x_i}})$$

https://habr.com/ru/post/485872/
https://habr.com/ru/company/ods/blog/323890/
https://dyakonov.org/логистическая-функция-ошибки/#more-6139

# SVM

7

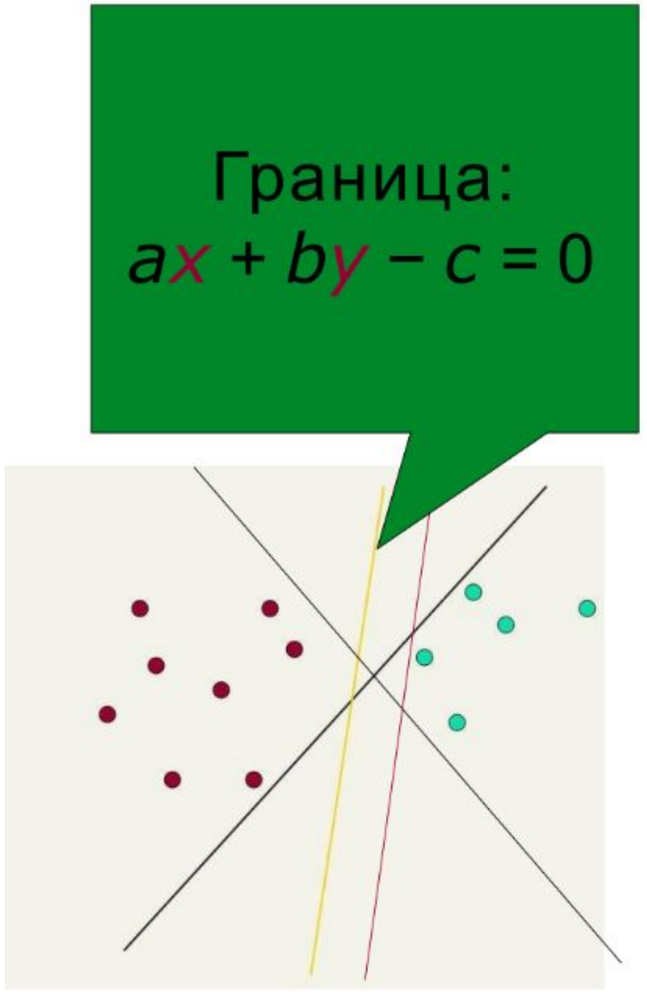
нетология

# Множество гиперплоскостей

Множество решений для *a,b,c.*

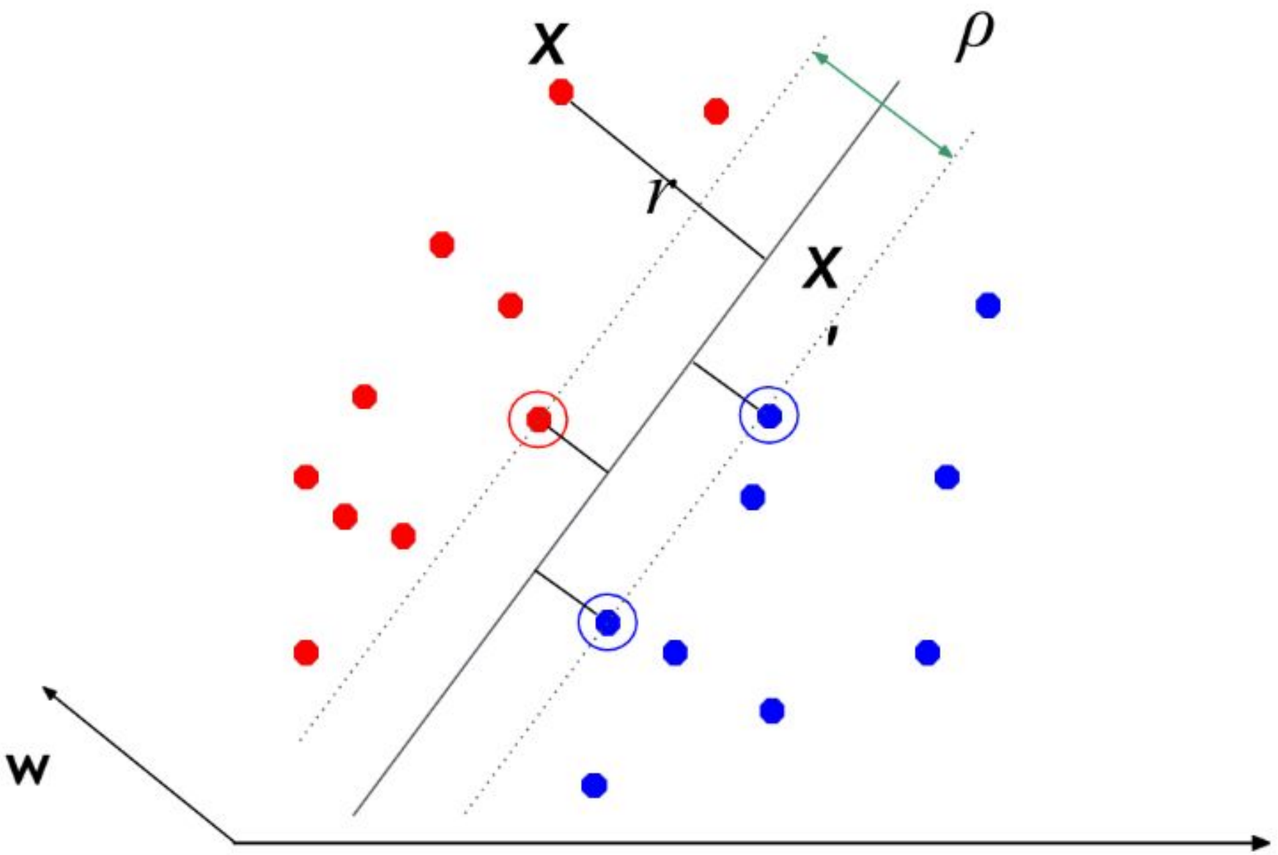
SVM находит оптимальную разделяющую поверхность

Максимизирует «зазор»



# Максимальный зазор

- w – нормаль к разделяющей плоскости
- $x_i$ - sample
- $y_i$: - класс sample i {+1, -1} (важно, не {0, 1})
- Классификатор: $f(x_i) = \text{sign}( \mathbf{w}^T\mathbf{x}+b )$
- Зазор для точки x $r = y\frac{\mathbf{w}^T\mathbf{x}+b}{\|\mathbf{w}\|}$
- Зазор всего датасета – минимум зазора для всех точек



# Формула

Итого получаем задачу оптимизации:

Найти $\mathbf{w}$ и $b$ такие что

максимально; и для всех $\{(\mathbf{x_i}, y_i)\}$

$\mathbf{w^T x_i} + b \geq 1$ если $y_i = 1$; $\mathbf{w^T x_i} + b \leq -1$ если $y_i = -1$

Перепишем в более понятном виде:

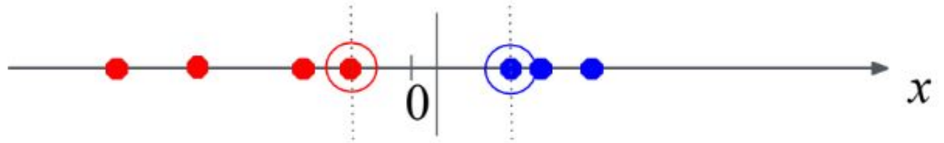
Найти $\mathbf{w}$ и $b$ такие что

$\Phi(\mathbf{w}) = 0.5\, \mathbf{w}^T\mathbf{w}$ максимально

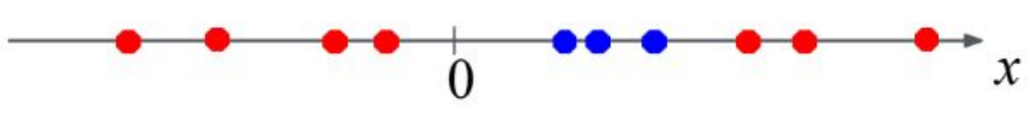
И для всех $\{(\mathbf{x_i}, y_i)\}$: $y_i(\mathbf{w^T x_i} + b) \geq 1$

# Non-linear SVMs

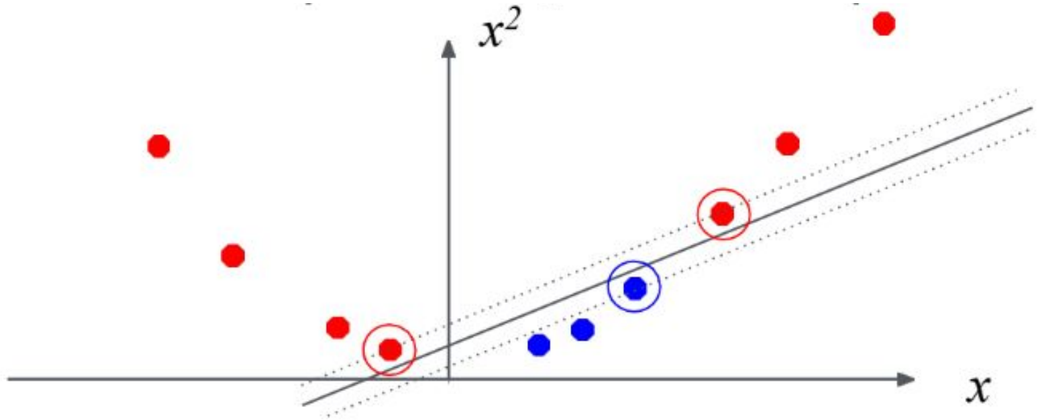
Линейно разделимые датасеты хорошо классифицируются




Но что делать, если они не линейно разделимы?




Можно попробовать отобразить данные в пр-во более высокой размерности



# The «Kernel Trick»

- SVM зависит от скалярного произведения $K(x_i, x_j) = x_i^T x_j$
- Если каждая точка отображается в пр-во более высокой размерности при помощи Φ: x ➔ φ(x), тогда скалярное произведение становится:
- $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$
- Функция ядра - это функция, соответствующая скалярному произведению в пр-ве более высокой размерности

# Kernels

- Полиномиальное $k(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}')^d$
- Полиномиальное со смещением $k(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}' + 1)^d$

# Kernels

- Радиальная базисная функция $k(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2)$, для $\gamma > 0$
- Радиальная базисная функция Гаусса $k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right)$

# Практика

# Что мы сегодня узнали

8

нетология

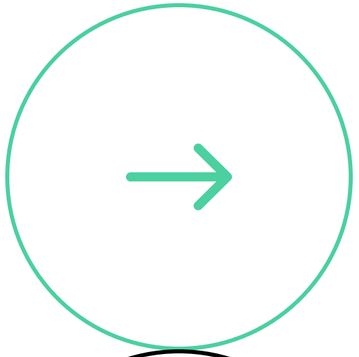
# Итоги занятия

- Вспомнили основы теории вероятностей
- Изучили линейные модели и требования к ним на основе функции правдоподобия
- Реализовали логистическую регрессию
- Изучили алгоритм градиентного спуска и потренировались в его реализации

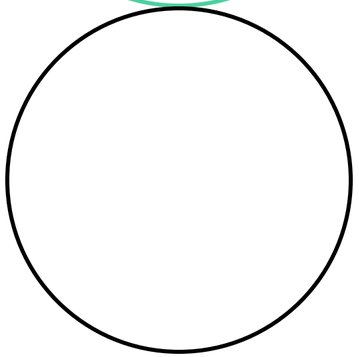
# Полезные материалы

9

нетология



Статья о линейных моделях в ODS
https://habrahabr.ru/company/ods/blog/323890/



Курс «Основы статистики» на Stepik.org
https:/stepik.org/course/Основы-статистики-76

# Спасибо за внимание!

fb.com/

mail@mail.com

нетология