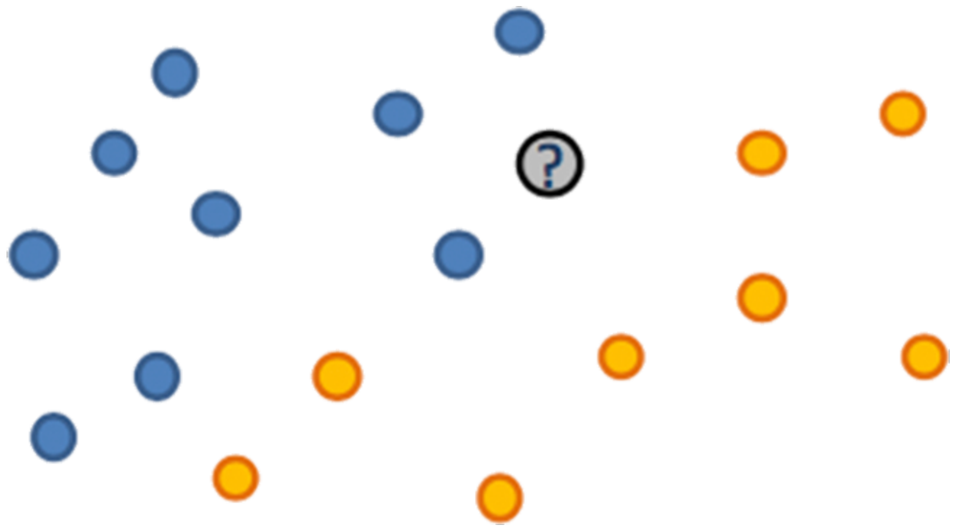
МЕТОД К БЛИЖАЙШИХ СОСЕДЕЙ

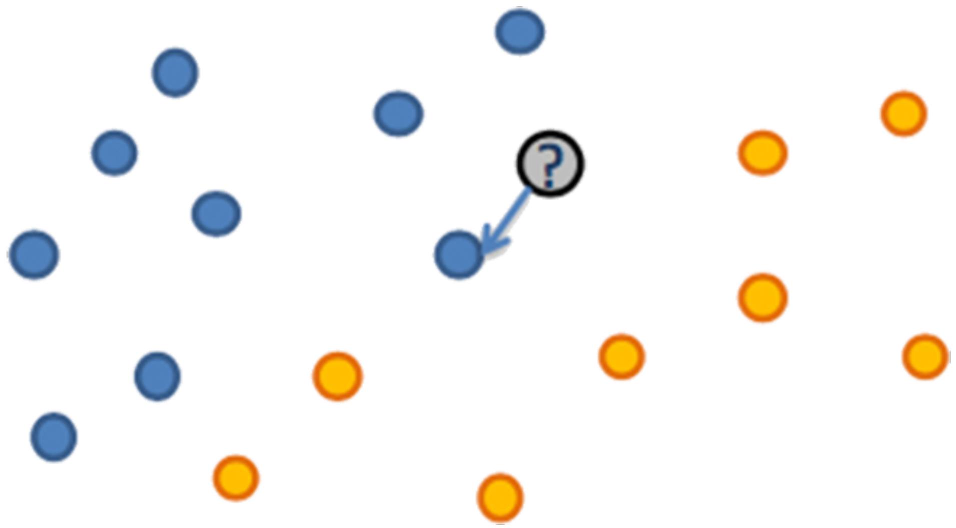
МЕТРИЧЕСКИЕ АЛГОРИТМЫ

- Метрические алгоритмы: в пространстве признаков введено понятие метрики
- Начнём с метода ближайшего соседа в задаче классификации

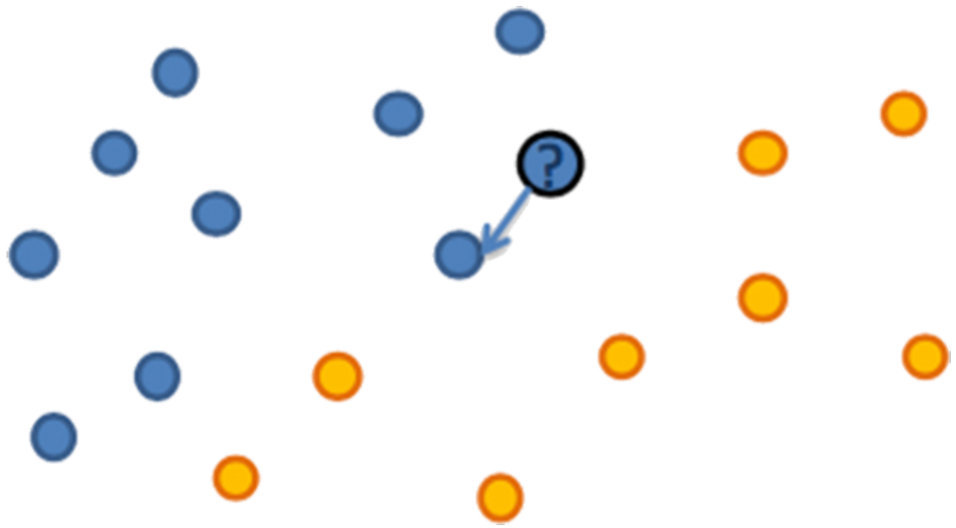
МЕТОД БЛИЖАЙШЕГО СОСЕДА



МЕТОД БЛИЖАЙШЕГО СОСЕДА

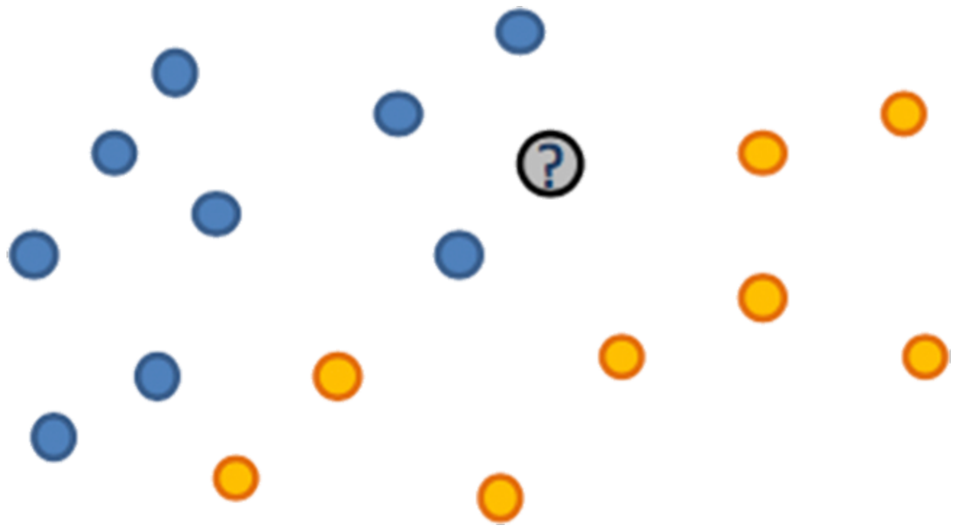


МЕТОД БЛИЖАЙШЕГО СОСЕДА



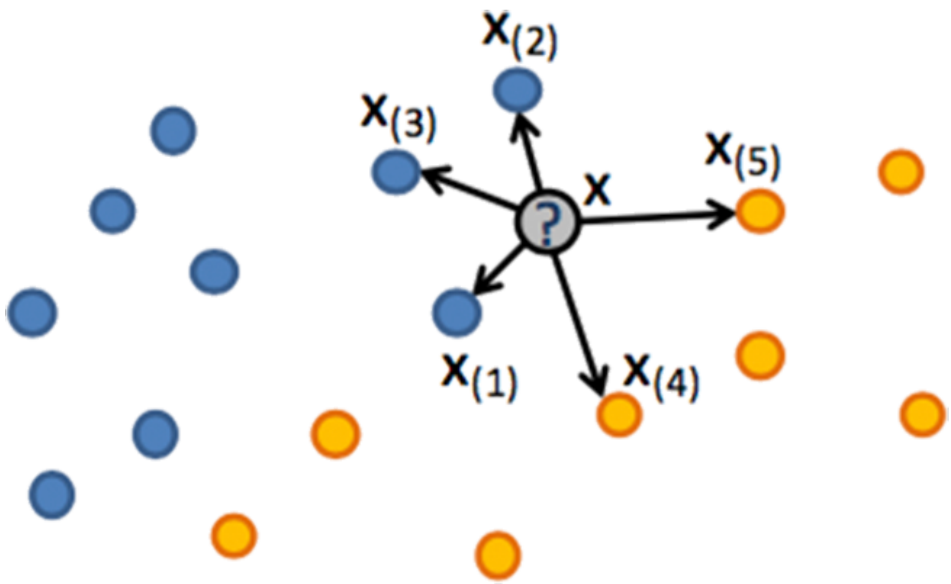
МЕТОД К БЛИЖАЙШИХ СОСЕДЕЙ (KNN)

Пример классификации ($k = 5$)



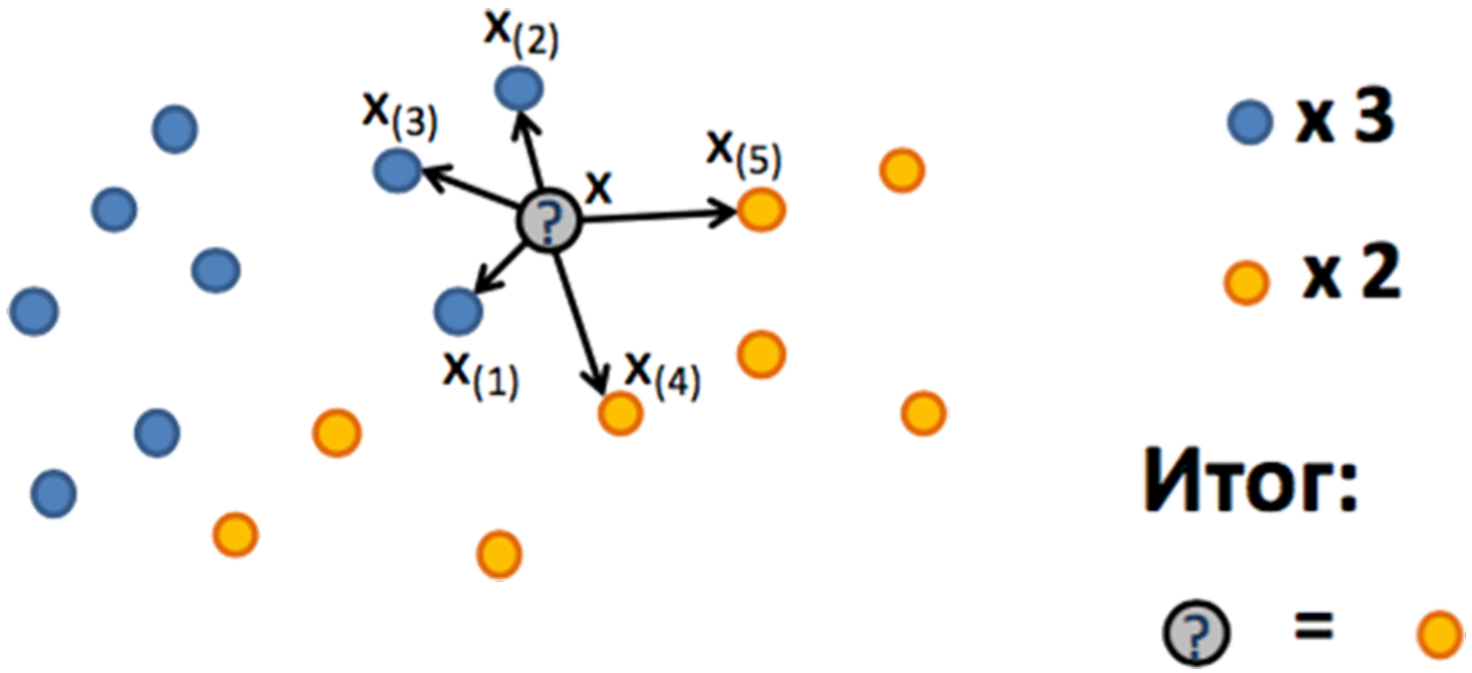
Метод к ближайших соседей (KNN)

Пример классификации ($k = 5$)



МЕТОД K БЛИЖАЙШИХ СОСЕДЕЙ (KNN)

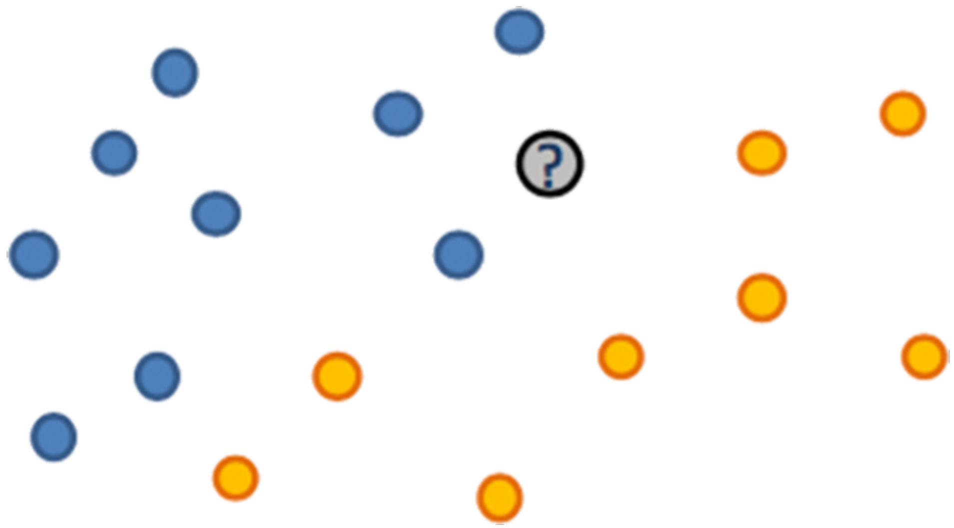
Пример классификации ($k = 5$)



МФТИ

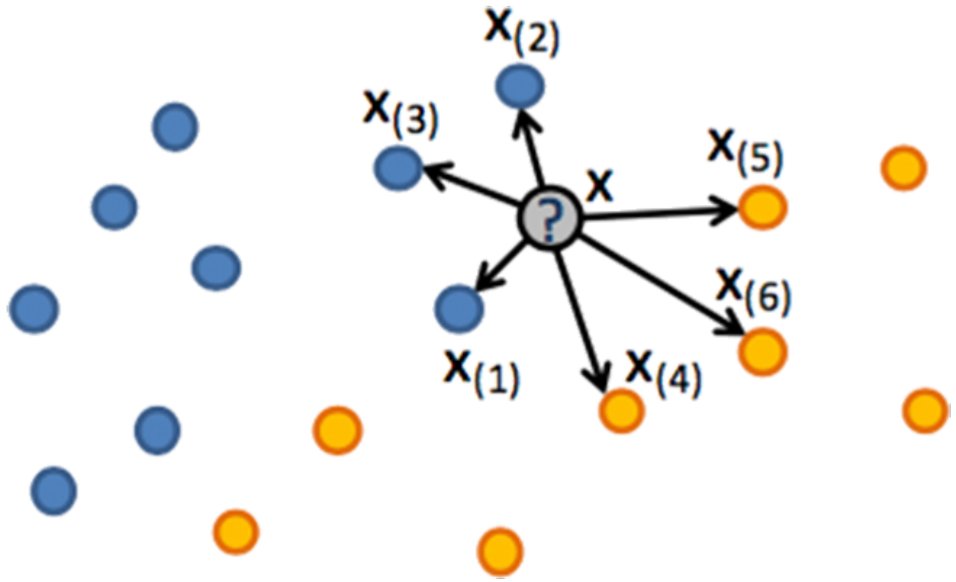
ВЗВЕШЕННЫЙ KNN

Пример классификации ($k = 6$)



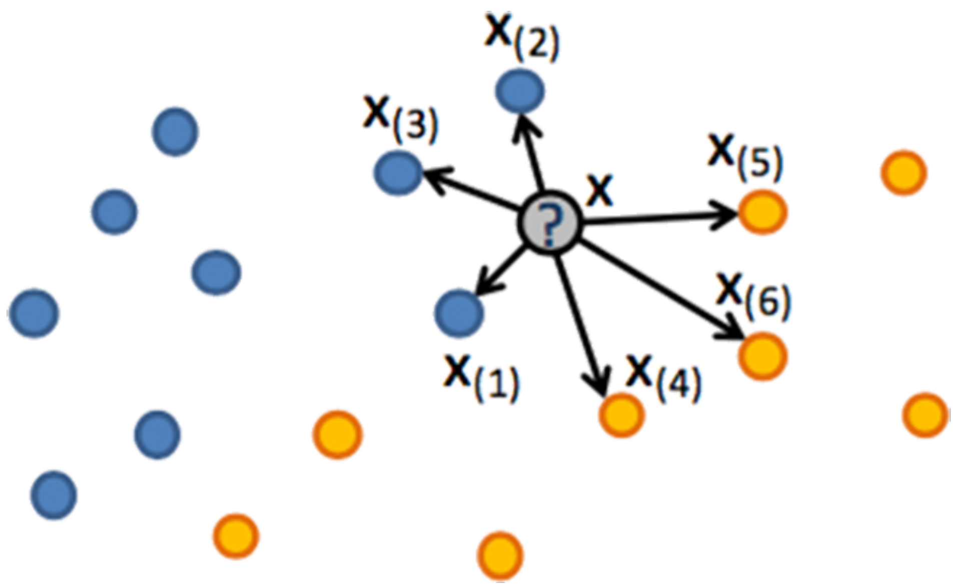
ВЗВЕШЕННЫЙ KNN

Пример классификации ($k = 6$)



ВЗВЕШЕННЫЙ KNN

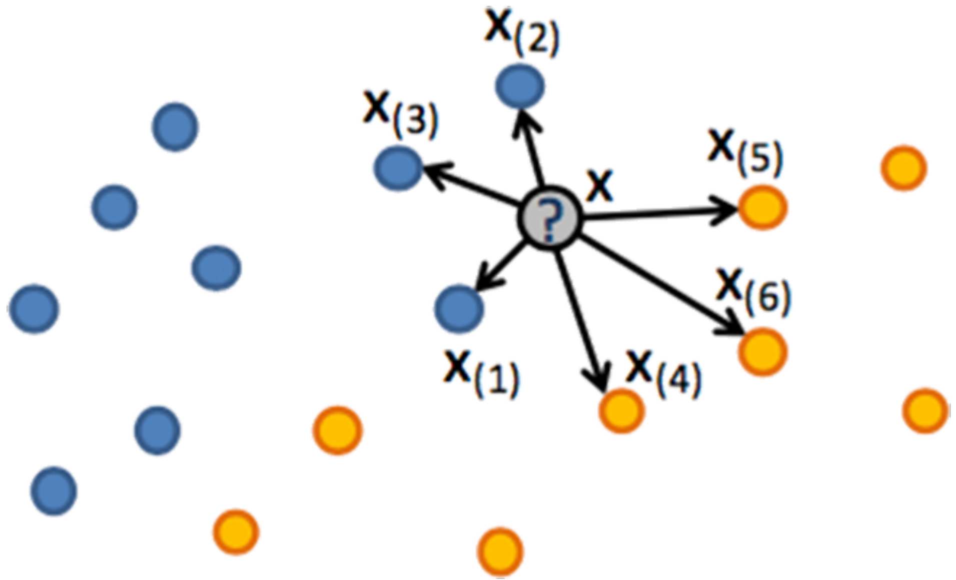
Пример классификации ($\boldsymbol{k = 6}$)



Веса можно определять как функцию от соседа или его номера: $\mathbf{w}(x_{(i)}) = \mathbf{w}(i)$

ВЗВЕШЕННЫЙ KNN

Пример классификации ($k = 6$)



Веса можно определять как функцию от соседа или его номера: $\mathbf{w}(x_{(i)}) = \mathbf{w}(i)$

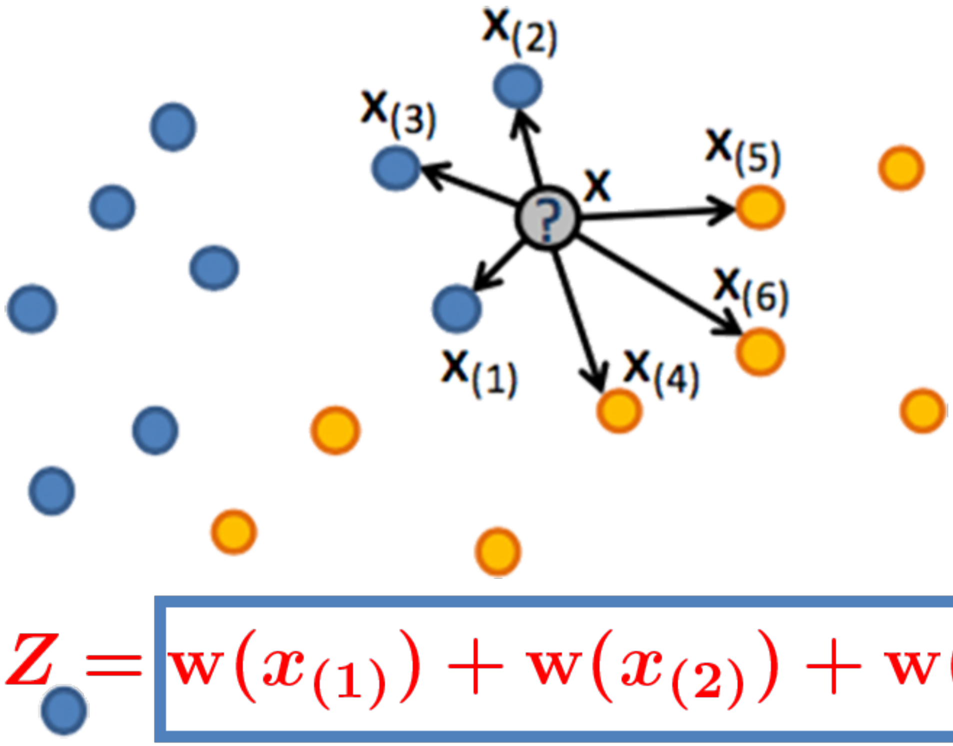
Или как функцию расстояния:

$$\mathbf{w}(x(i)) = \mathbf{w}(d(x, x_{(i)}))$$

МФТИ

ВЗВЕШЕННЫЙ KNN

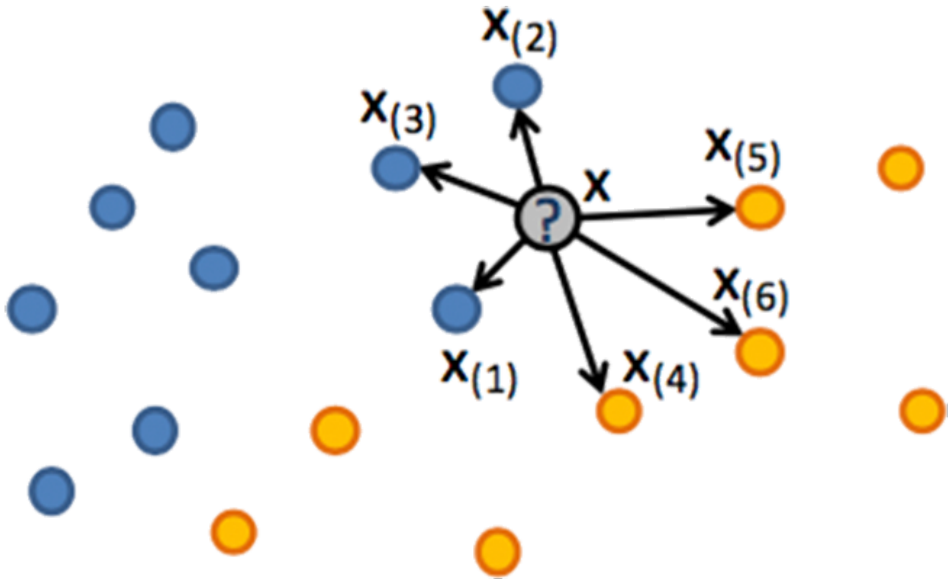
Пример классификации ($k = 6$)



$$Z = \boxed{\mathbf{w}(x_{(1)}) + \mathbf{w}(x_{(2)}) + \mathbf{w}(x_{(3)})}$$

ВЗВЕШЕННЫЙ KNN

Пример классификации ($\boldsymbol{k = 6}$)

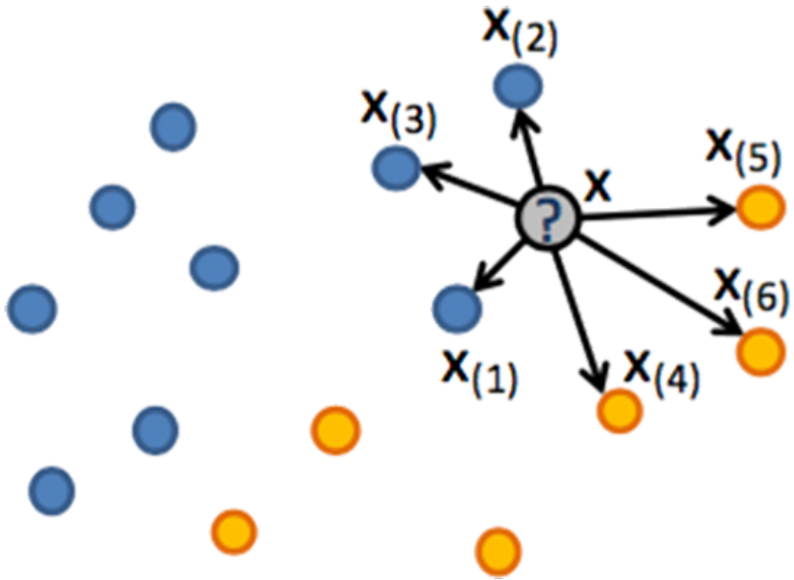


$$Z = \boxed{\mathbf{w}(x_{(1)}) + \mathbf{w}(x_{(2)}) + \mathbf{w}(x_{(3)})}$$

$$Z = \boxed{\mathbf{w}(x_{(4)}) + \mathbf{w}(x_{(5)}) + \mathbf{w}(x_{(6)})}$$

ВЗВЕШЕННЫЙ KNN

Пример классификации ($k = 6$)



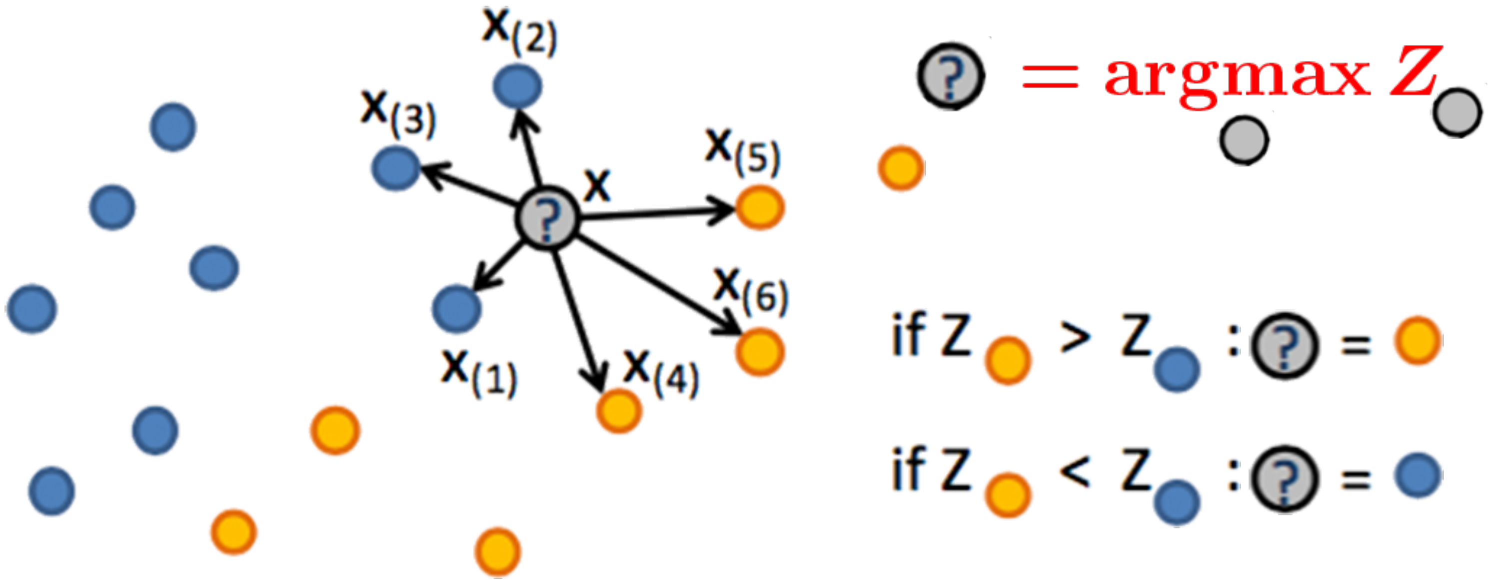
$$? = \operatorname{argmax} Z$$

$$Z = \boxed{\mathrm{w}(x_{(1)}) + \mathrm{w}(x_{(2)}) + \mathrm{w}(x_{(3)})}$$

$$Z = \boxed{\mathrm{w}(x_{(4)}) + \mathrm{w}(x_{(5)}) + \mathrm{w}(x_{(6)})}$$

ВЗВЕШЕННЫЙ KNN

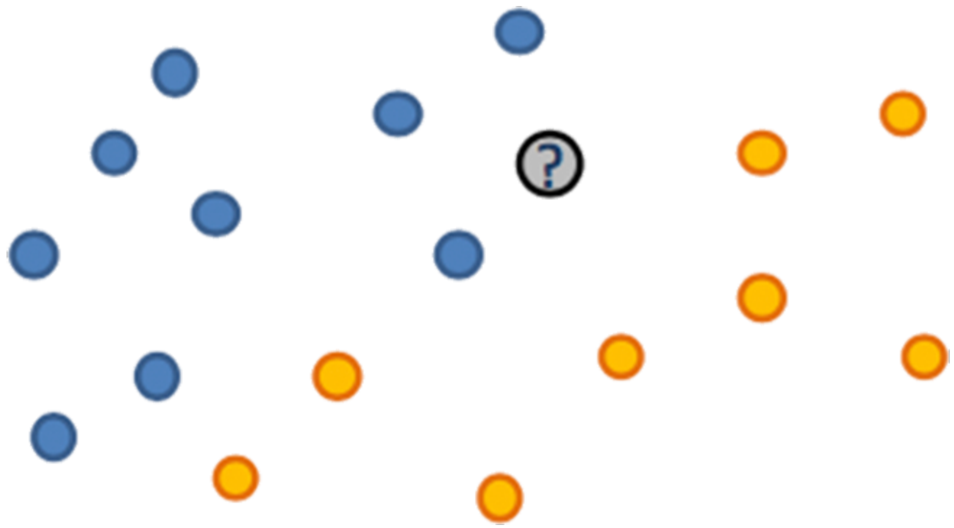
Пример классификации ($k = 6$)



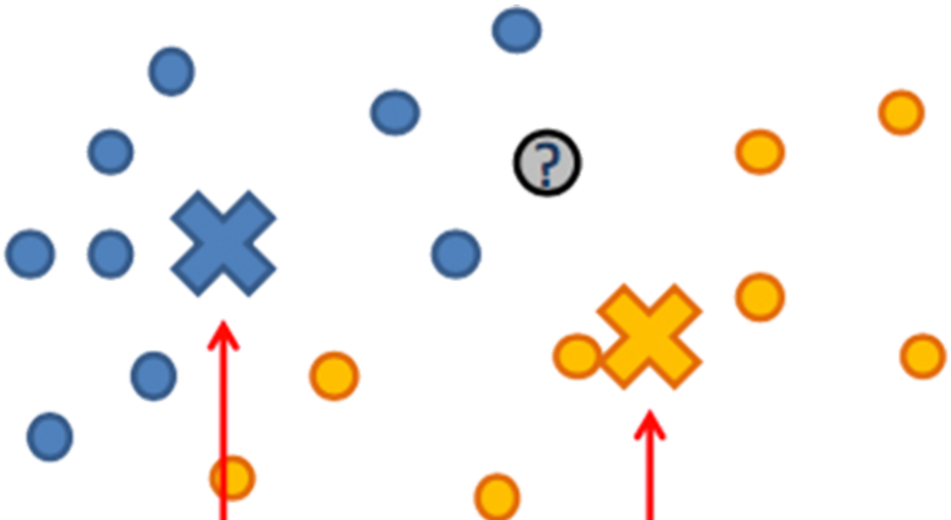
$$Z = \boxed{\mathrm{w}(x_{(1)}) + \mathrm{w}(x_{(2)}) + \mathrm{w}(x_{(3)})}$$

$$Z = \boxed{\mathrm{w}(x_{(4)}) + \mathrm{w}(x_{(5)}) + \mathrm{w}(x_{(6)})}$$

ЦЕНТРОИДНЫЙ КЛАССИФИКАТОР



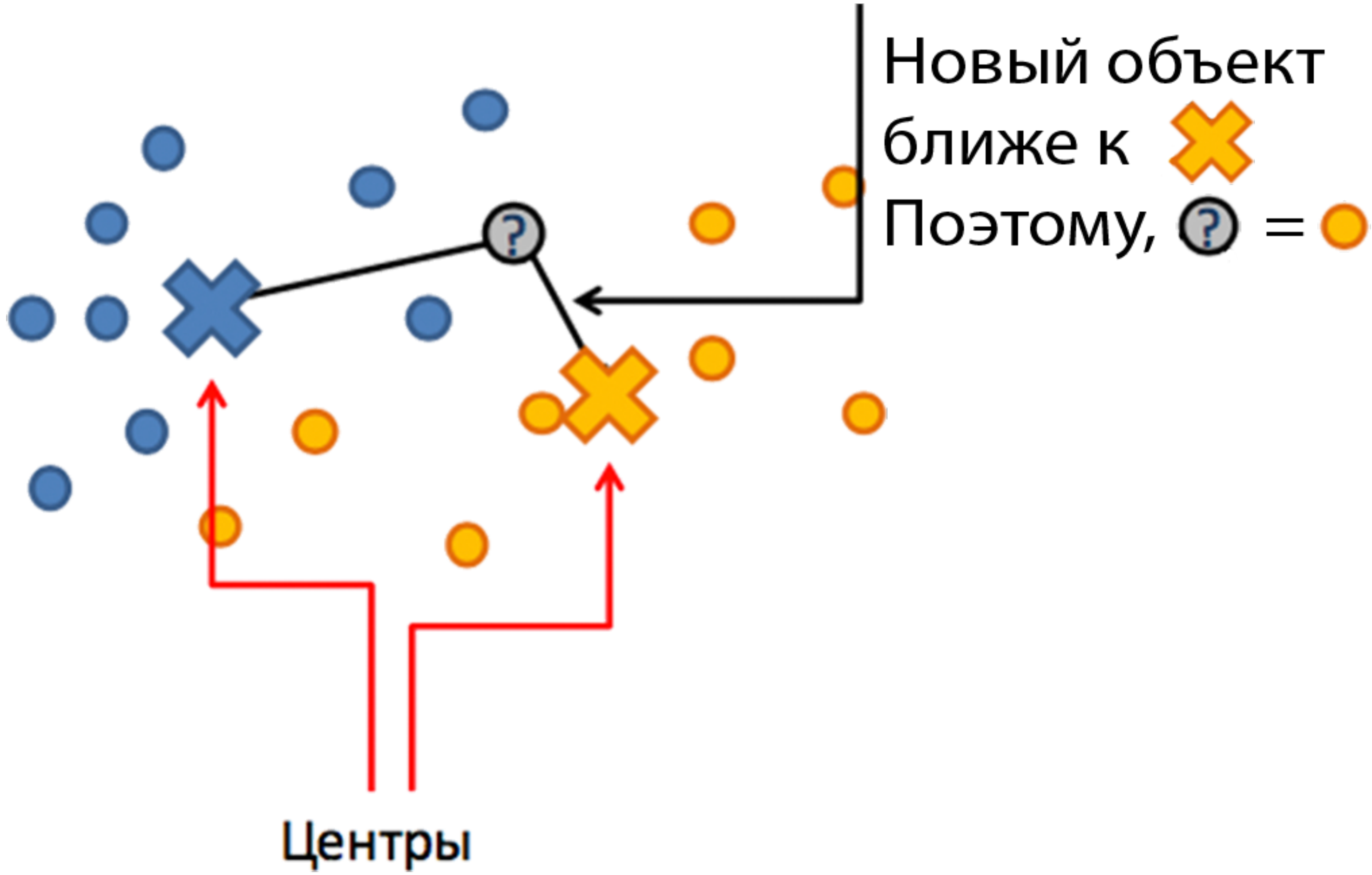
ЦЕНТРОИДНЫЙ КЛАССИФИКАТОР



$$\mu_{-1} = \frac{1}{l_{-1}} \sum_{i:y_i=-1} x_i \qquad \mu_{+1} = \frac{1}{l_{+1}} \sum_{i:y_i=+1} x_i$$

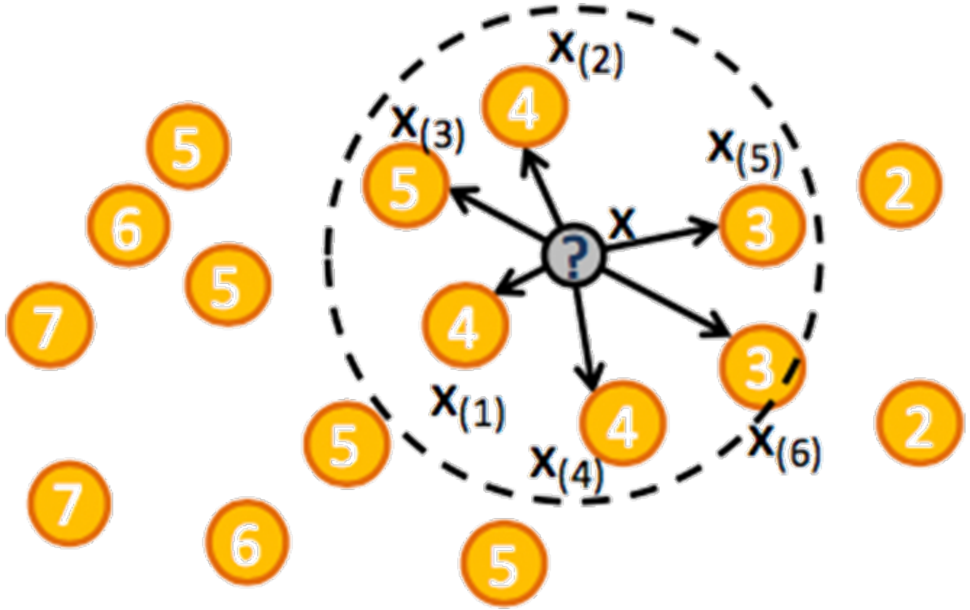
Центры

ЦЕНТРОИДНЫЙ КЛАССИФИКАТОР



Взвешенный KNN для регрессии

Пример классификации ($k = 6$)



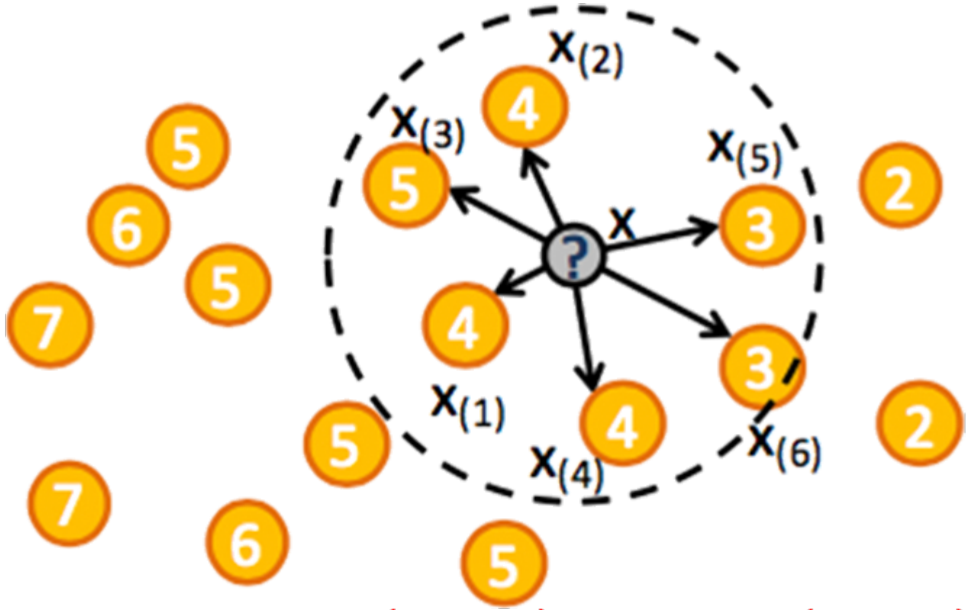
Веса можно определять как функцию от соседа или его номера: $\mathbf{w}(x_{(i)}) = \mathbf{w}(i)$

Или как функцию расстояния:

$$\mathbf{w}(x(i)) = \mathbf{w}(d(x, x_{(i)}))$$

ВЗВЕШЕННЫЙ KNN ДЛЯ РЕГРЕССИИ

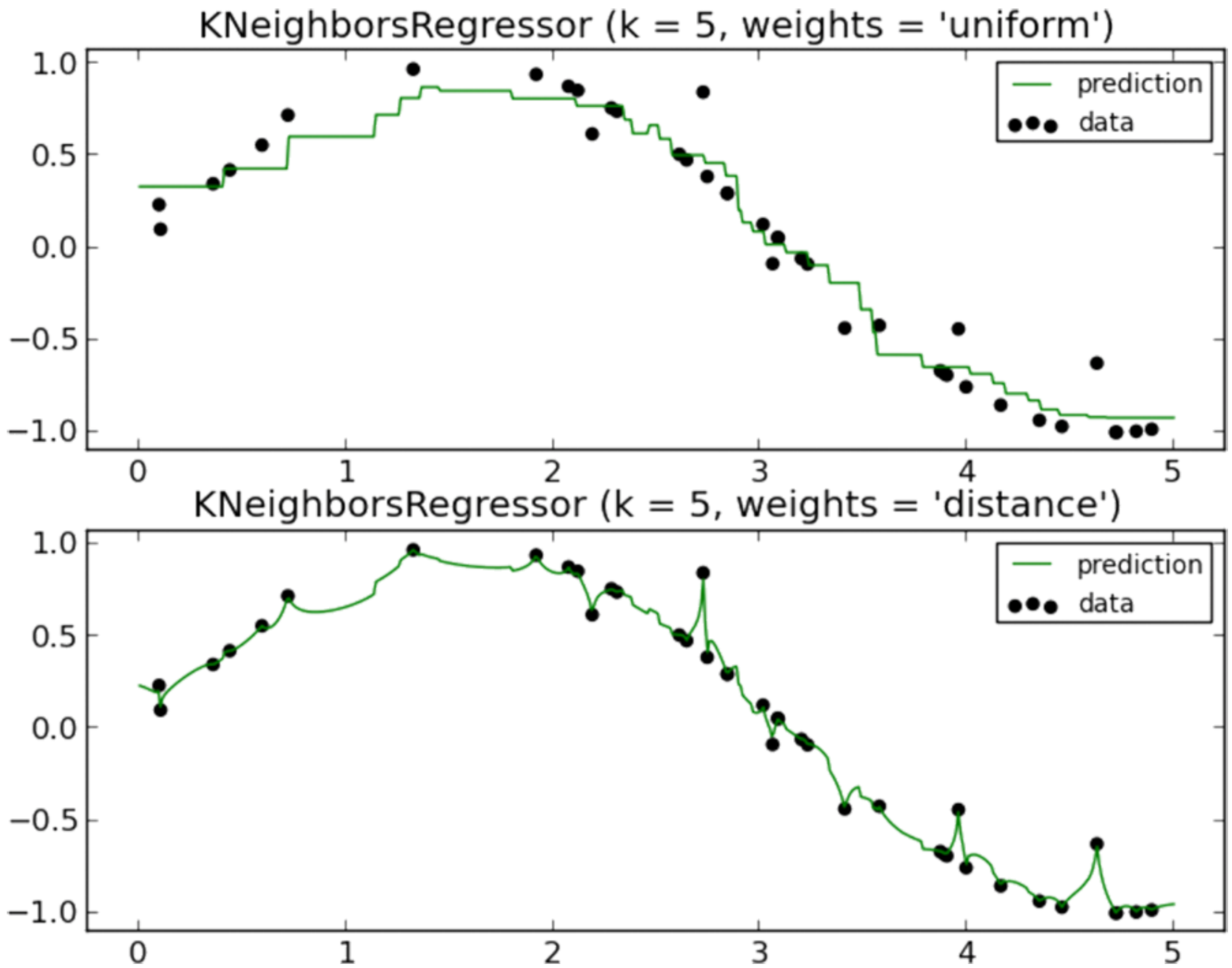
Пример классификации ($k = 6$)



$$? = \frac{4\,w(x_{(1)}) + 4\,w(x_{(2)}) + 5\,w(x_{(3)}) + 4\,w(x_{(4)}) + 3\,w(x_{(5)}) + 3\,w(x_{(6)})}{w(x_{(1)}) + w(x_{(2)}) + w(x_{(3)}) + w(x_{(4)}) + w(x_{(5)}) + w(x_{(6)})}$$

МФТИ

ВЫБОР ВЕСОВ В KNN



ВЫБОР ВЕСОВ В KNN

- Метод k ближайших соседей
- Веса соседей
- kNN в задаче классификации и в задаче регрессии

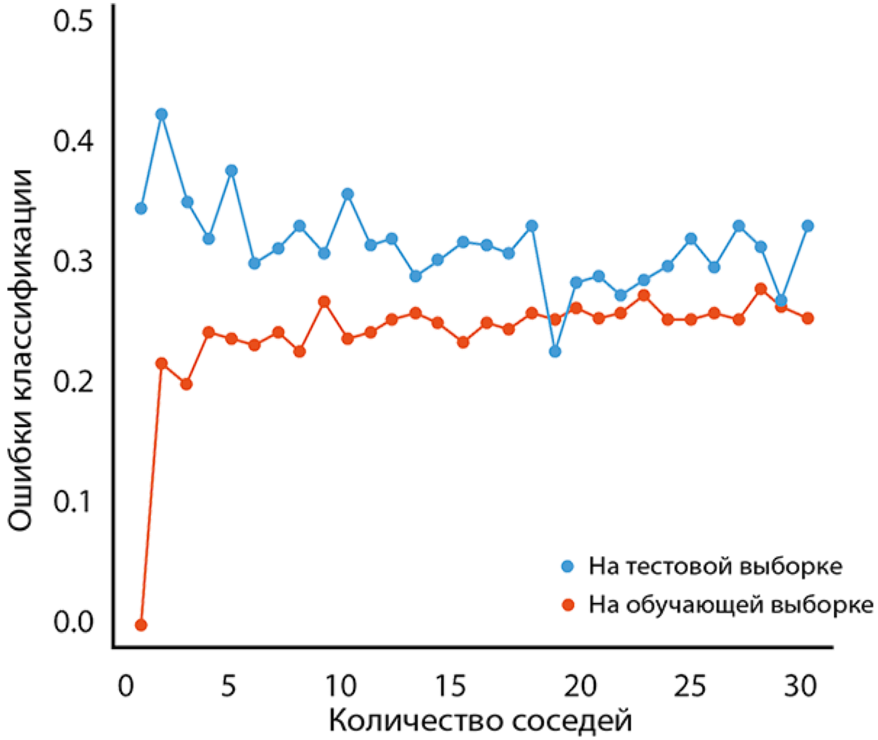
НАСТРОЙКА ПАРАМЕТРОВ В kNN

НАСТРОЙКА ПАРАМЕТРОВ В kNN

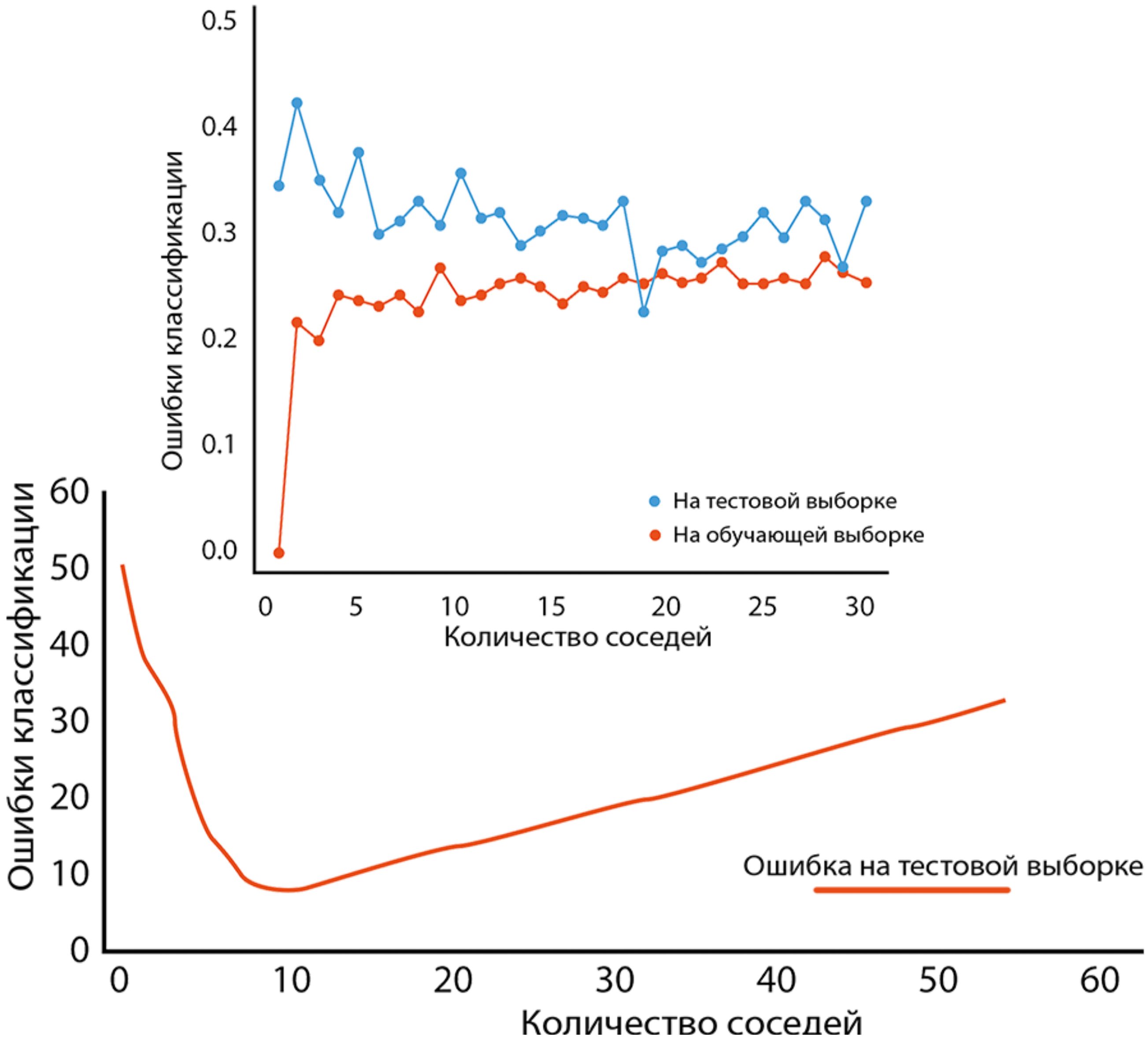
- У kNN нет параметров, настраиваемых на обучающей выборке
- Но есть параметры, выбираемые исследователем
- Количество соседей, веса объектов и расстояние
- Эти параметры можно подобрать

МФТИ

Количество соседей



КОЛИЧЕСТВО СОСЕДЕЙ



Веса соседей как функция от номера

- $\mathrm{w}(i)$ должна не возрастать
- Простейший вариант: $\mathrm{w}(i) = 1$
- $\mathrm{w}(i) = q^i,\ 0 < q < 1$
- $\mathrm{w}(i) = \frac{1}{i},\ \mathrm{w}(i) = \frac{1}{i+a},\ \mathrm{w}(i) = \frac{1}{(i+a)^b}$
- Не очень удачный вариант: $\mathrm{w}(i) = 1 - \frac{i-1}{k}$

МФТИ

ВЕСА СОСЕДЕЙ КАК ФУНКЦИЯ ОТ РАССТОЯНИЯ

- $\mathbf{w}(d) = 1$
- $\mathbf{w}(d) = \frac{1}{d}$

ВЕСА СОСЕДЕЙ КАК ФУНКЦИЯ ОТ РАССТОЯНИЯ

- $\mathrm{w}(d) = 1$
- $\mathrm{w}(d) = \frac{1}{d}$
- $\mathrm{w}(d) = \frac{1}{(q+a)^b}$
- $\mathrm{w}(d) = q^d,\ 0 < q < 1$
- Ядра

ВЕСА СОСЕДЕЙ КАК ФУНКЦИЯ ОТ РАССТОЯНИЯ

- В kNN можно настраивать количество соседей, функции весов объектов и метрику
- Делать это нужно не на обучающей выборке, а по качеству в кросс-валидации
- Уже обсудили: выбор количества соседей и весов объектов
- Далее: метрики

МЕТРИКИ

ЧТО ТАКОЕ РАССТОЯНИЕ?

- $\rho(x, y) \geq 0$
- $\rho(x, y) = \rho(y, x)$
- $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$

ЧТО ТАКОЕ РАССТОЯНИЕ?

- $\rho(x,y) \geq 0$
- $\rho(x,y) = \rho(y,x)$
- $\rho(x,y) \leq \rho(x,z) + \rho(z,y)$
- Для нас это не важно, главное — чтобы работало

МФТИ

САМОЕ ОБЫЧНОЕ РАССТОЯНИЕ

- Евклидова метрика:

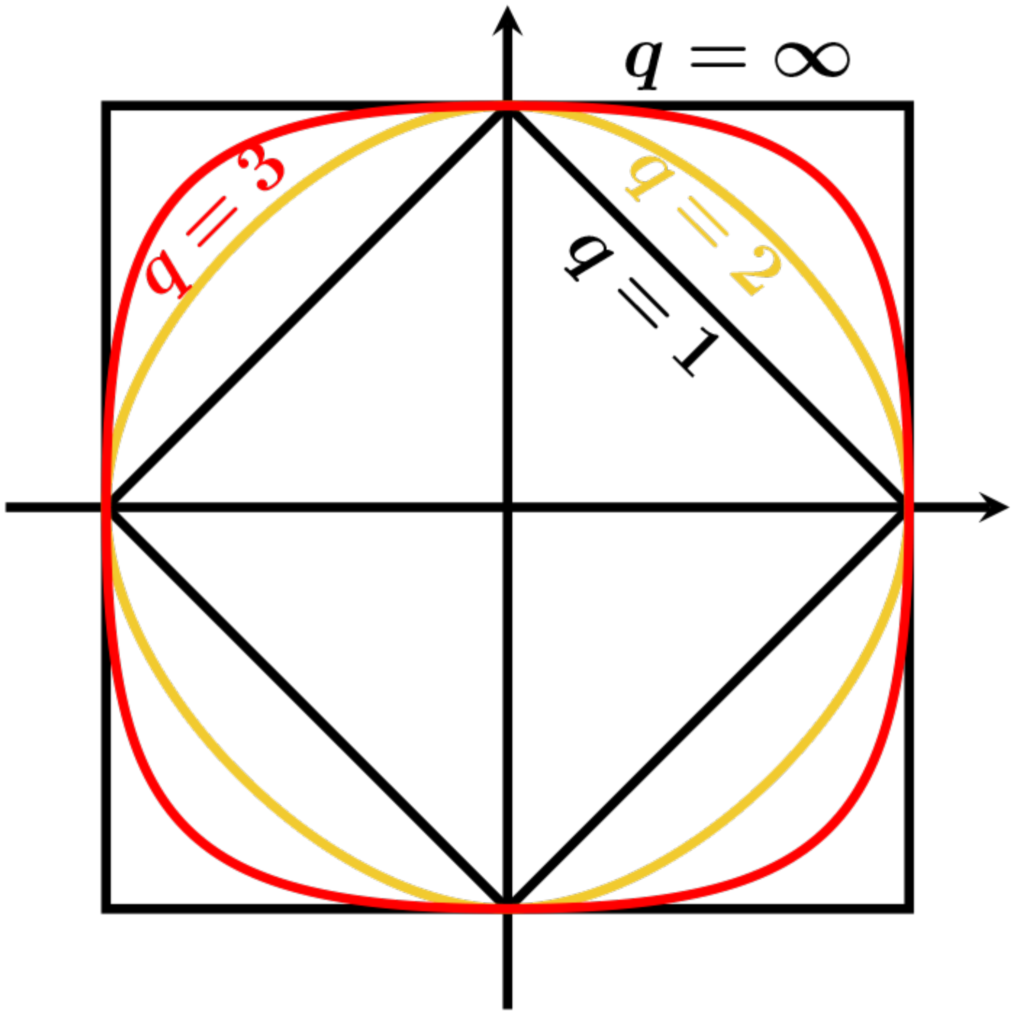
$$\rho(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Метрика Минковского

- Евклидова метрика: $$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$
- Манхэттенская метрика: $$\sum_{i=1}^{n}|x_i - y_i|$$
- Метрика Минковского: $$\left(\sum_{i=1}^{n}(|x_i - y_i|)^q\right)^{\frac{1}{q}}$$

МФТИ

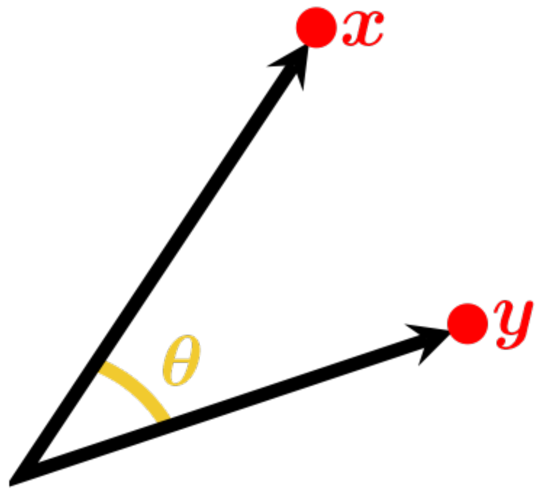
КАК «ПОСМОТРЕТЬ» НА МЕТРИКУ



КОСИНУСНАЯ МЕРА

$$\text{similarity} = \cos(\theta) = \frac{x \cdot y}{\|x\| \cdot \|y\|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$

- Функция близости, а не расстояние



Коэффициент корреляции

$$r = \frac{\sum_{i=1}^{n}((x_i - \bar{x})(y_i - \bar{y}))}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

ДРУГИЕ ФУНКЦИИ БЛИЗОСТИ

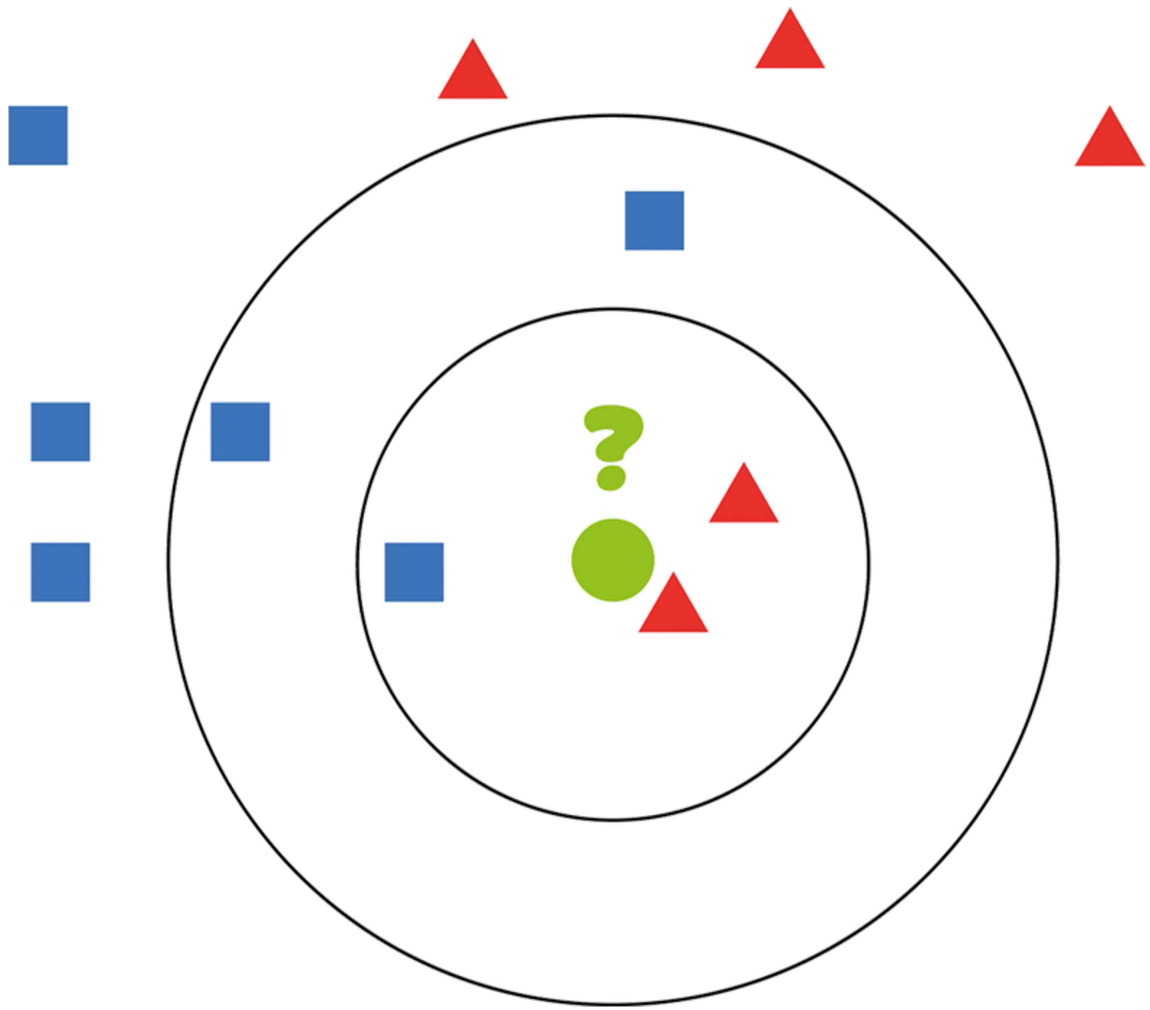
- Скалярное произведение: $\sum x_i \cdot y_i$
- Коэффициент Дайса: $\frac{2\sum x_i \cdot y_i}{\sum x_i^2 + \sum y_i^2}$
- Косинусная мера: $\frac{\sum x_i \cdot y_i}{\sqrt{\sum x_i^2} \cdot \sqrt{\sum y_i^2}}$
- Коэффициент Жаккара: $\frac{\sum x_i \cdot y_i}{\sum x_i^2 + \sum y_i^2 - \sum x_i \cdot y_i}$

РЕЗЮМЕ

- Можно придумать множество разных метрик
- Чтобы понять, какая метрика лучше подходит в задаче, нужно просто попробовать разные
- Некоторые метрики и функции близости часто используются в задачах из какой-то конкретной области, но все равно полезно экспериментировать

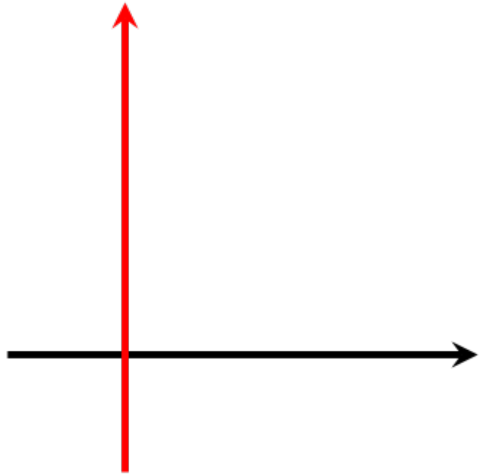
ПРОКЛЯТИЕ РАЗМЕРНОСТИ

РОЛЬ РАССТОЯНИЯ В МЕТРИЧЕСКИХ МЕТОДАХ

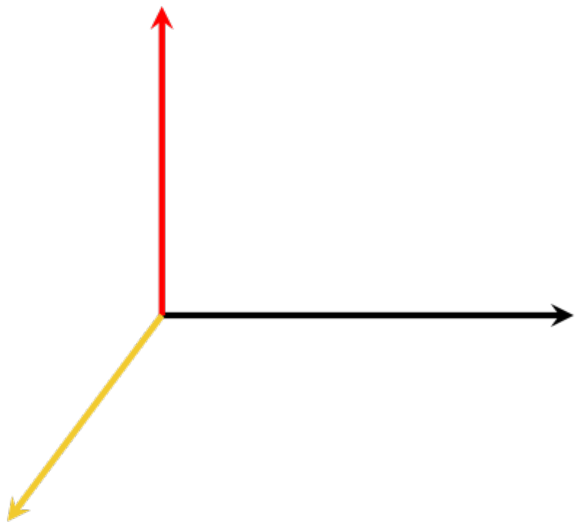


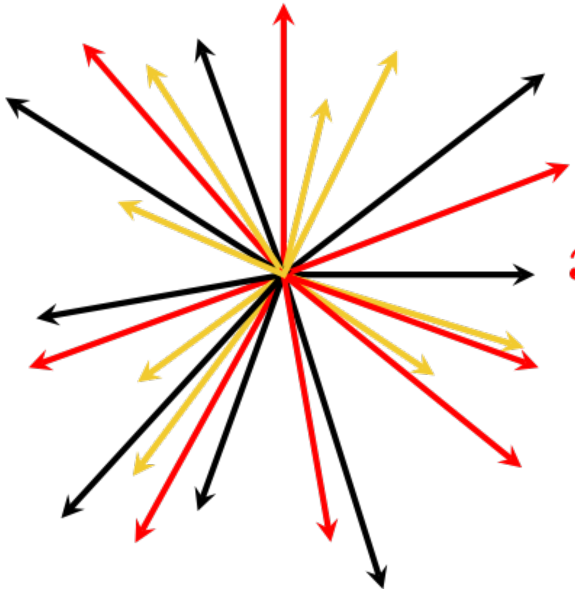
Роль расстояния в метрических методах

$$x = (x_1, x_2)$$



$$x = (x_1, x_2, x_3)$$





$$x = (x_1, x_2, x_3, \ldots, x_N)$$

РАССТОЯНИЕ В ПРОСТРАНСТВЕ ВЫСОКОЙ РАЗМЕРНОСТИ

- Нет ли чего-то необычного в свойствах расстояния в пространствах высокой размерности?

МФТИ

ИДЕЯ 1: НЕБОЛЬШИЕ ОТЛИЧИЯ В БОЛЬШОМ ЧИСЛЕ КООРДИНАТ

$$x_1 = (a_1, a_2, ..., a_N)$$

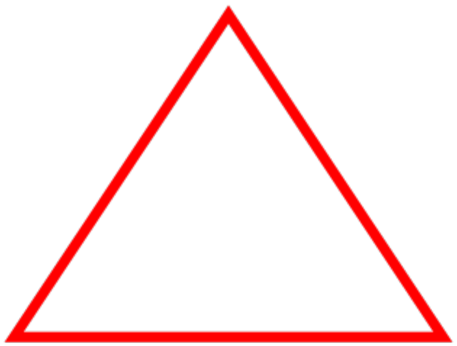
$$x_2 = (a_1 + \varepsilon, a_2 + \varepsilon, ..., a_N + \varepsilon)$$

$$x_3 = (a_1, a_2 + \Delta, a_3, ..., a_N)$$

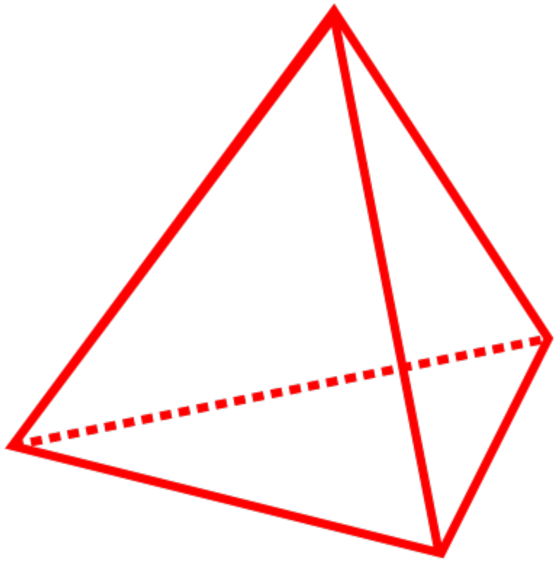
МФТИ

ИДЕЯ 2: ПОЧТИ ОДИНАКОВЫЕ РАССТОЯНИЯ

Треугольник



Тетраэдр



В N-мерном пространстве — до $N+1$ равноудалённой точки

ИДЕЯ 3: ЭКСПОНЕНЦИАЛЬНЫЙ РОСТ НЕОБХОДИМЫХ ДАННЫХ

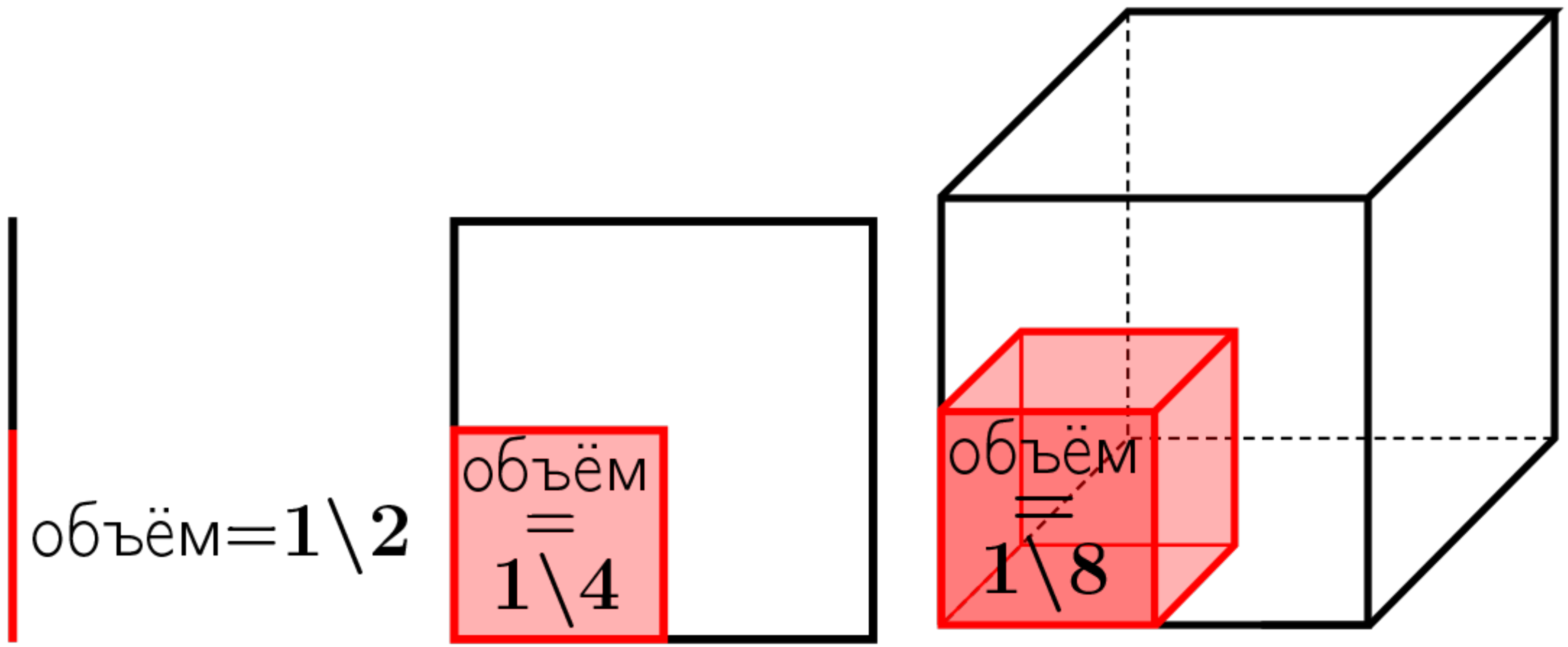
- Рассмотрим векторы размерности N из бинарных признаков:

$$X = (0, 0, 1, 0, 1, 1, \ldots, 1)$$

- Всевозможных комбинаций значений признаков — 2^N

- С ростом N экспоненциально увеличивается необходимое количество данных

ПРИМЕР: ВЕРОЯТНОСТЬ ПОПАДАНИЯ В КУБ



ПРИМЕР: ВЕРОЯТНОСТЬ ПОПАДАНИЯ В КУБ

- Куб с длиной ребра 0.99 будет составлять 0.99^n часть от объема единичного куба

$$\lim_{n\to\infty} 0.99^n = 0$$

РЕЗЮМЕ

- Расстояние в пространствах высокой размерности
- Рост необходимого объёма данных
- Примерно одинаковое расстояние между большинством пар объектов
- Проклятие размерности

РЕКОМЕНДАЦИИ ФИЛЬМОВ С ПОМОЩЬЮ KNN

ЗАДАЧА

- Есть известные оценки, которые пользователи поставили уже просмотренным фильмам

- Нужно:
  - Спрогнозировать оценки, которые поставили бы пользователи другим фильмам
  - Порекомендовать пользователям то, что им больше понравится

ЗАДАЧА

	Пила	Улица Вязов	Ванильное небо	1+1
Маша	5	4	1	2
Юля		5	2	
Вова			3	5
Коля	3		4	5
Петя				4
Ваня		5	3	3

МФТИ

ЗАДАЧА

	Пила	Улица Вязов	Ванильное небо	1+1
Маша	5	4	1	2
Юля		5	2	?
Вова			3	5
Коля	3		4	5
Петя				4
Ваня		5	3	3

USER-BASED

	Пила	Улица Вязов	Ванильное небо	1+1
Маша	5	4	1	2
Юля		5	2	
Вова			3	5
Коля			4	5
Петя				4
Ваня		5	3	3

USER-BASED

	Пила	Улица Вязов	Ванильное небо	1+1
Маша	5	4	1	2
Юля		5	2	?
Вова			3	5
Коля	3		4	5
Петя				4
Ваня		5	3	3

ITEM-BASED

	Пила	Улица Вязов	Ванильное небо	1+1
Маша	5	4	1	2
Юля		5	2	?
Вова			3	5
Коля	3		4	5
Петя				4
Ваня		5	3	3

МФТИ

ПОХОЖЕСТЬ ПОЛЬЗОВАТЕЛЕЙ

				a	
		Пила	Улица Вязов	Ванильное небо	1+1
	Маша	5	4	1	2
	Юля		5	2	
	Вова			3	5
i	Коля	3		$r_{i,a}$	5
	Петя				4
j	Ваня		5	$r_{j,a}$	3

$$\bar{r}_i = \frac{3 + r_{i,a} + 5}{3} \qquad \bar{r}_j = \frac{5 + r_{j,a} + 3}{3}$$

ПОХОЖЕСТЬ ПОЛЬЗОВАТЕЛЕЙ

				a	
		Пила	Улица Вязов	Ванильное небо	1+1
	Маша	5	4	1	2
	Юля		5	2	
	Вова			3	5
i	Коля	3		$r_{i,a}$	5
	Петя				4
j	Ваня		5	$r_{j,a}$	3

$$\bar{r}_i = \frac{3 + r_{i,a} + 5}{3} \qquad \bar{r}_j = \frac{5 + r_{j,a} + 3}{3}$$

МФТИ

ПОХОЖЕСТЬ ПОЛЬЗОВАТЕЛЕЙ

				a	
		Пила	Улица Вязов	Ванильное небо	1+1
	Маша	5	4	1	2
	Юля		5	2	
	Вова			3	5
i	Коля	3		$r_{i,a}$	5
	Петя				4
j	Ваня		5	$r_{j,a}$	3

$$\mathbf{w}_{i,j} = \frac{\sum\limits_a (r_{i,a} - \bar{r}_i)(r_{j,a} - \bar{r}_j)}{\sqrt{\sum\limits_a (r_{i,a} - \bar{r}_i)^2}\sqrt{\sum\limits_a (r_{j,a} - \bar{r}_j)^2}}$$

ПОХОЖЕСТЬ ПОЛЬЗОВАТЕЛЕЙ

$$\mathrm{w}_{i,j} = \frac{\sum_a (r_{i,a} - \bar{r}_i)(r_{j,a} - \bar{r}_j)}{\sqrt{\sum_a (r_{i,a} - \bar{r}_i)^2} \sqrt{\sum_a (r_{j,a} - \bar{r}_j)^2}}$$

Можно делать суммирование не по всем фильмам (заменяя каким-то числом пропуски), а только по тем, которые посмотрели и i-тый и j-тый пользователи

МФТИ

ПОХОЖЕСТЬ ПОЛЬЗОВАТЕЛЕЙ

$$\mathbf{w}_{i,j} = \frac{\sum_a r_{i,a} r_{j,a}}{\sqrt{\sum_a r_{i,a}^2}\sqrt{\sum_a r_{j,a}^2}}$$

ПРОГНОЗИРОВАНИЕ РЕЙТИНГА

$$\hat{r}_{i,a} = \bar{r}_i + \frac{\sum_j (r_{j,a} - \bar{r}_j)\,\mathbf{w}_{i,j}}{\sum_j |\,\mathbf{w}_{i,j}\,|}$$

ПРОГНОЗИРОВАНИЕ РЕЙТИНГА

$$\hat{r}_{i,a} = \bar{r}_i + \frac{\sum_j (r_{j,a} - \bar{r}_j)\,\mathbf{w}_{i,j}}{\sum_j |\,\mathbf{w}_{i,j}\,|}$$

МФТИ

ПРОГНОЗИРОВАНИЕ РЕЙТИНГА

$$\cancel{\hat{r}_{i,a} = \bar{r}_i + \frac{\sum_j (r_{j,a} - \bar{r}_j)\, \mathbf{w}_{i,j}}{\sum_j |\mathbf{w}_{i,j}|}}$$

$$\hat{r}_{i,a} = \bar{r}_i + \frac{\sum_{j \in \mathrm{kNN}(i)} (r_{j,a} - \bar{r}_j)\, \mathbf{w}_{i,j}}{\sum_{j \in \mathrm{kNN}(i)} |\mathbf{w}_{i,j}|}$$

Резюме

- kNN — не только для классификации и регрессии
- Другие задачи тоже могут легко сводиться к той же идее

Метод опорных векторов (SVM)

МЕТОД ОПОРНЫХ ВЕКТОРОВ

- Линейный классификатор:

$$a(x) = \mathrm{sign}(< \mathrm{w}, x > - \mathrm{w}_0)$$

- Использующий кусочно-линейную функцию потерь и $L2$-регуляризатор:

$$\sum_{i=1}^{l} \boxed{L(M_i)} + \boxed{\gamma \| \mathrm{w} \|^2} \to \min_{\mathrm{w}}$$

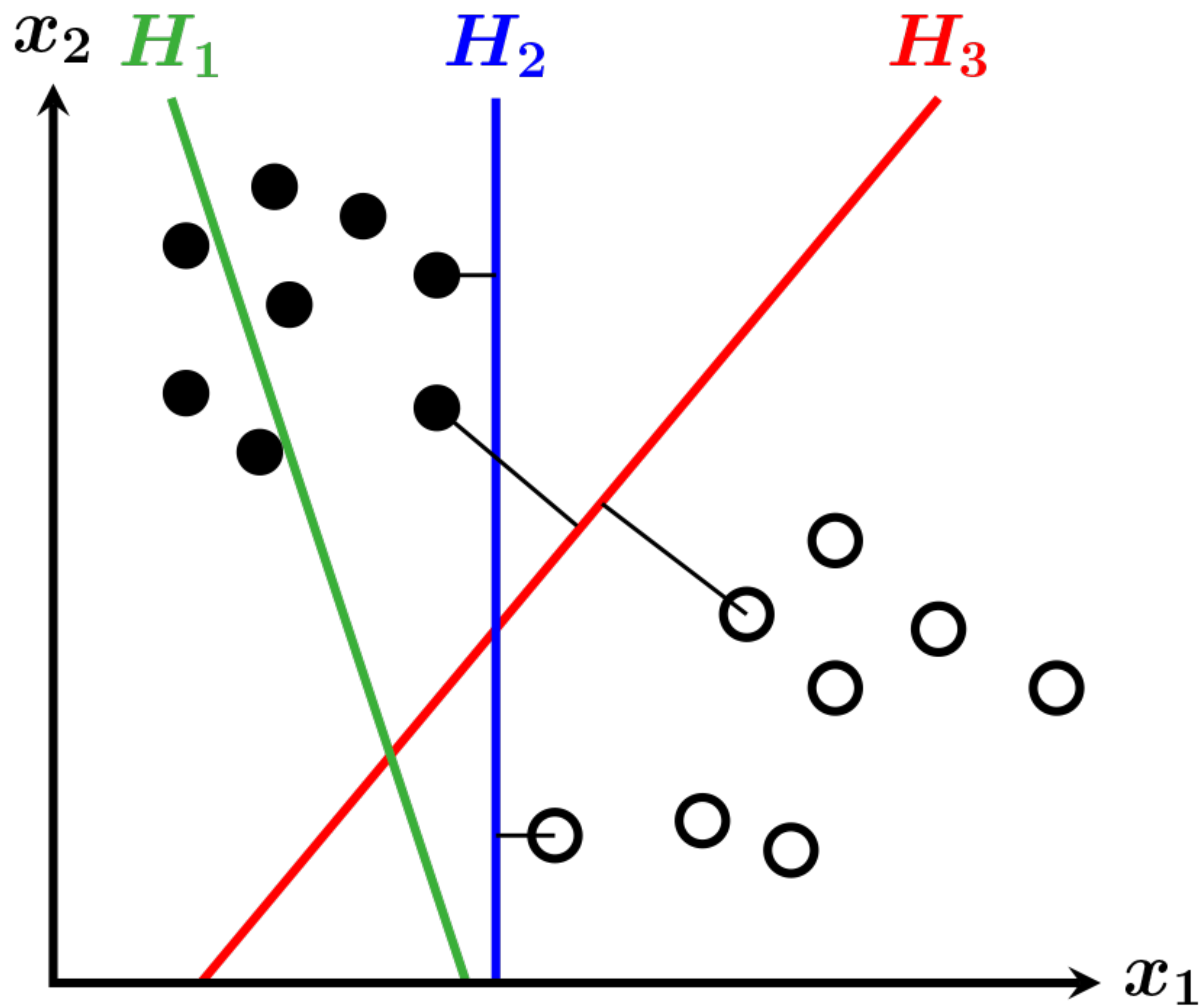
функция потерь квадратичный регуляризатор

- Кусочно-линейная функция потерь:

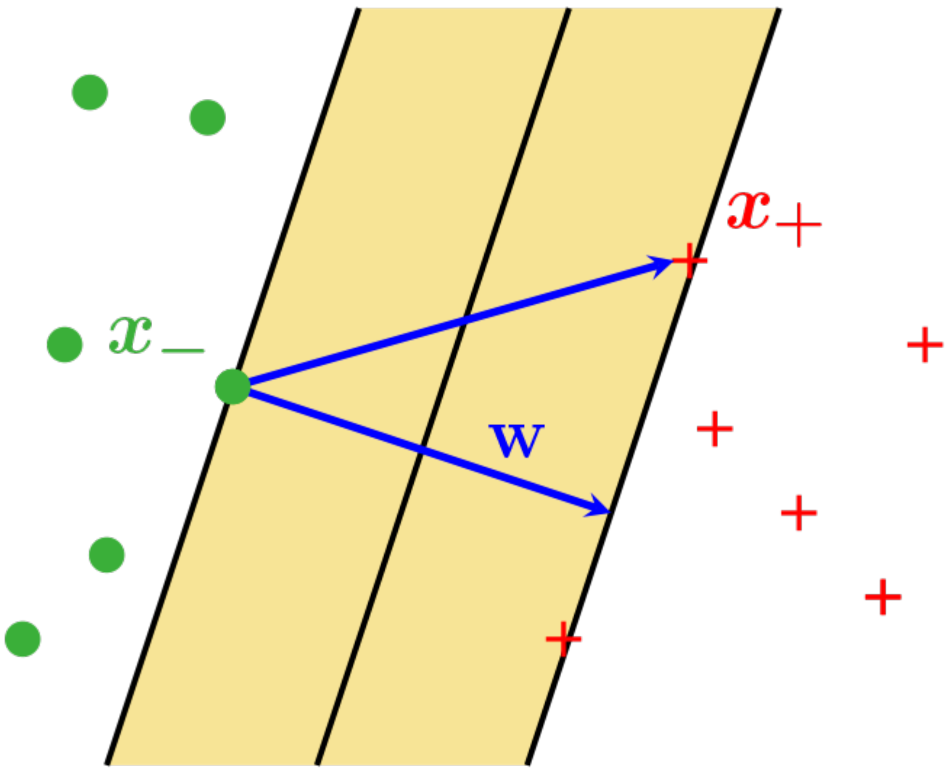
$$L(M_i) = \max\{0, 1 - M_i\} = (1 - M_i)_+$$

МФТИ

ПОСТРОЕНИЕ РАЗДЕЛЯЮЩЕЙ ГИПЕРПЛОСКОСТИ

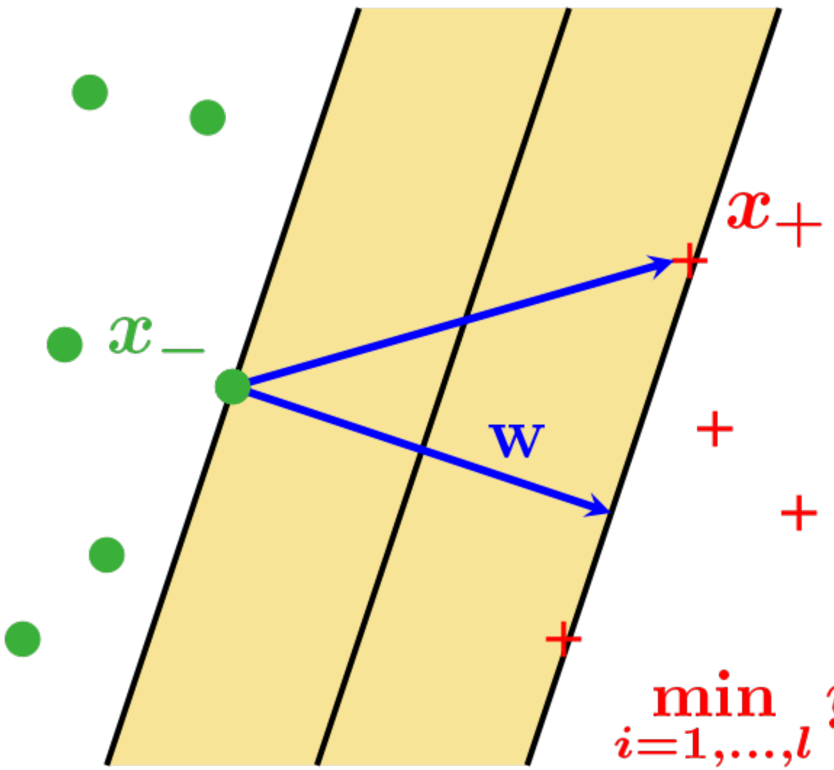


РАЗДЕЛЯЮЩАЯ ПОЛОСА



$$a(x) = \mathrm{sign}\left(\sum_{j=1}^{n} \mathrm{w}_j\, x^j - \mathrm{w}_0\right) = \mathrm{sign}\left(\langle \mathrm{w}, x\rangle - \mathrm{w}_0\right)$$

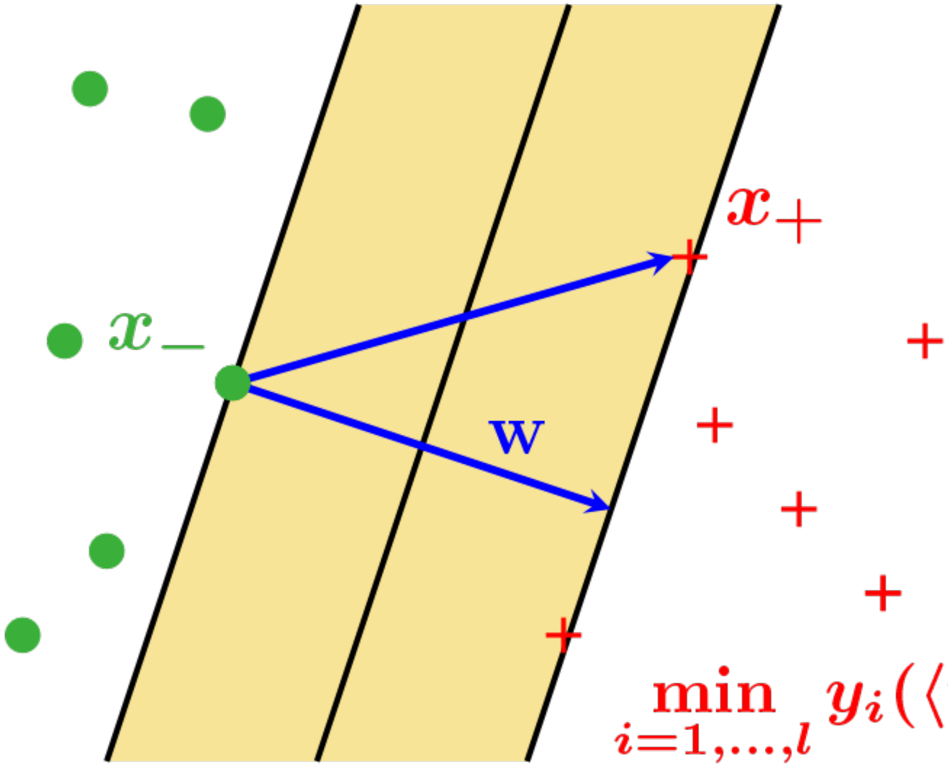
РАЗДЕЛЯЮЩАЯ ПОЛОСА



$$\min_{i=1,\dots,l} y_i(\langle \mathrm{w}, x_i\rangle - \mathrm{w}_0) = 1$$

$$a(x) = \operatorname{sign}\left(\sum_{j=1}^{n} \mathrm{w}_j\, x^j - \mathrm{w}_0\right) = \operatorname{sign}\left(\langle \mathrm{w}, x\rangle - \mathrm{w}_0\right)$$

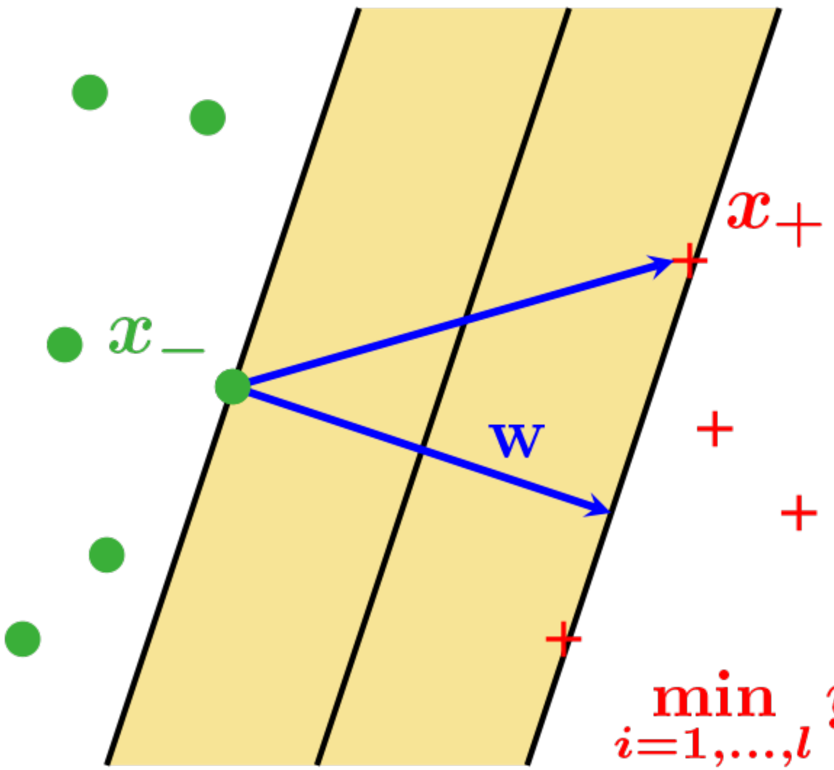
ШИРИНА РАЗДЕЛЯЮЩЕЙ ПОЛОСЫ



$$\min_{i=1,\dots,l} y_i(\langle \mathrm{w}, x_i\rangle - \mathrm{w}_0) = 1$$

$$\left\langle (x_+ - x_-), \frac{\mathrm{w}}{\|\mathrm{w}\|} \right\rangle$$

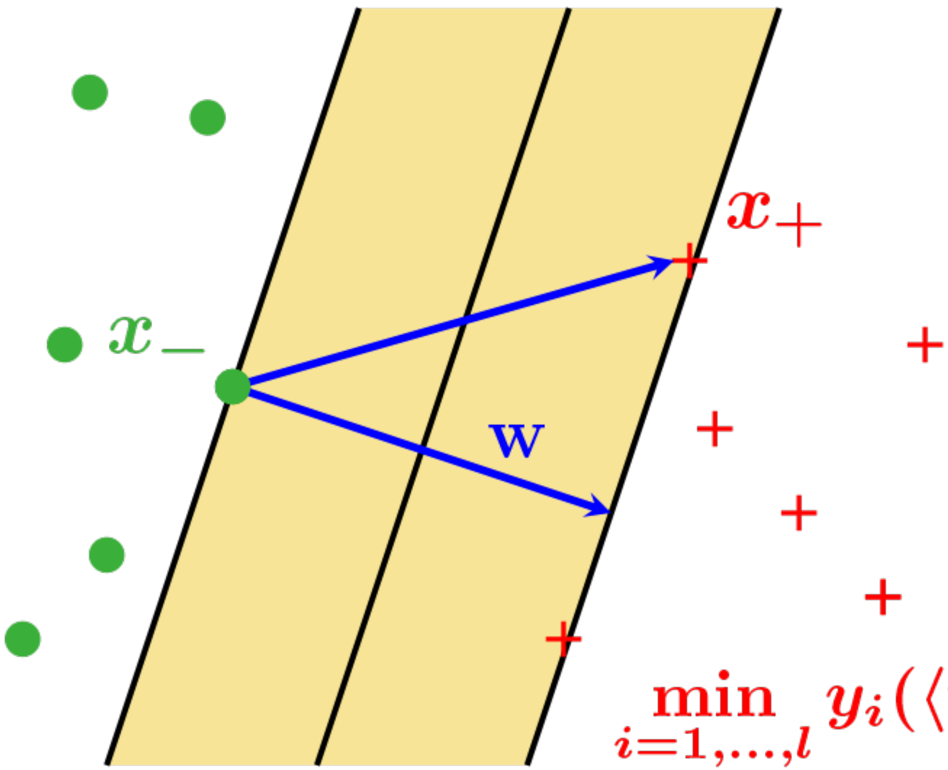
Ширина разделяющей полосы



$$\min_{i=1,\ldots,l} y_i(\langle \mathrm{w}, x_i\rangle - \mathrm{w}_0) = 1$$

$$\left\langle (x_+ - x_-), \frac{\mathrm{w}}{\|\mathrm{w}\|} \right\rangle = \frac{\langle \mathrm{w}, x_+\rangle - \langle \mathrm{w}, x_-\rangle}{\|\mathrm{w}\|}$$

ШИРИНА РАЗДЕЛЯЮЩЕЙ ПОЛОСЫ



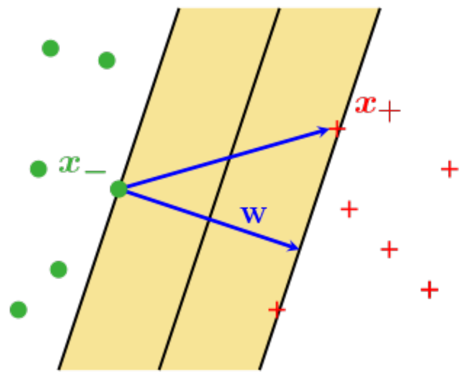
$$\min_{i=1,\dots,l} y_i(\langle \mathrm{w}, x_i\rangle - \mathrm{w}_0) = 1$$

$$\left\langle (x_+ - x_-), \frac{\mathrm{w}}{\|\mathrm{w}\|} \right\rangle = \frac{\langle \mathrm{w}, x_+\rangle - \langle \mathrm{w}, x_-\rangle}{\|\mathrm{w}\|} =$$

$$= \frac{(\mathrm{w}_0 + 1) - (\mathrm{w}_0 - 1)}{\|\mathrm{w}\|} = \frac{2}{\|\mathrm{w}\|}$$

МФТИ

Ширина разделяющей полосы

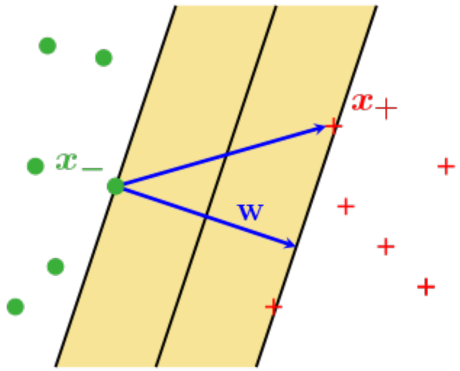


$$\min_{i=1,\dots,l} y_i(\langle \mathrm{w}, x_i\rangle - \mathrm{w}_0) = 1$$

$$\left\langle (x_+ - x_-), \frac{\mathrm{w}}{\|\mathrm{w}\|} \right\rangle = \frac{\langle \mathrm{w}, x_+\rangle - \langle \mathrm{w}, x_-\rangle}{\|\mathrm{w}\|} =$$

$$= \frac{(\mathrm{w}_0 + 1) - (\mathrm{w}_0 - 1)}{\|\mathrm{w}\|} = \frac{2}{\|\mathrm{w}\|}$$

МАКСИМИЗАЦИЯ ЗАЗОРА



$$\min_{i=1,\dots,l} y_i(\langle \mathbf{w}, x_i \rangle - \mathbf{w}_0) = 1$$

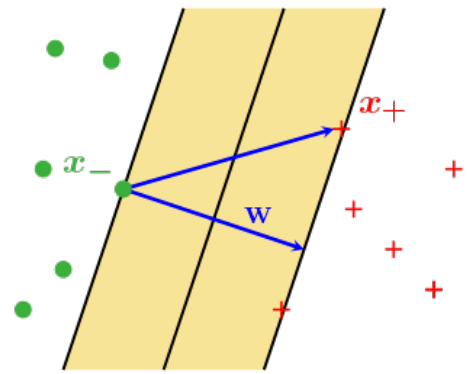
$$\left\langle (x_+ - x_-), \frac{\mathbf{w}}{\|\mathbf{w}\|} \right\rangle = \frac{\langle \mathbf{w}, x_+ \rangle - \langle \mathbf{w}, x_- \rangle}{\|\mathbf{w}\|} =$$

$$= \frac{(\mathbf{w}_0 + 1) - (\mathbf{w}_0 - 1)}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|}$$

$$\begin{cases} \langle \mathbf{w}, \mathbf{w} \rangle \to \min \\ y_i\left(\langle \mathbf{w}, x_i \rangle - \mathbf{w}_0\right) \geq 1, \; i = 1, \; \dots, l \end{cases}$$

МФТИ

Случай линейно неразделимой выборки



$$\begin{cases} \langle \mathbf{w}, \mathbf{w} \rangle \to \min \\ y_i \left(\langle \mathbf{w}, x_i \rangle - \mathbf{w}_0 \right) \geq 1, \; i = 1, \, \ldots \, , l \end{cases}$$

$$\begin{cases} \frac{1}{2} \langle \mathbf{w}, \mathbf{w} \rangle + C \sum\limits_{i=1}^{l} \xi_i \to \min\limits_{\mathbf{w}, \mathbf{w}_0, \xi} \\ y_i \left(\langle \mathbf{w}, x_i \rangle - \mathbf{w}_0 \right) \geq 1 - \xi_i, \; i = 1, \, \ldots \, , l \\ \xi_i \geq 0, \; i = 1, \, \ldots \, , l \end{cases}$$

МФТИ

ОПТИМИЗИЦИОННАЯ ЗАДАЧА В SVM

$$\begin{cases} \frac{1}{2}\langle \mathrm{w}, \mathrm{w} \rangle + C \sum\limits_{i=1}^{l} \xi_i \to \min\limits_{\mathrm{w}, \mathrm{w}_0, \xi} \\ y_i \left(\langle \mathrm{w}, x_i \rangle - \mathrm{w}_0 \right) \geq 1 - \xi_i, \; i = 1, \; \ldots \; , l \\ \xi_i \geq 0, \; i = 1, \; \ldots \; , l \end{cases}$$

Причём здесь линейный классификатор в привычном нам виде?

БЕЗУСЛОВНАЯ ОПТИМИЗАЦИОННАЯ ЗАДАЧА В SVM

$$
\begin{cases}
\frac{1}{2}\langle \mathbf{w}, \mathbf{w} \rangle + C \sum\limits_{i=1}^{l} \xi_i \to \min\limits_{\mathbf{w}, \mathbf{w}_0, \xi} \\
y_i \left(\langle \mathbf{w}, x_i \rangle - \mathbf{w}_0 \right) \geq 1 - \xi_i, \; i = 1, \ldots, l \\
\xi_i \geq 0, \; i = 1, \ldots, l
\end{cases}
$$

Напоминание:

$M_i = y_i \left(\langle \mathbf{w}, x_i \rangle - \mathbf{w}_0 \right)$

отступ на i-том объекте

БЕЗУСЛОВНАЯ ОПТИМИЗАЦИОННАЯ ЗАДАЧА В SVM

$$\begin{cases} \frac{1}{2}\langle \mathrm{w}, \mathrm{w}\rangle + C\sum\limits_{i=1}^{l} \xi_i \to \min\limits_{\mathrm{w}, \mathrm{w}_0, \xi} \\ y_i\left(\langle \mathrm{w}, x_i\rangle - \mathrm{w}_0\right) \geq 1 - \xi_i,\ i = 1,\ \ldots, l \\ \xi_i \geq 0,\ i = 1,\ \ldots, l \end{cases}$$

Напоминание:

$M_i = y_i\left(\langle \mathrm{w}, x_i\rangle - \mathrm{w}_0\right)$

отступ на i-том объекте

$$\xi_i \geq 0$$

$$\xi_i \geq 1 - M_i$$

$$\sum_{i=1}^{l} \xi_i \to \min$$

Безусловная оптимизационная задача в SVM

$$\begin{cases} \frac{1}{2}\langle \mathrm{w}, \mathrm{w}\rangle + C\sum\limits_{i=1}^{l}\xi_i \to \min\limits_{\mathrm{w}, \mathrm{w}_0, \xi} \\ y_i\left(\langle \mathrm{w}, x_i\rangle - \mathrm{w}_0\right) \geq 1 - \xi_i, \;\; i = 1, \;\ldots\;, l \\ \xi_i \geq 0, \;\; i = 1, \;\ldots\;, l \end{cases}$$

Напоминание:

$M_i = y_i\left(\langle \mathrm{w}, x_i\rangle - \mathrm{w}_0\right)$

отступ на i-том объекте

$$\xi_i \geq 0$$

$$\xi_i \geq 1 - M_i$$

$$\Rightarrow \;\; \xi_i = \max\{0, 1 - M_i\} = (1 - M_i)_+$$

$$\sum_{i=1}^{l}\xi_i \to \min$$

МФТИ

Безусловная оптимизационная задача в SVM

$$\begin{cases} \frac{1}{2}\langle \mathrm{w}, \mathrm{w}\rangle + C\sum_{i=1}^{l}\xi_i \to \min\limits_{\mathrm{w},\mathrm{w}_0,\xi} \\ y_i\left(\langle \mathrm{w}, x_i\rangle - \mathrm{w}_0\right) \geq 1-\xi_i, \;\; i=1,\;\ldots\;,l \\ \xi_i \geq 0, \;\; i=1,\;\ldots\;,l \end{cases}$$

Напоминание:

$M_i = y_i\left(\langle \mathrm{w}, x_i\rangle - \mathrm{w}_0\right)$

отступ на i-том объекте

$$\xi_i \geq 0$$

$$\xi_i \geq 1 - M_i$$

$$\Rightarrow \;\; \xi_i = \max\{0, 1 - M_i\} = (1 - M_i)_+$$

$$\sum_{i=1}^{l}\xi_i \to \min$$

$$Q(\mathrm{w}, \mathrm{w}_0) = \sum_{i=1}^{l}\left(1 - M_i(\mathrm{w}, \mathrm{w}_0)\right)_+ + \frac{1}{2C}\|\,\mathrm{w}\,\|^2 \to \min_{\mathrm{w},\mathrm{w}_0}$$

БЕЗУСЛОВНАЯ ОПТИМИЗАЦИОННАЯ ЗАДАЧА В SVM

$$Q(\mathrm{w}, \mathrm{w}_0) = \sum_{i=1}^{l} (1 - M_i(\mathrm{w}, \mathrm{w}_0))_+ + \frac{1}{2C} \| \mathrm{w} \|^2 \to \min_{\mathrm{w}, \mathrm{w}_0}$$

РЕЗЮМЕ

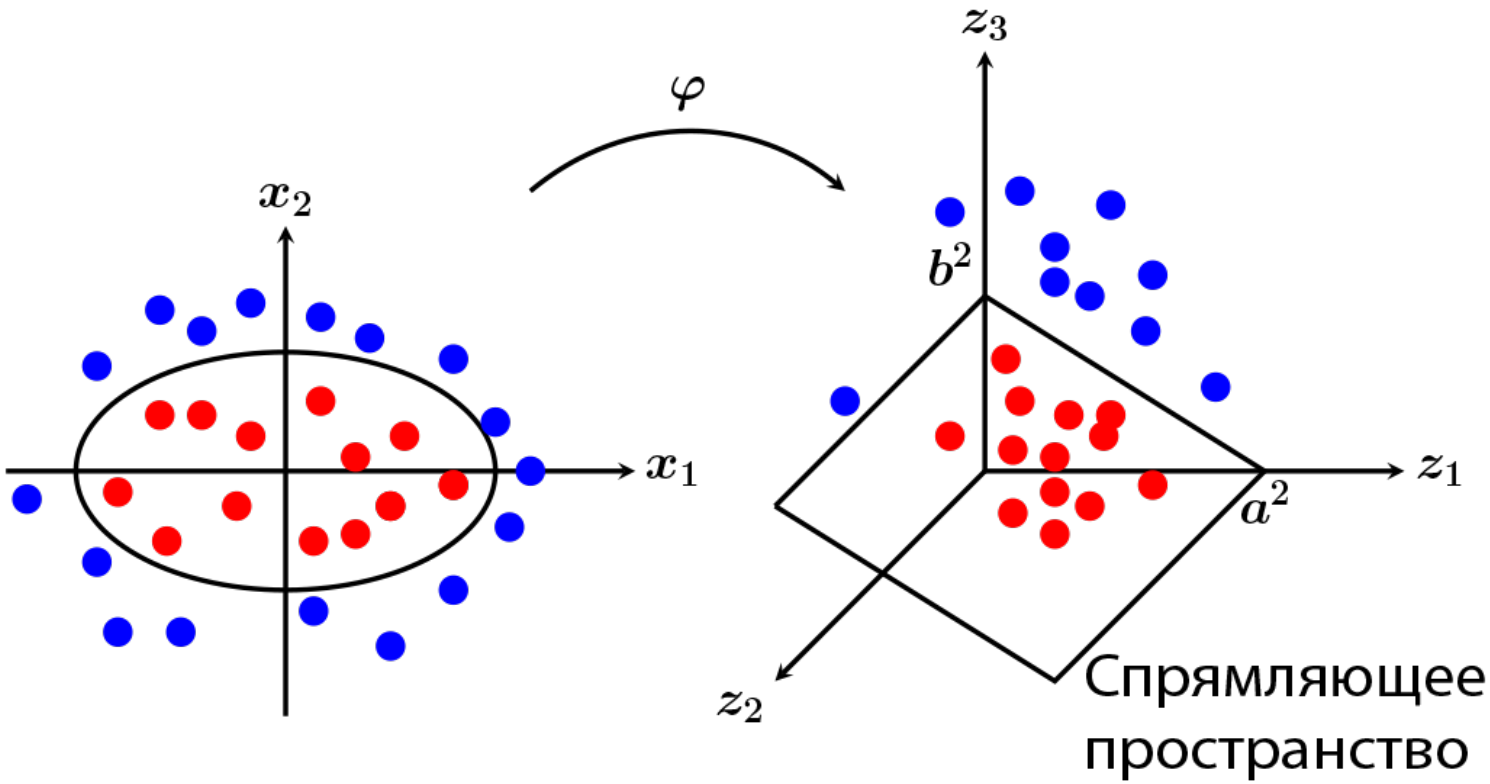
- Метод опорных векторов — линейный классификатор с кусочно-линейной функцией потерь (hinge loss) и $L2$ -регуляризатором
- Придуман метод был из соображений максимизации зазора между классами
- В случае линейно разделимой выборки это означает просто максимизацию ширины разделяющей полосы
- А в случае линейно неразделимой выборки просто добавляется возможность попадания объектов в полосу и штрафы за это

МФТИ

ЯДРА В МЕТОДЕ ОПОРНЫХ ВЕКТОРОВ

МФТИ

Добавление новых признаков



$$\varphi : (x_1, x_2) \to (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$

$$\left(\frac{x_1}{a}\right)^2 + \left(\frac{x_2}{b}\right)^2 = 1 \;\to\; \frac{z_1}{a^2} + \frac{z_3}{b^2} = 1$$

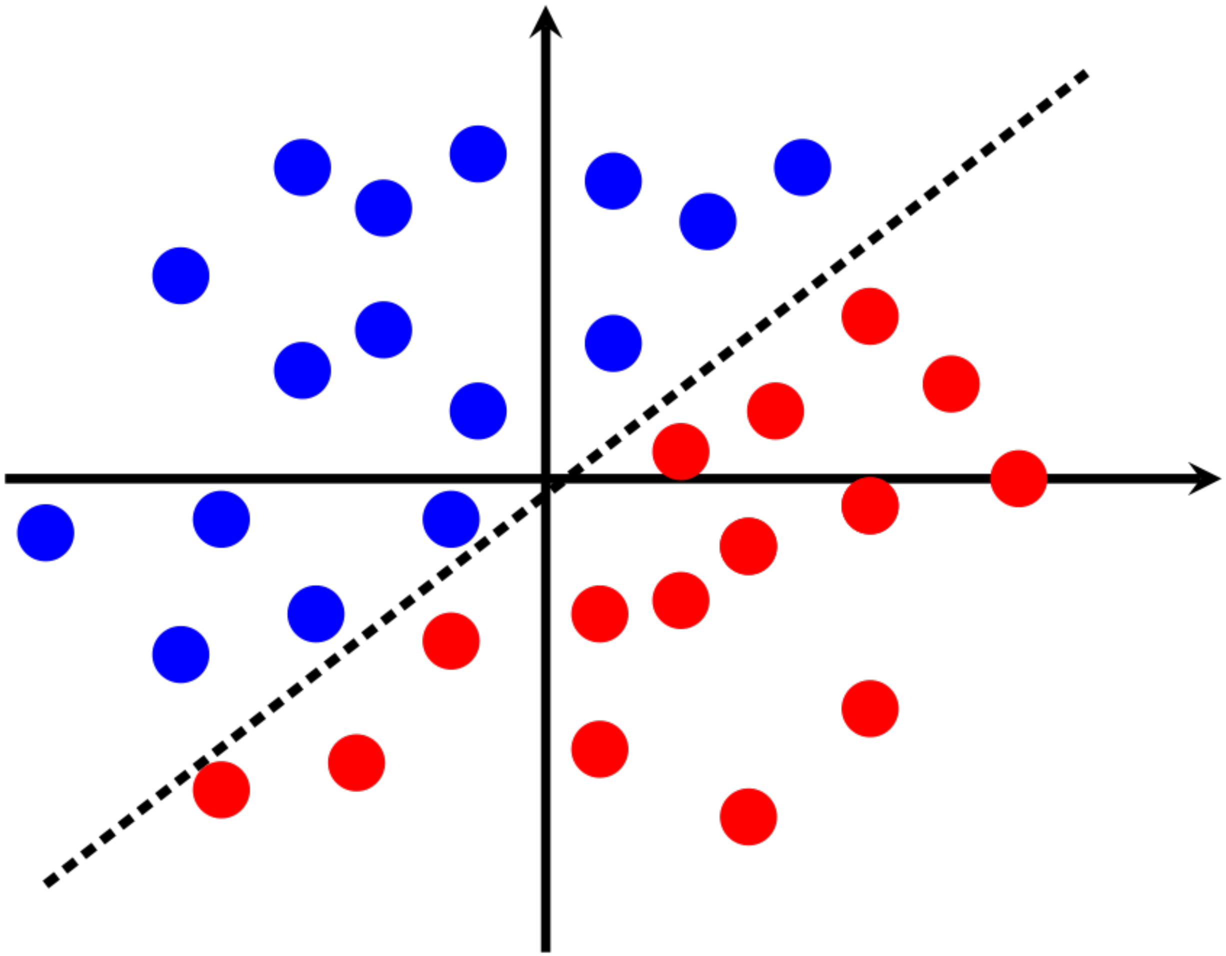
KERNEL TRICK

$$\begin{array}{l} x \mapsto \varphi(x) \\ \mathbf{w} \mapsto \varphi(\mathbf{w}) \end{array} \Rightarrow \langle \mathbf{w}, x \rangle \mapsto \langle \varphi(\mathbf{w}), \varphi(x) \rangle$$

Можно не делать преобразование признаков явно, а вместо скалярного произведения $\langle \mathbf{w}, x \rangle$ использовать функцию $K(\mathbf{w}, x)$, представимую в виде:

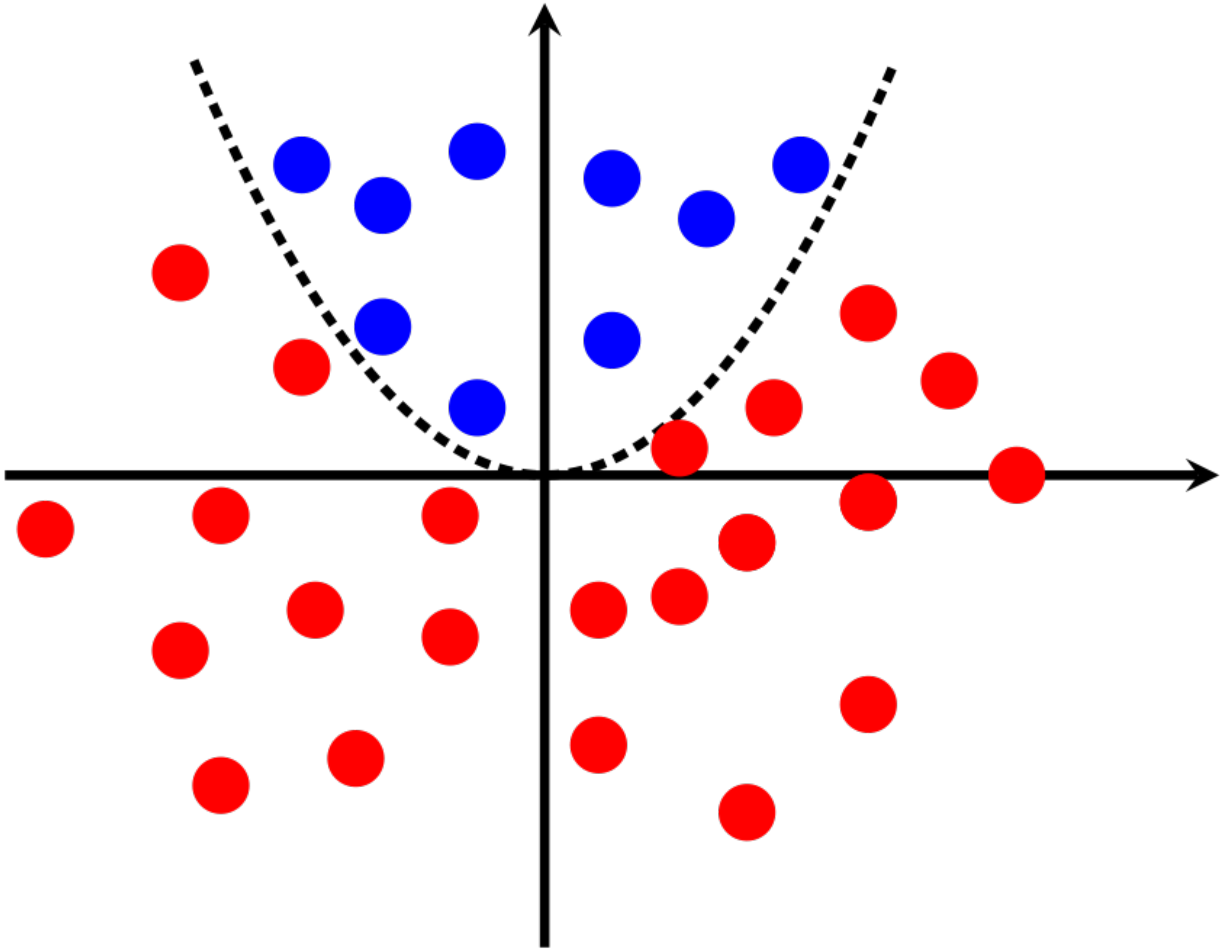
$$K(\mathbf{w}, x) = \langle \varphi(\mathbf{w}), \varphi(x) \rangle$$

ЛИНЕЙНОЕ ЯДРО



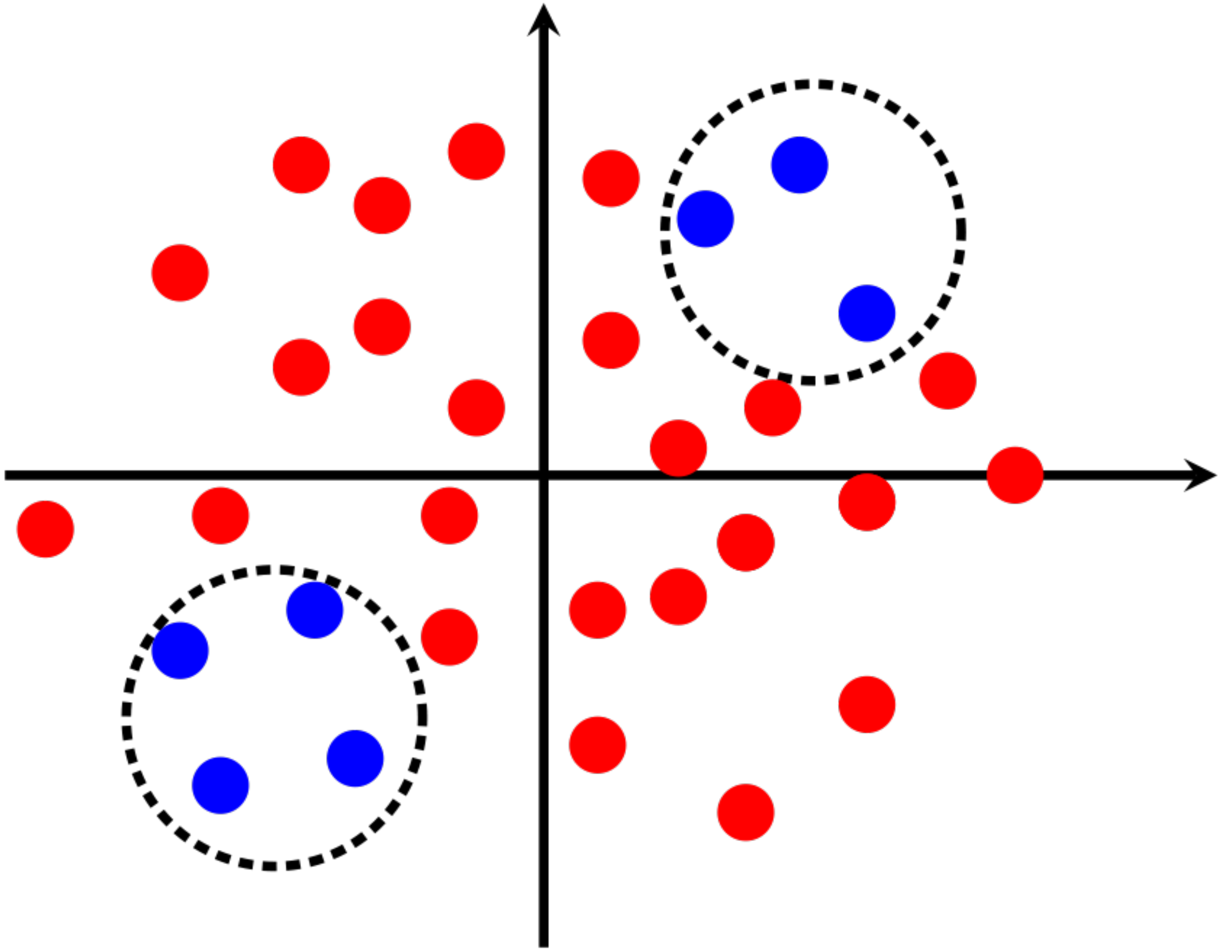
$$K(\mathbf{w}, x) = \langle \mathbf{w}, x \rangle$$

ПОЛИНОМИАЛЬНОЕ ЯДРО



$$K(\mathrm{w}, x) = (\gamma\langle \mathrm{w}, x\rangle + r)^d$$

РАДИАЛЬНОЕ ЯДРО



$$K(\mathrm{w}, x) = e^{-\gamma \| \mathrm{w} - x \|^2}$$

ЯДРА И БИБЛИОТЕКИ

- LibSVM — можно выбирать ядра
- LibLinear — только линейное ядро
- Scikit-learn — обёртка над LibSVM и LibLinear
- Vowpal Wabbit — только линейное ядро

Резюме

- Kernel trick
- Линейное, полиномиальное и радиальное ядра
- Библиотеки