

Что такое машинное обучение

Сергей Бартунов
ДБДиИП, ФКН НИУ ВШЭ

Машинное обучение

Машинное обучение

Научная дисциплина, изучающая алгоритмы,
способные настраиваться по данным

Как отличить смородину от бузины?



Объекты одного класса должны иметь похожие
признаковые описания (гипотеза компактности)

Объекты одного класса должны иметь похожие **признаковые описания** (гипотеза компактности)

$$\phi(\text{img1}) = (0.3, 10, 0.2)$$


$$\phi(\text{img2}) = (2.5, 14.2, 1.1)$$


Объекты одного класса должны иметь похожие **признаковые описания** (гипотеза компактности)

$$\phi(\text{img1}) = (0.3, 10, 0.2)$$

отношение ширины
листа к длине высота куста ширина ветки
с ягодами

$$\phi(\text{img2}) = (2.5, 14.2, 1.1)$$

Обучение с учителем

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$$

ϕ (



)

-1 (смородина)

Минимизация эмпирического риска

обучение: по обучающей выборке построить классификатор, минимизирующий эмп. риск

$$f^* = \arg \min_{f \in F} \frac{1}{N} \sum_{i=1}^N l(f(\mathbf{x}_i), y_i)$$

функция потерь $l(\hat{y}, y)$ отражает насколько сильно мы ошиблись на данном примере

Линейный классификатор

Рассмотрим семейство линейных функций классификации:

$$f(\mathbf{x}; \mathbf{w}) = \text{sign}\left(\sum_{d=1}^D w_d x_d + w_0\right)$$

Очевидная функция потерь $l(\hat{y}, y) = \mathbb{1}[\hat{y} \neq y]$

Линейный классификатор

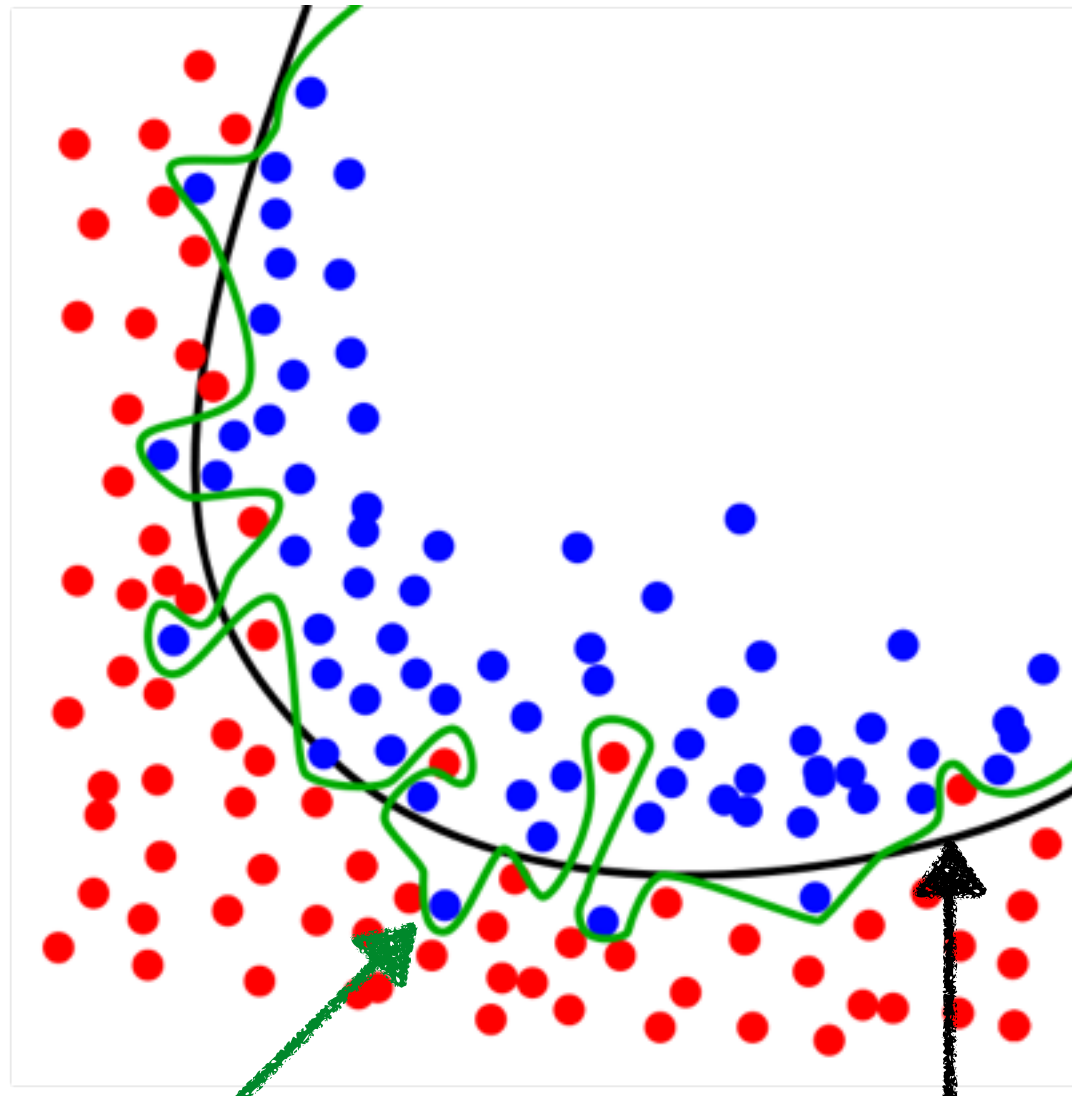
Рассмотрим семейство линейных функций классификации:

$$f(\mathbf{x}; \mathbf{w}) = \text{sign}\left(\sum_{d=1}^D w_d x_d + w_0\right)$$

Очевидная функция потерь $l(\hat{y}, y) = \mathbb{1}[\hat{y} \neq y]$

Обучение: $\mathbf{w}^* = \arg \min_{\mathbf{w}} \frac{1}{N} \sum_{i=1}^N l(f(\mathbf{x}_i; \mathbf{w}), y_i)$

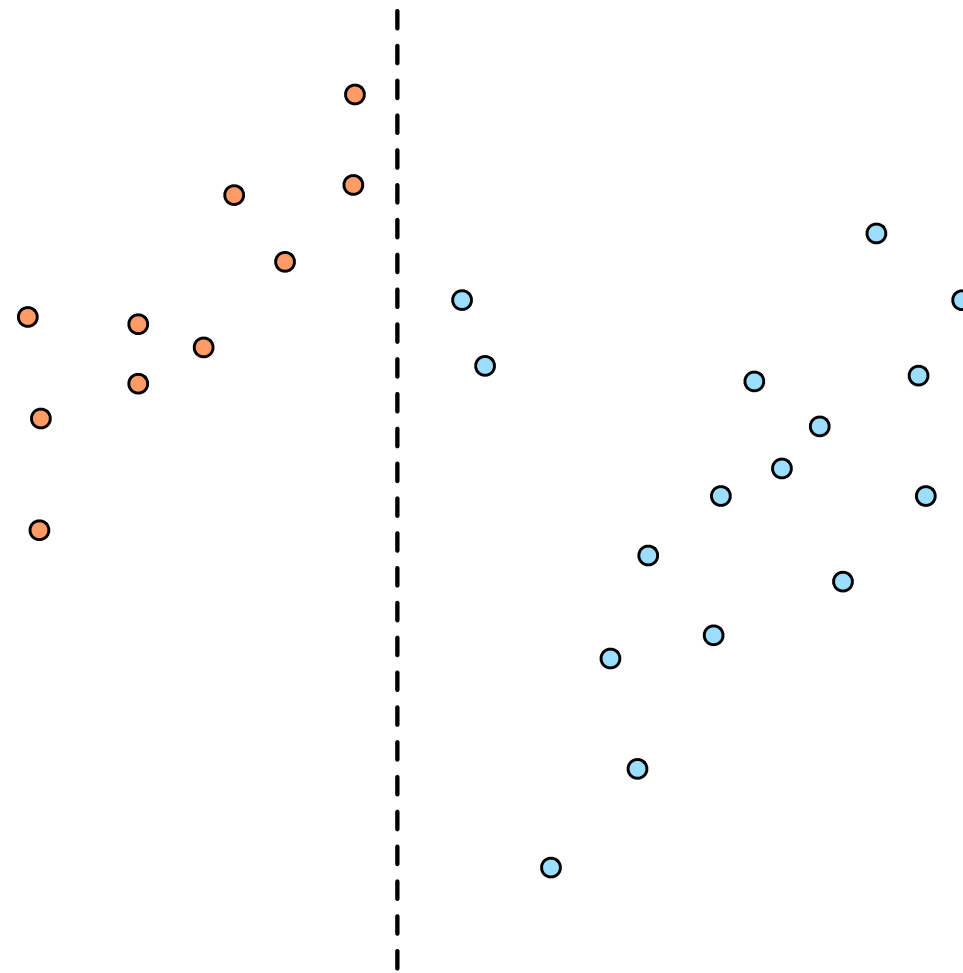
Переобучение



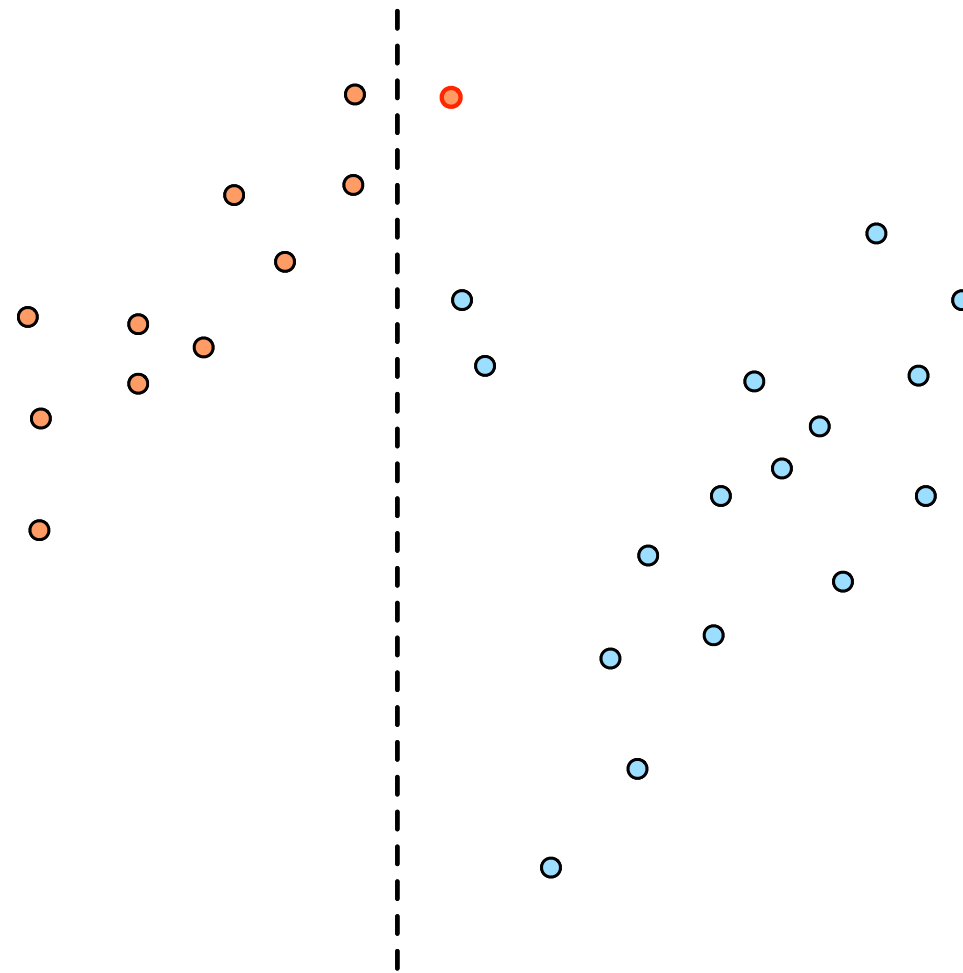
переобучение правило

реальная
закономерность

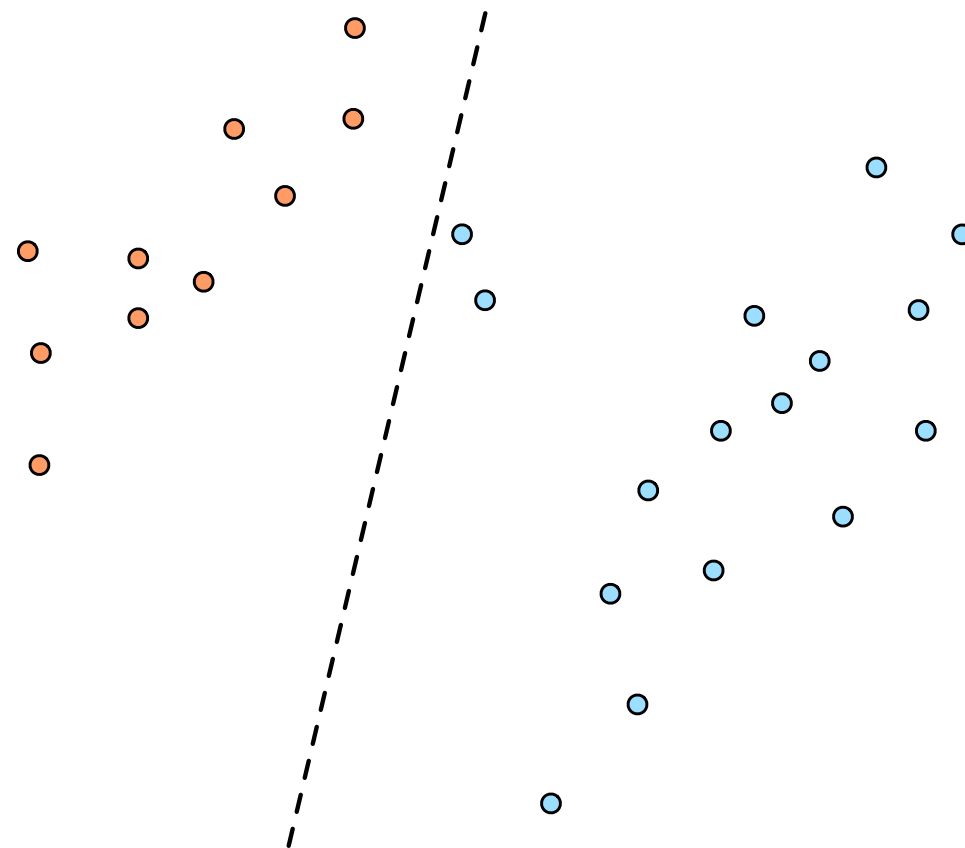
Переобучение



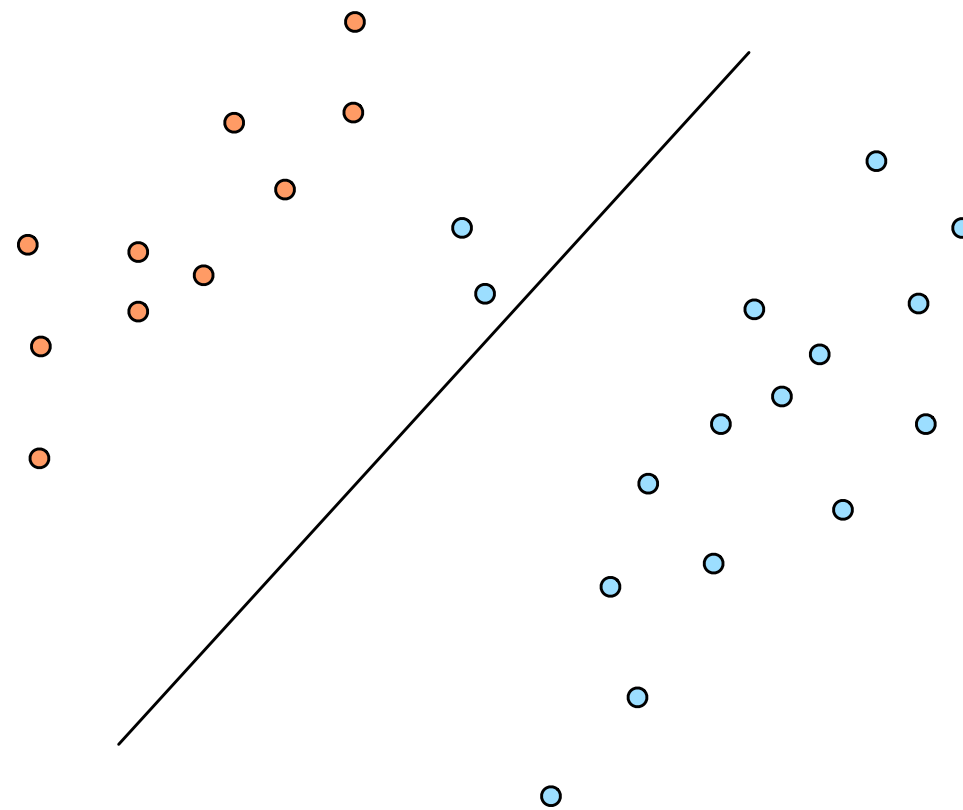
Переобучение



Переобучение



Переобучение



Решение: **регуляризация** решающего правила

$$f^* = \arg \min_{f \in F} \frac{1}{N} \sum_{i=1}^N l(f(\mathbf{x}_i), y_i) + r(f)$$

Насколько обучение надежно?

Эмпирический риск $\mathcal{R}_N(f) = \frac{1}{N} \sum_{i=1}^N l(f(\mathbf{x}_i), y_i)$

Ожидаемый риск $\mathcal{R}(f) = \mathbb{E}_{(\mathbf{x}, y)} l(f(\mathbf{x}), y)$

Насколько обучение надежно?

Эмпирический риск $\mathcal{R}_N(f) = \frac{1}{N} \sum_{i=1}^N l(f(\mathbf{x}_i), y_i)$

Ожидаемый риск $\mathcal{R}(f) = \mathbb{E}_{(\mathbf{x}, y)} l(f(\mathbf{x}), y)$

$$f_N^* = \arg \min_f \mathcal{R}_N(f)$$

$$f^* = \arg \min_f \mathcal{R}(f)$$

$$f_N^* \xrightarrow{P} f^*$$

Регрессия

Пусть y - оценка за курс по машинному обучению,
а x_d - оценка за предыдущий курс d

Предположение: $y = \sum_{d=1}^D w_d x_d + w_0 + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$

Регрессия

Пусть y - оценка за курс по машинному обучению,
а x_d - оценка за предыдущий курс d

Предположение: $y = \sum_{d=1}^D w_d x_d + w_0 + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$

“Важность предмета”

“гарантированная”

оценка

не выпался,
строгий экзаменатор

Регрессия

Пусть y - оценка за курс по машинному обучению,
а x_d - оценка за предыдущий курс d

Предположение: $y = \sum_{d=1}^D w_d x_d + w_0 + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$

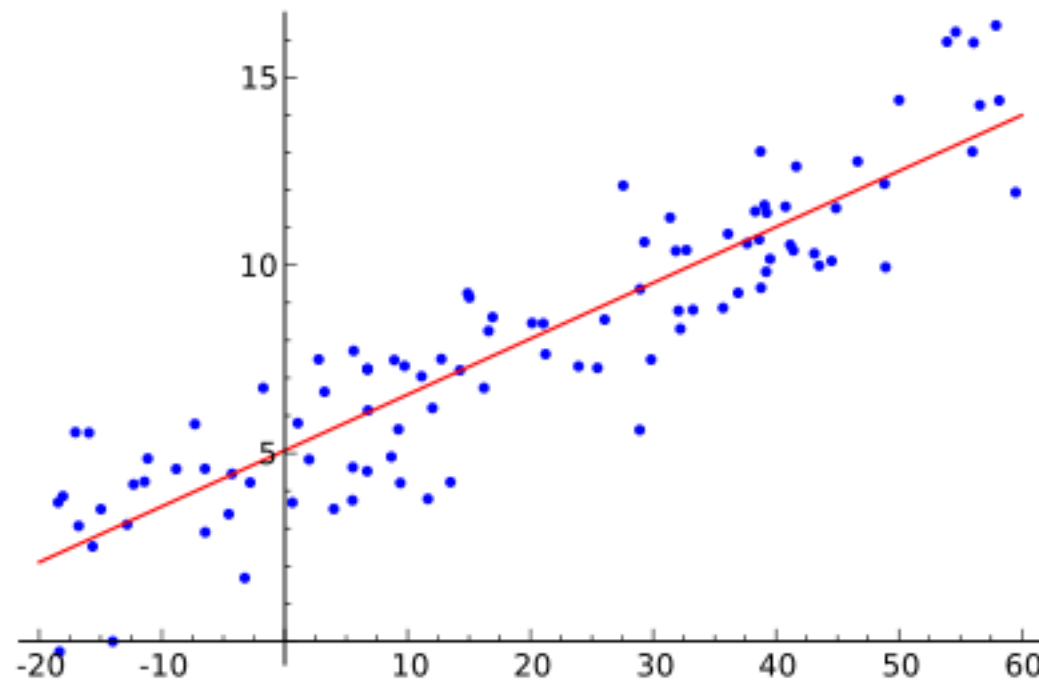
“Важность предмета”

“гарантированная”
оценка

не выпался,
строгий экзаменатор

Как настроить веса?

Регрессия



Решающее правило: $f(\mathbf{x}; \mathbf{w}) = \sum_{d=1}^D w_d x_d + w_0$

Функция потерь: $l(\hat{y}, y) = (\hat{y} - y)^2$

Обучение: $\mathbf{w}^* = \arg \min_{\mathbf{w}} \frac{1}{N} \sum_{i=1}^N l(f(\mathbf{x}_i; \mathbf{w}), y_i)$

Кластеризация

Пусть есть N измерений:

$$X = \{\mathbf{x}_i\}_{i=1}^N, \quad \mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iD})$$

Кластеризация

Пусть есть N измерений:

$$X = \{\mathbf{x}_i\}_{i=1}^N, \quad \mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iD})$$


длина хвоста

вес

Кластеризация

Пусть есть N измерений:

$$X = \{\mathbf{x}_i\}_{i=1}^N, \quad \mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iD})$$


длина хвоста

вес

Предположения

- есть K видов крокодилов
- каждый вид описывается средними показателями

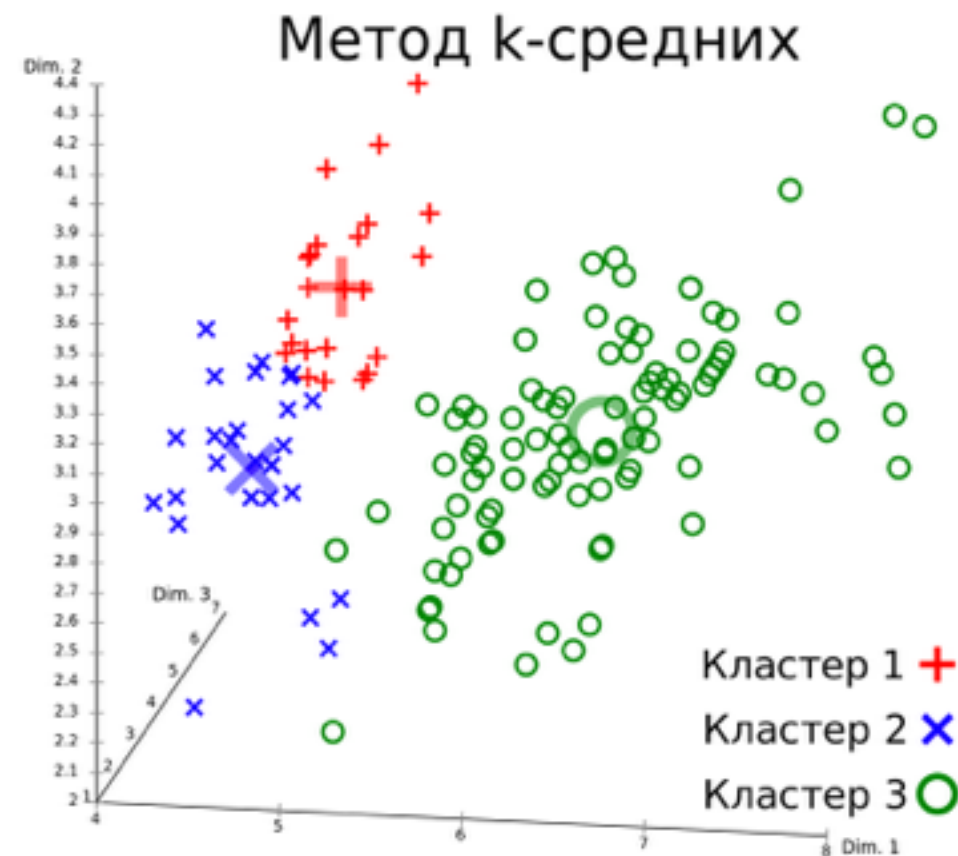
$$\mu_k = (\mu_{k1}, \mu_{k2}, \dots, \mu_{kD})$$

- принадлежность виду можно определить с помощью расстояния до средних показателей

Кластеризация

$z_i = k$ - крокодила номер i имеет вид k

Задача:
$$\min_{\mu, \mathbf{z}} \sum_{k=1}^K \sum_{z_i=k} \|\mathbf{x}_i - \mu_k\|_2^2$$



Коллаборативная фильтрация

	Бегущий по лезвию бритвы	Евротур	Звездные войны	Операция Ы	Стражи галактики
Олег		2	7	1	8
Юля		8		7	7
Федя	9			2	2
Женя	9	7			7

Матрица рейтингов R

Матричное разложение

Низкоранговое разложение матрицы рейтингов

$$R \approx U^T V \quad R \in \mathbb{R}^{N \times M}, U \in \mathbb{R}^{K \times N}, V \in \mathbb{R}^{K \times M}$$

Матричное разложение

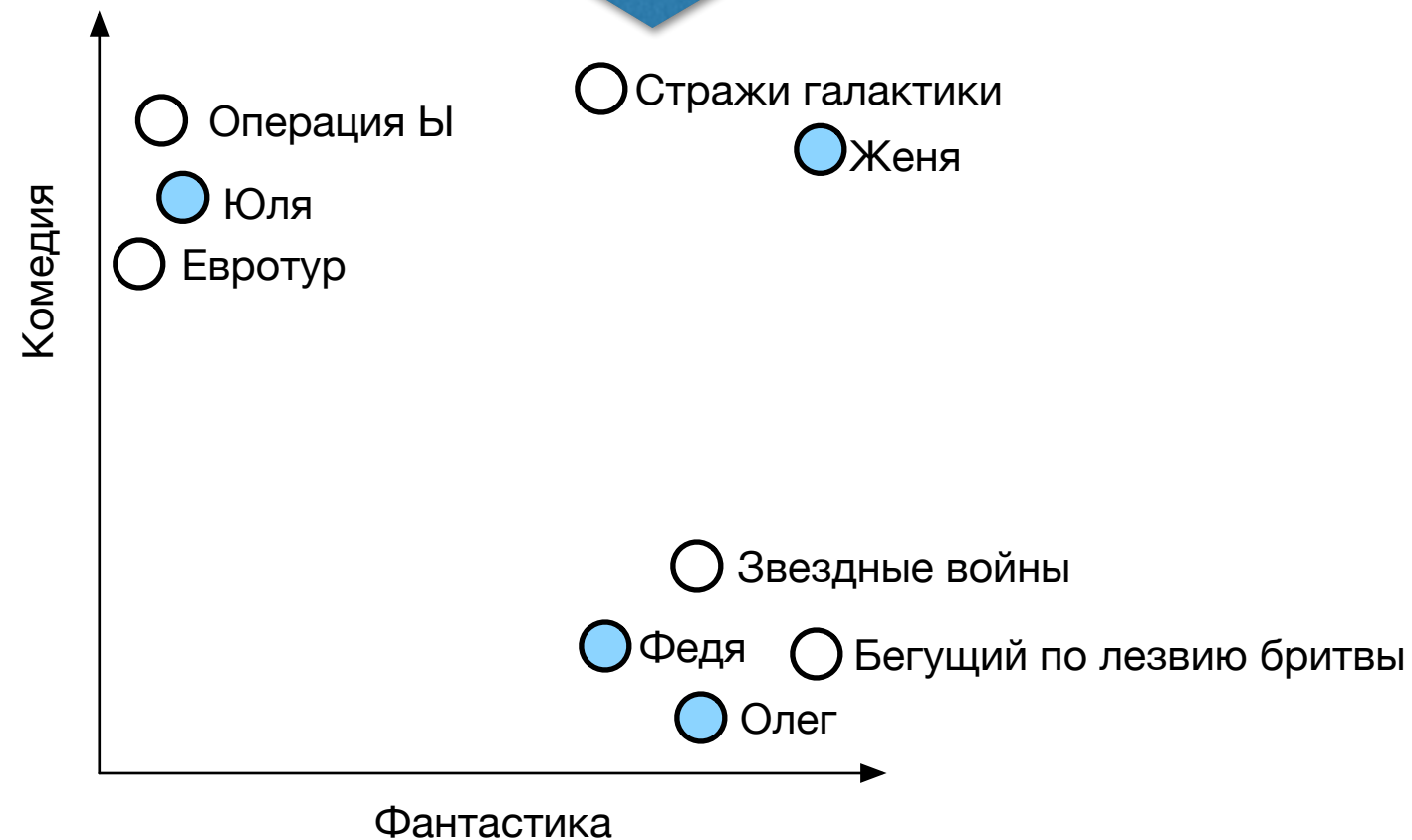
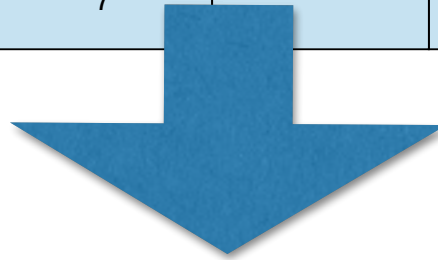
Низкоранговое разложение матрицы рейтингов

$$R \approx U^T V \quad R \in \mathbb{R}^{N \times M}, U \in \mathbb{R}^{K \times N}, V \in \mathbb{R}^{K \times M}$$

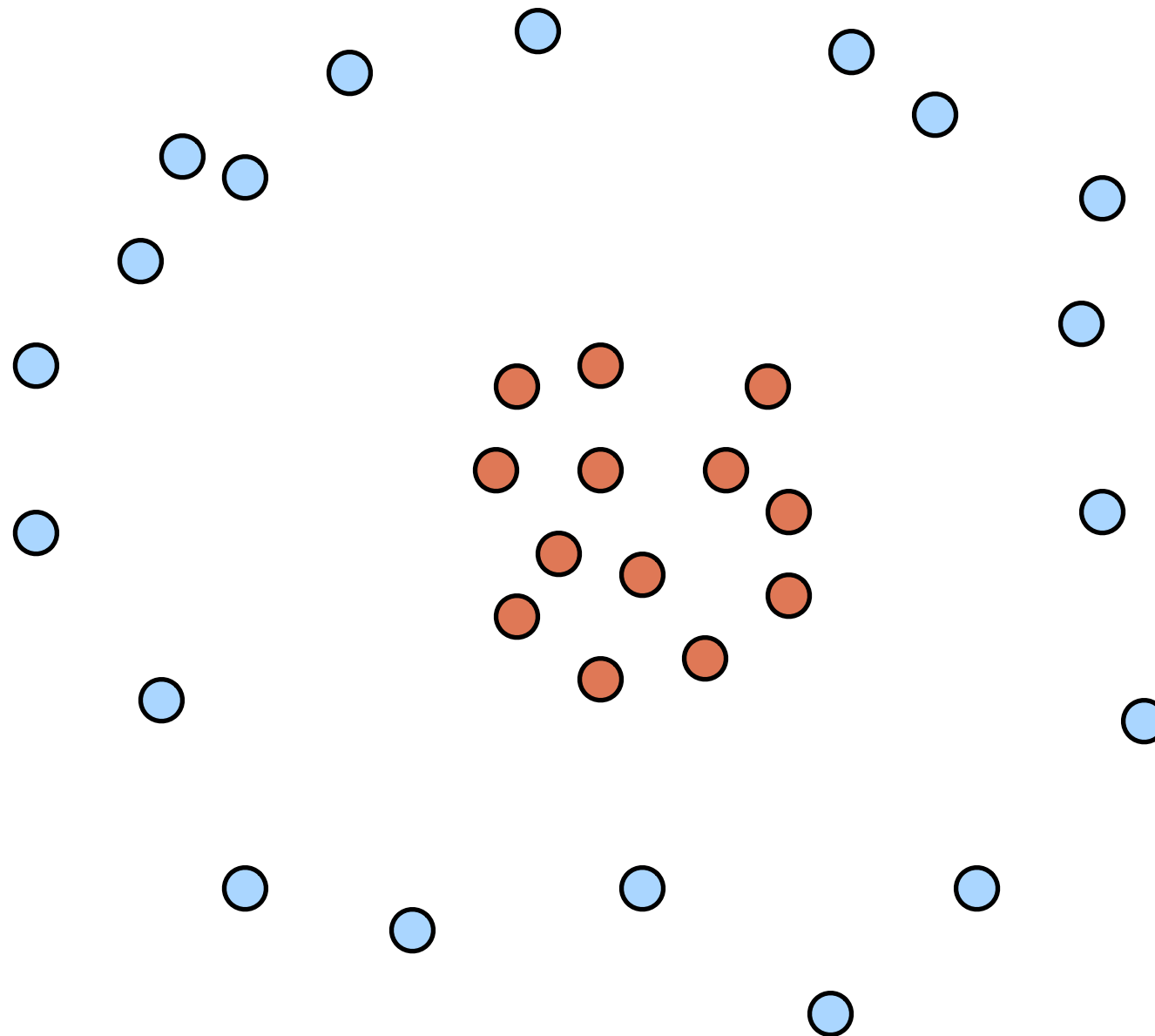
Задача оптимизации: $\min_{U, V} \sum_{i, j} (R_{ij} - \sum_{k=1}^K U_{ik} V_{jk})^2$

Латентное пространство

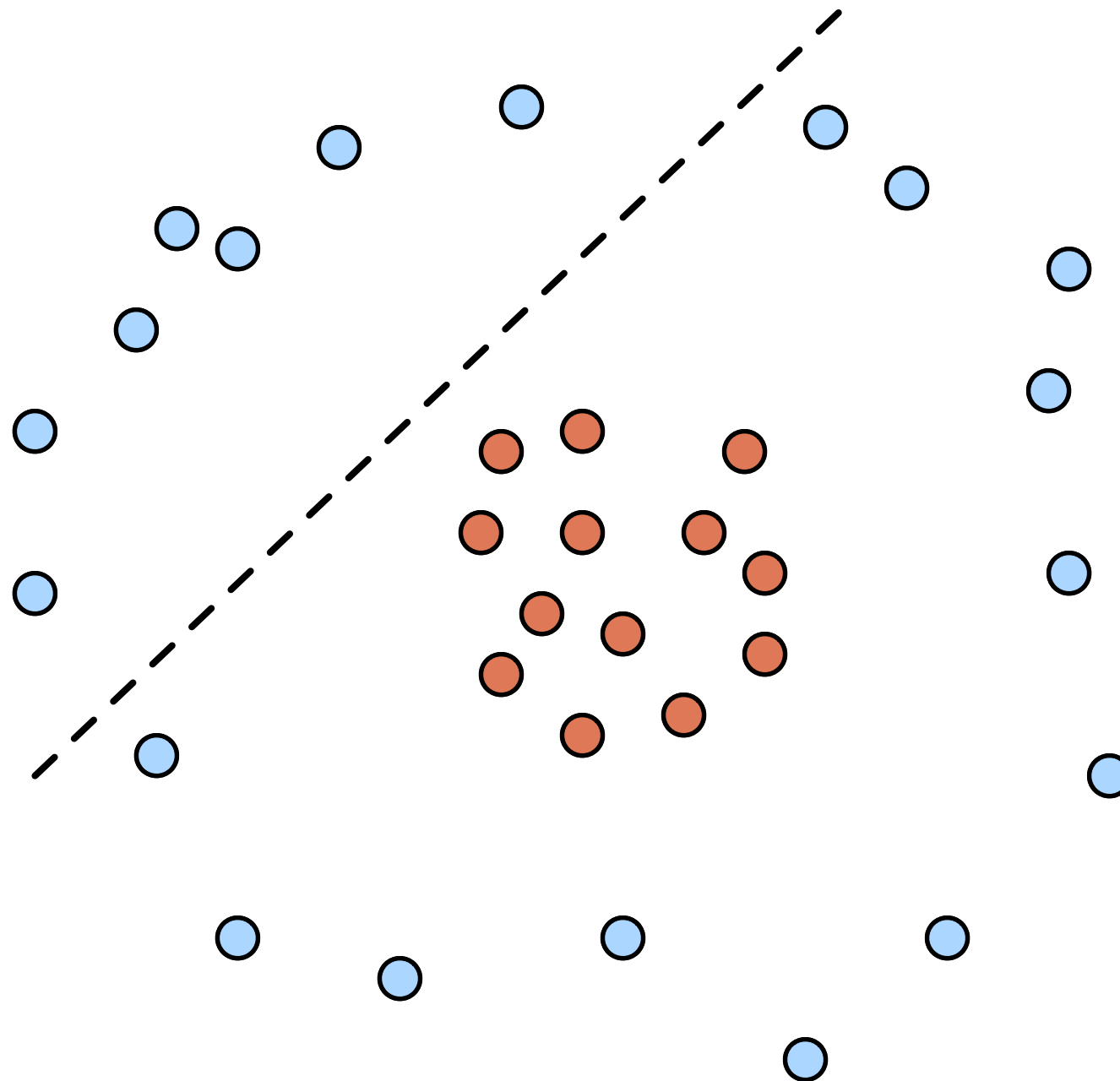
	Бегущий по лезвию бритвы	Евротур	Звездные войны	Операция Ы	Стражи галактики
Олег		2	7	1	8
Юля		8		7	7
Федя	9			2	2
Женя	9	7			7



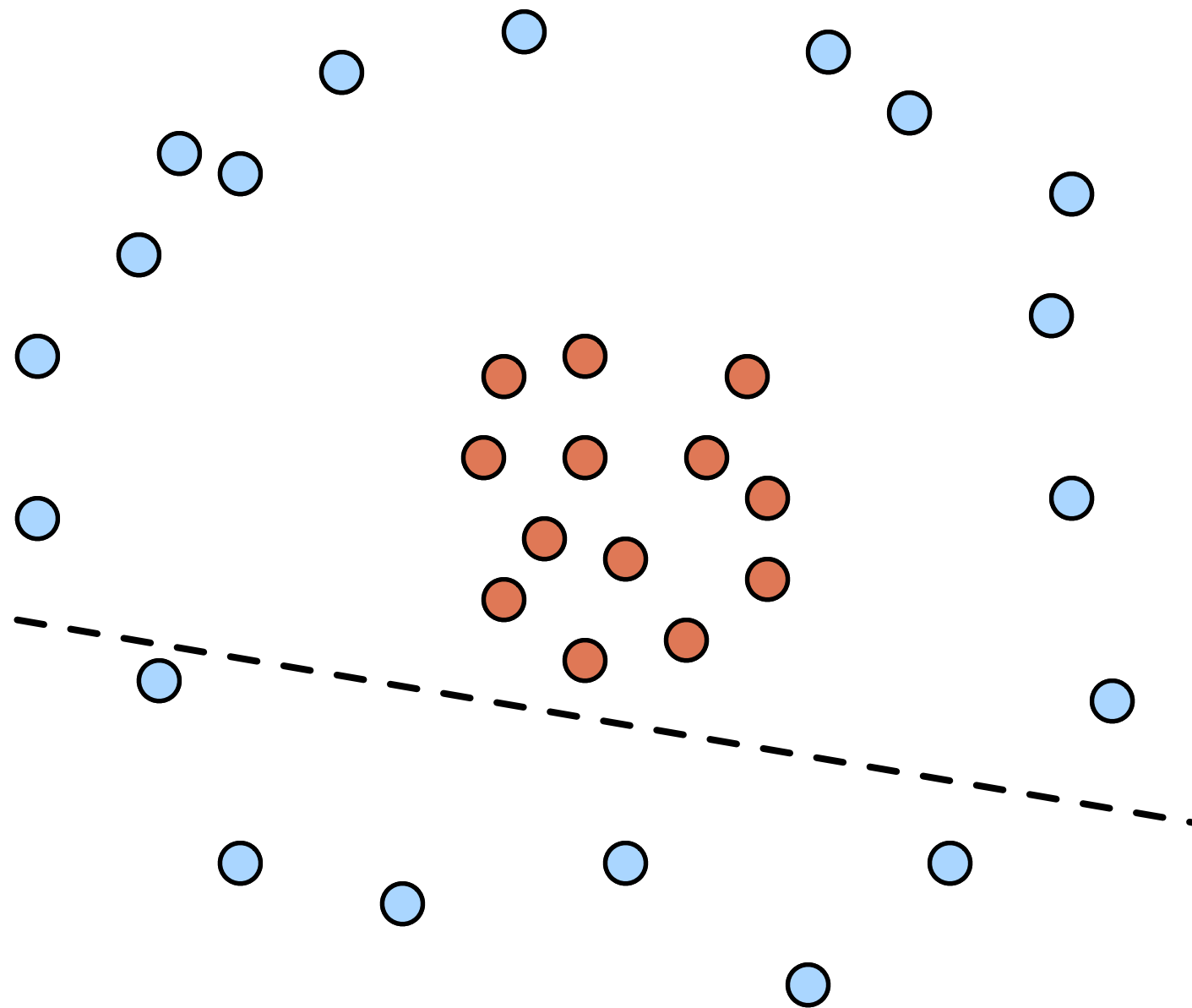
Метаобучение и усиление слабых алгоритмов



Метаобучение и усиление слабых алгоритмов



Метаобучение и усиление слабых алгоритмов



Метаобучение и усиление слабых алгоритмов

Проинициализируем важность каждого объекта:

$$w_i^{(0)} = 1/N$$

Определим ошибку на итерации

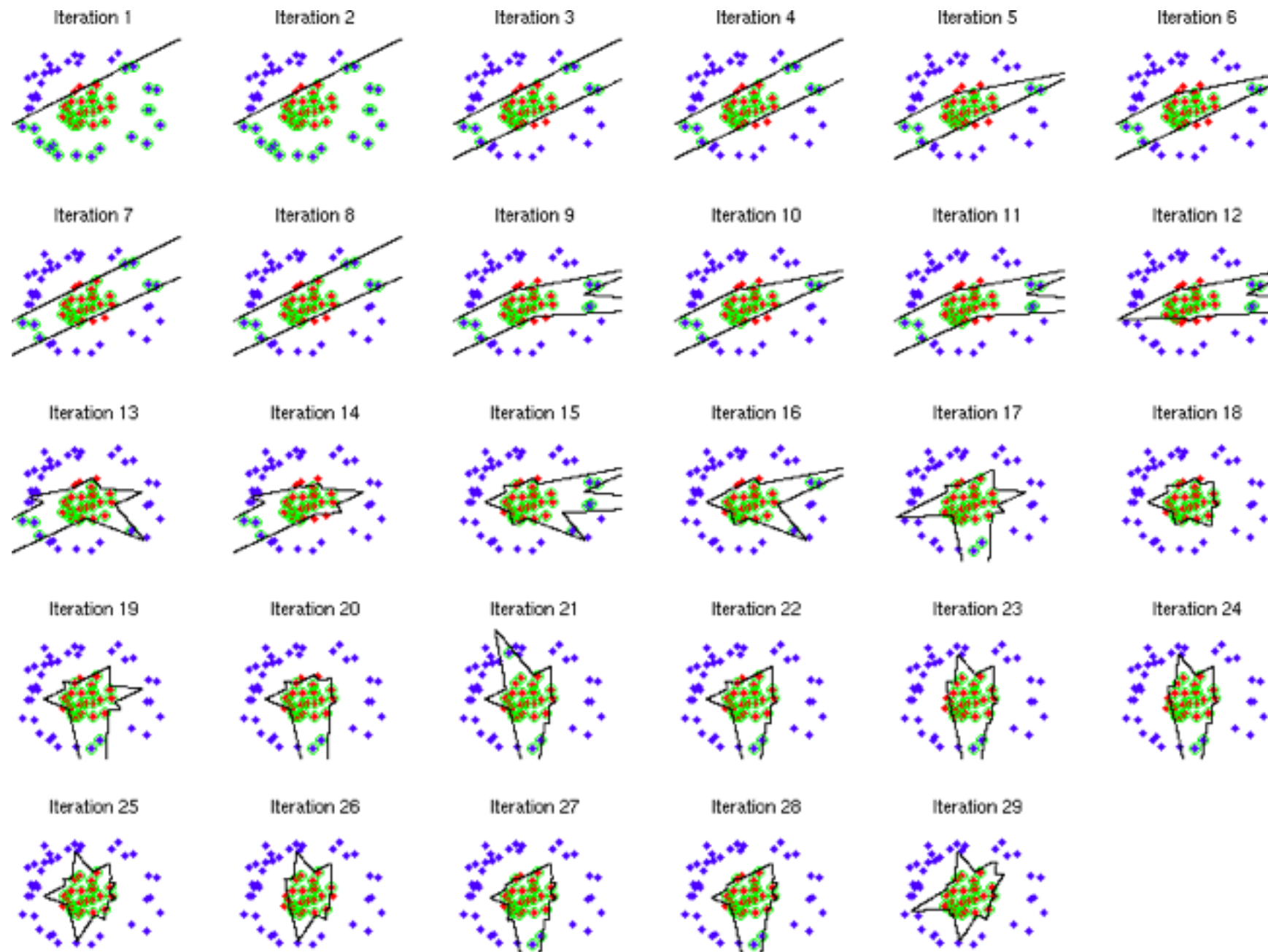
$$\epsilon_t(f) = \sum_{i=1}^N w_i^{(t)} \exp(-y_i f(\mathbf{x}_i))$$

На каждой итерации $t = 1, \dots, T$:

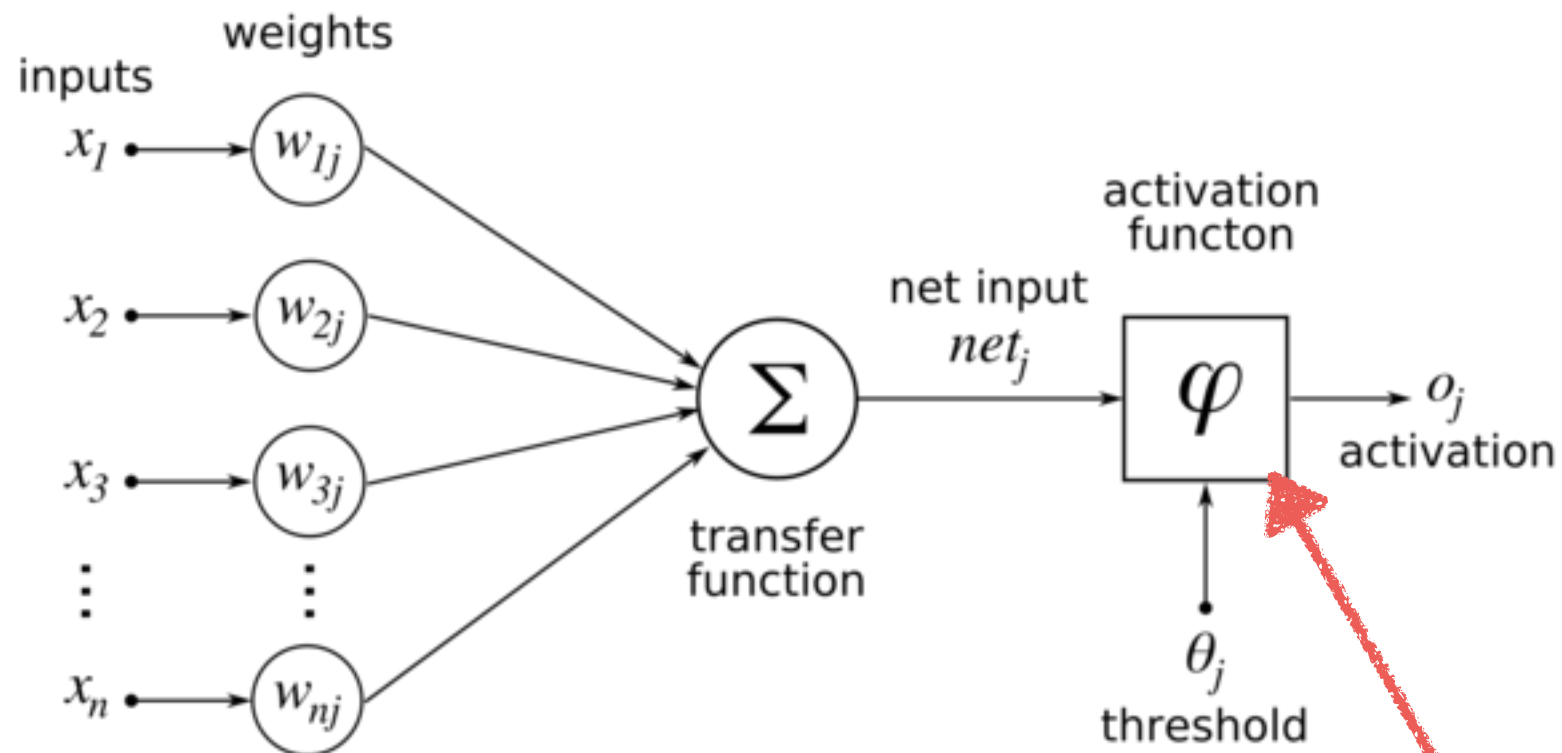
- Обучим слабый классификатор $f_t = \arg \min_{f \in \mathbb{F}} \epsilon_t(f)$
- Выберем вес $\alpha_t = \frac{1}{2} \log\left(\frac{1 - \epsilon_t(f_t)}{\epsilon_t(f_t)}\right)$
- Пересчитаем веса $w_i^{(t)} = w_i^{(t-1)} \exp(-y_i \alpha_t f_t(\mathbf{x}_i))$

Сильный классификатор: $\hat{f}_t = \sum_t \alpha_t f_t$

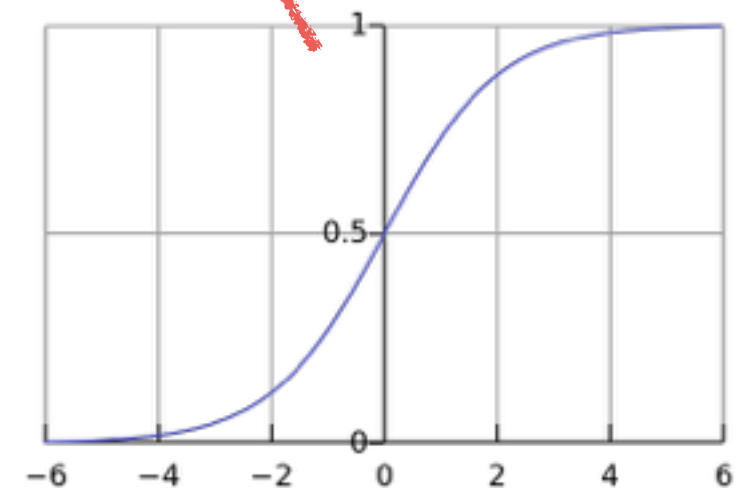
Метаобучение и усиление слабых алгоритмов



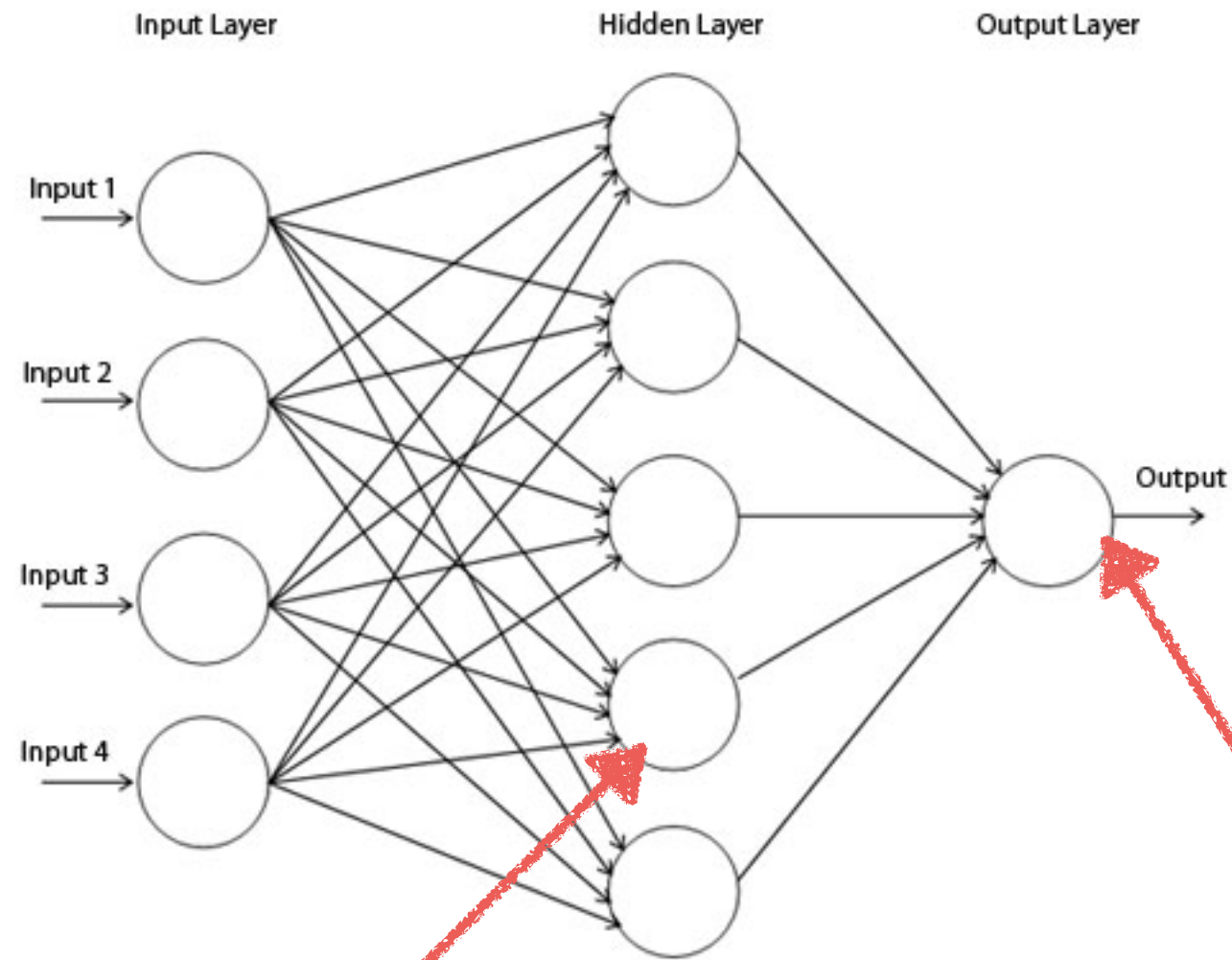
Модель нейрона



$$h = \sigma\left(\sum_{d=1}^D w_d x_d + w_0\right)$$



Перцептрон



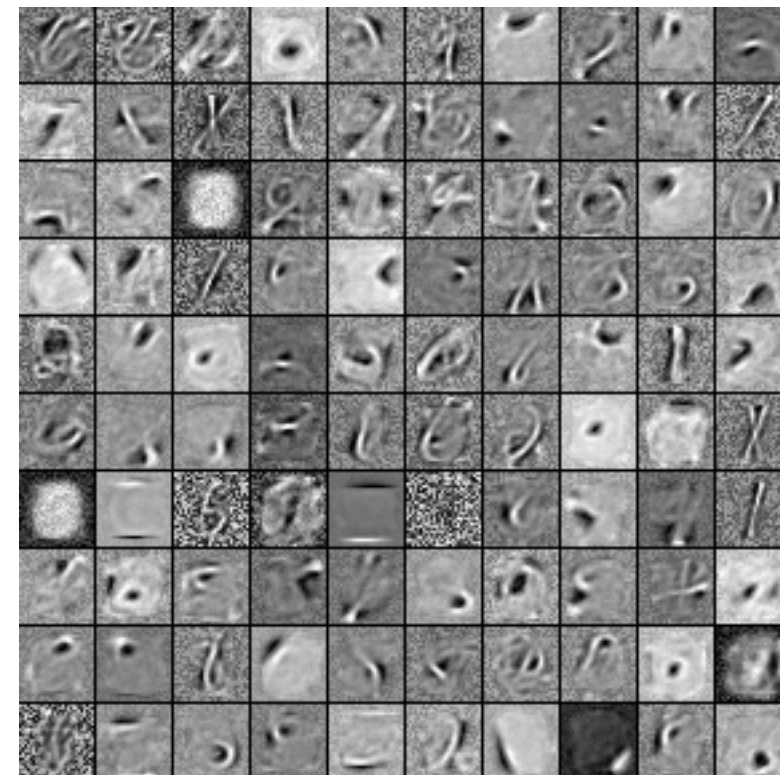
$$h_j = \sigma\left(\sum_{d=1}^D w_{jd}x_d + w_{j0}\right)$$

$$y = \sigma\left(\sum_{j=1}^K v_j y_j + v_0\right)$$

Перцептрон



данные



значения на скрытых слоях

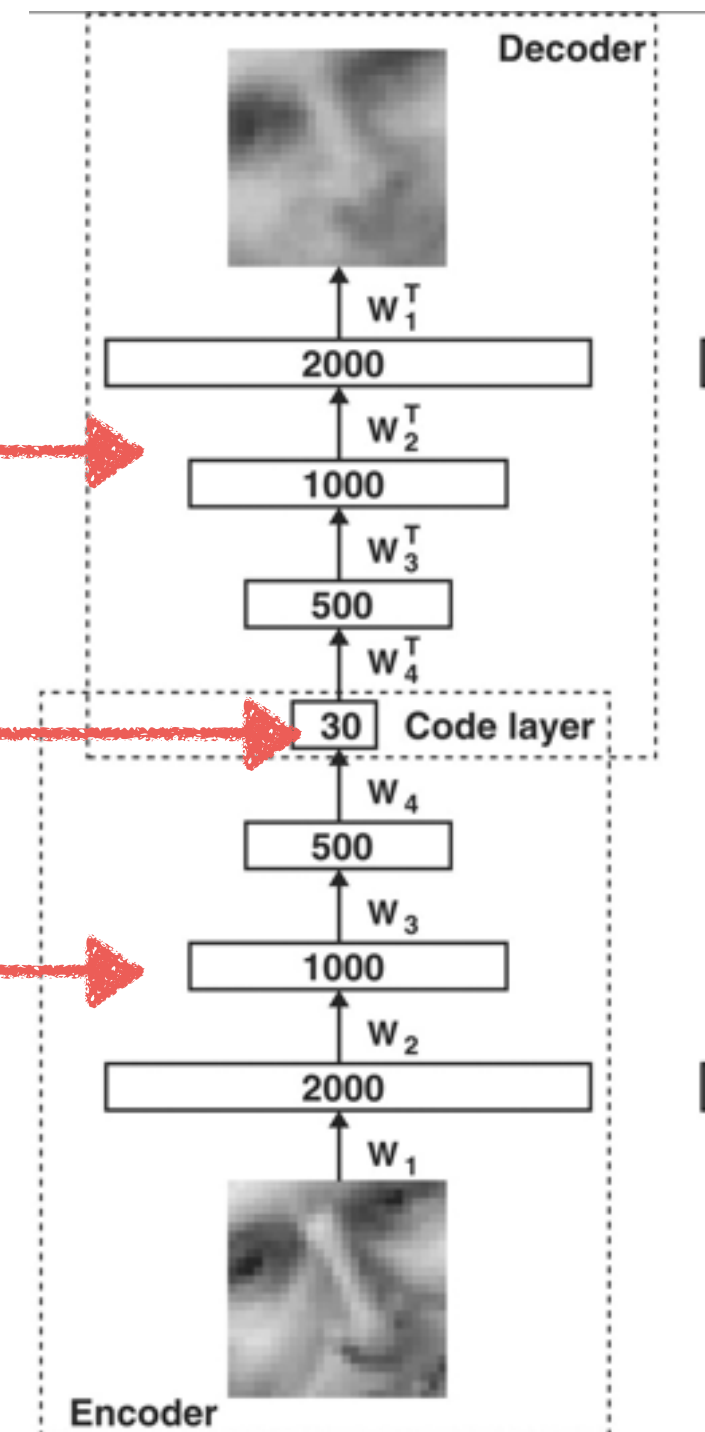
АВТОКОДИРОВЩИК

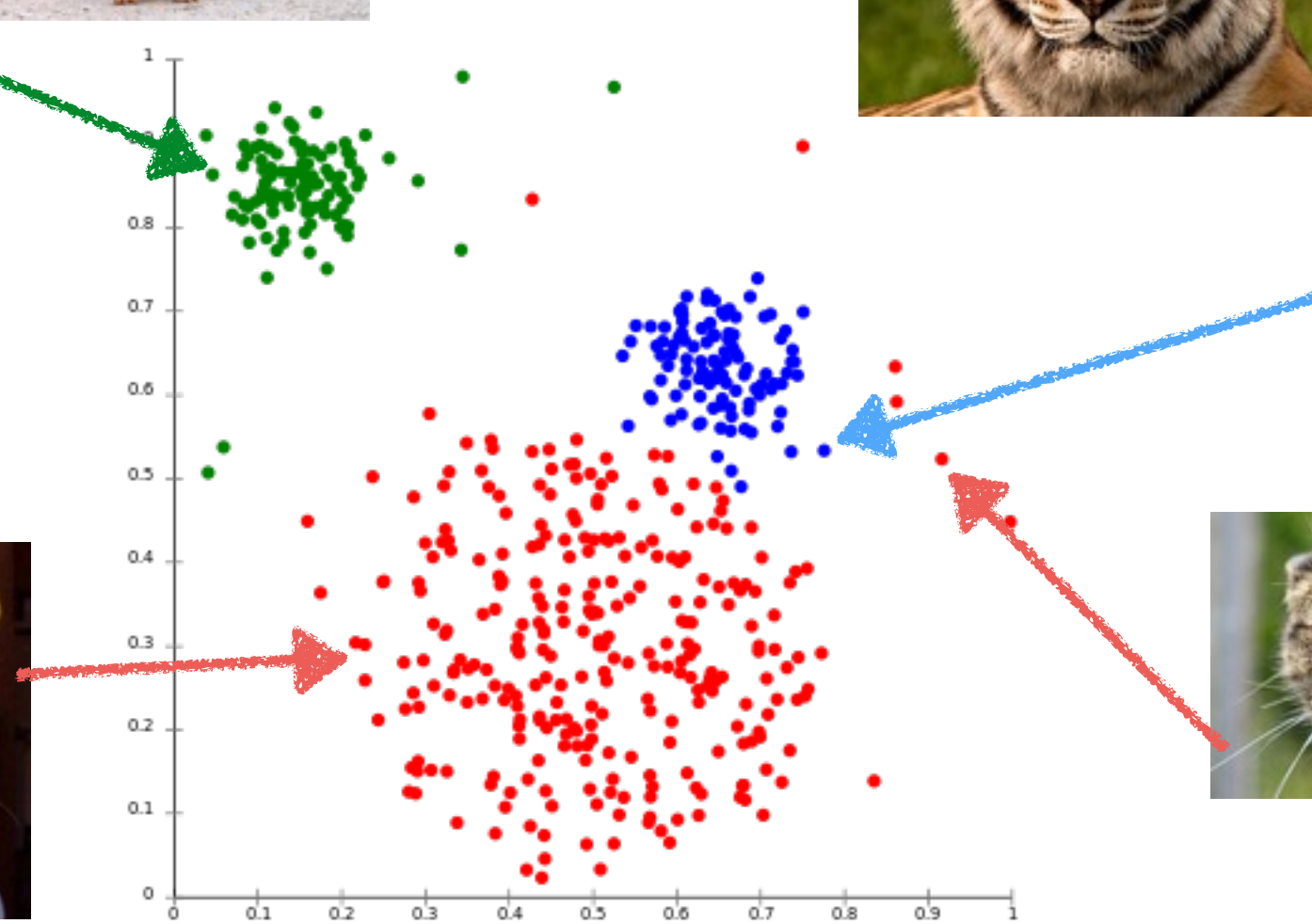
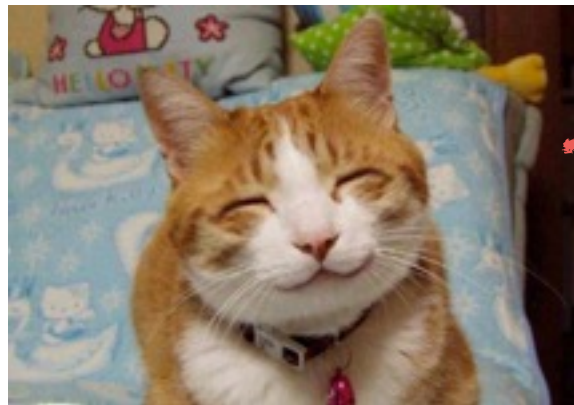
декодировщик

КОД

кодировщик

Цель:
$$\min_W \sum_{i=1}^N l(\mathbf{x}_i, de(en(\mathbf{x}_i)))$$





Обучение в реальном времени

В очередной момент времени t :

- Среда задает вопрос $\mathbf{x}_t$
- Алгоритм дает ответ $\hat{y}_t$
- Среда сообщает правильный ответ y_t

Цель: минимизировать $\text{regret}_{\mathbb{F}}(T) = \sum_{t=1}^T l(\hat{y}_t, y_t) - \min_{f \in \mathbb{F}} \sum_{t=1}^T l(f(\mathbf{x}_t), y_t)$

Обучение в реальном времени

В очередной момент времени t :

- Среда задает вопрос $\mathbf{x}_t$
- Алгоритм дает ответ $\hat{y}_t$
- Среда сообщает правильный ответ y_t

Цель: минимизировать $\text{regret}_{\mathbb{F}}(T) = \sum_{t=1}^T l(\hat{y}_t, y_t) - \min_{f \in \mathbb{F}} \sum_{t=1}^T l(f(\mathbf{x}_t), y_t)$

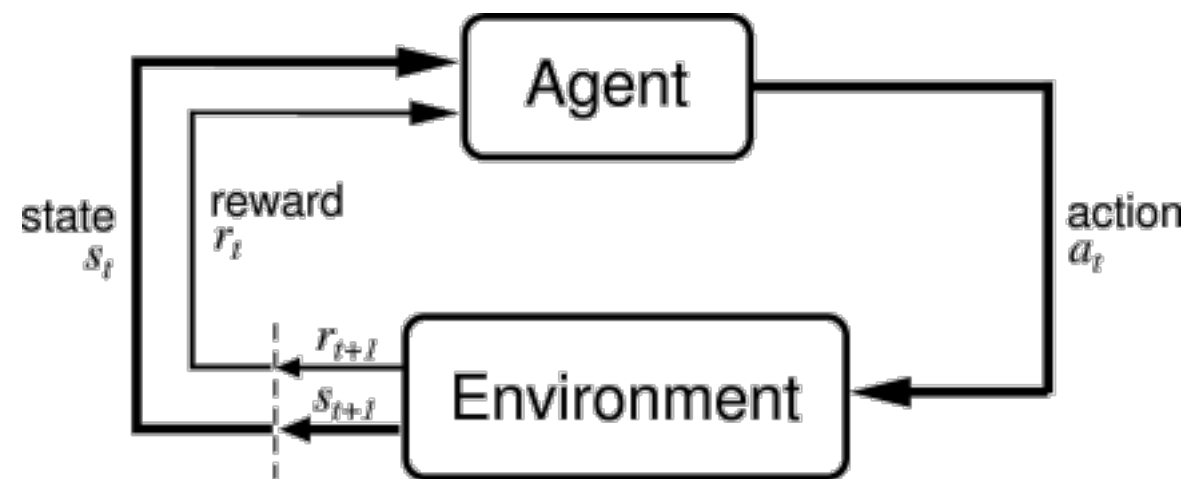
В случае $|\mathbb{F}| < \infty$ можно показать, что $\text{regret}_{\mathbb{F}}(T) \leq \sqrt{0.5 \log(|\mathbb{F}|)T}$

если среда не имеет доступа к нашему генератору случайных чисел
и $y_t = f^*(\mathbf{x}_t)$, $f^* \in \mathbb{F}$

Обучение с подкреплением

В очередной момент времени t :

- Алгоритм измеряет состояние $s_t \in \mathcal{S}$
- С помощью стратегии π алгоритм выбирает действие $a_t \in \mathcal{A}$
- Алгоритм получает вознаграждение r_t



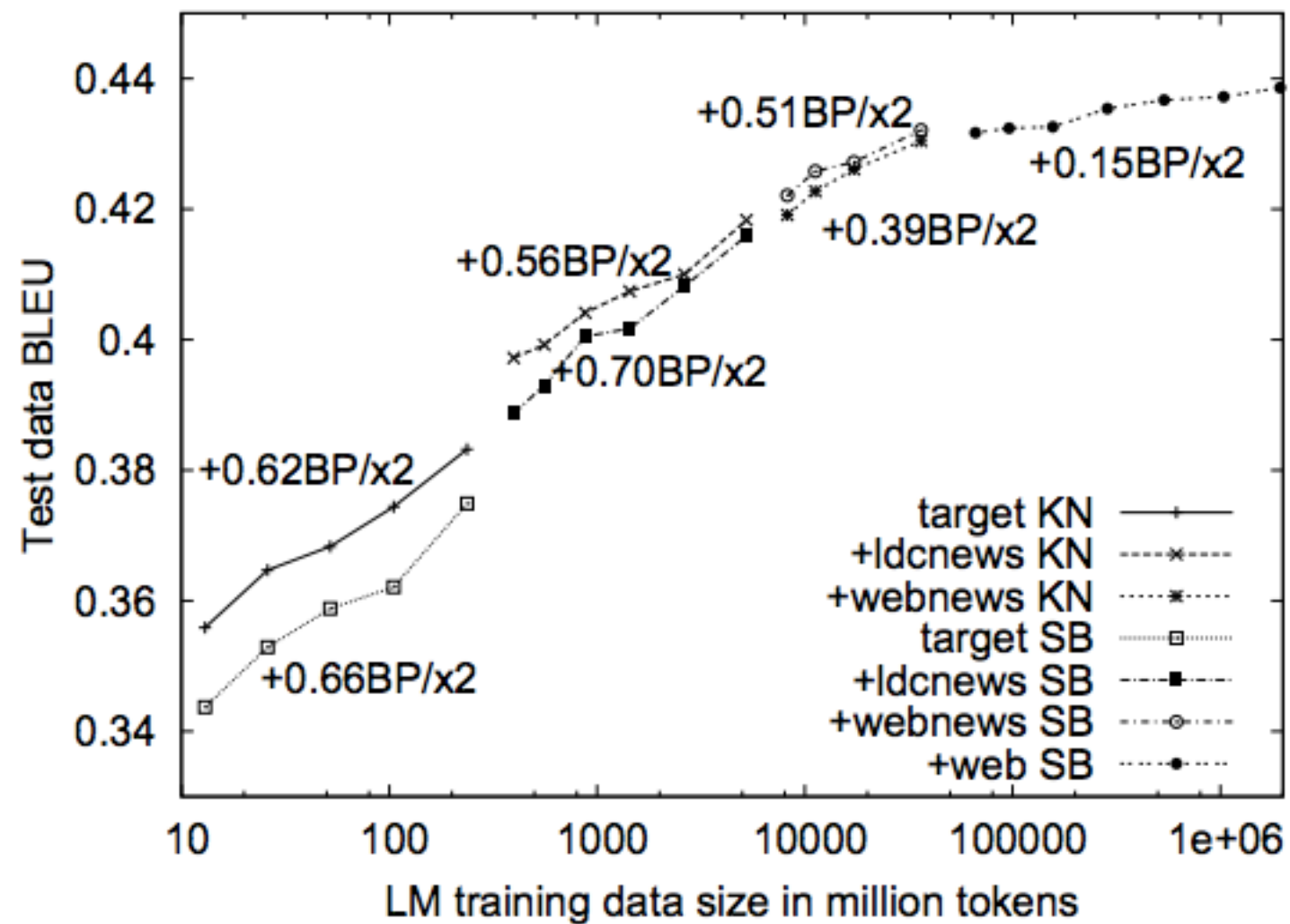
Цель: $\max_{\pi} \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t r_t | \pi \right]$

Big data?

Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...

Dan Ariely

Big data?



Корни машинного обучения

- Теория вероятностей и математическая статистика
- Линейная алгебра
- Математический анализ
- Методы оптимизации
- Теория игр
- Алгоритмы и структуры данных
- Параллельные вычисления