

Принципы активного обучения

Иванова Елизавета

Санкт-Петербург, 2016

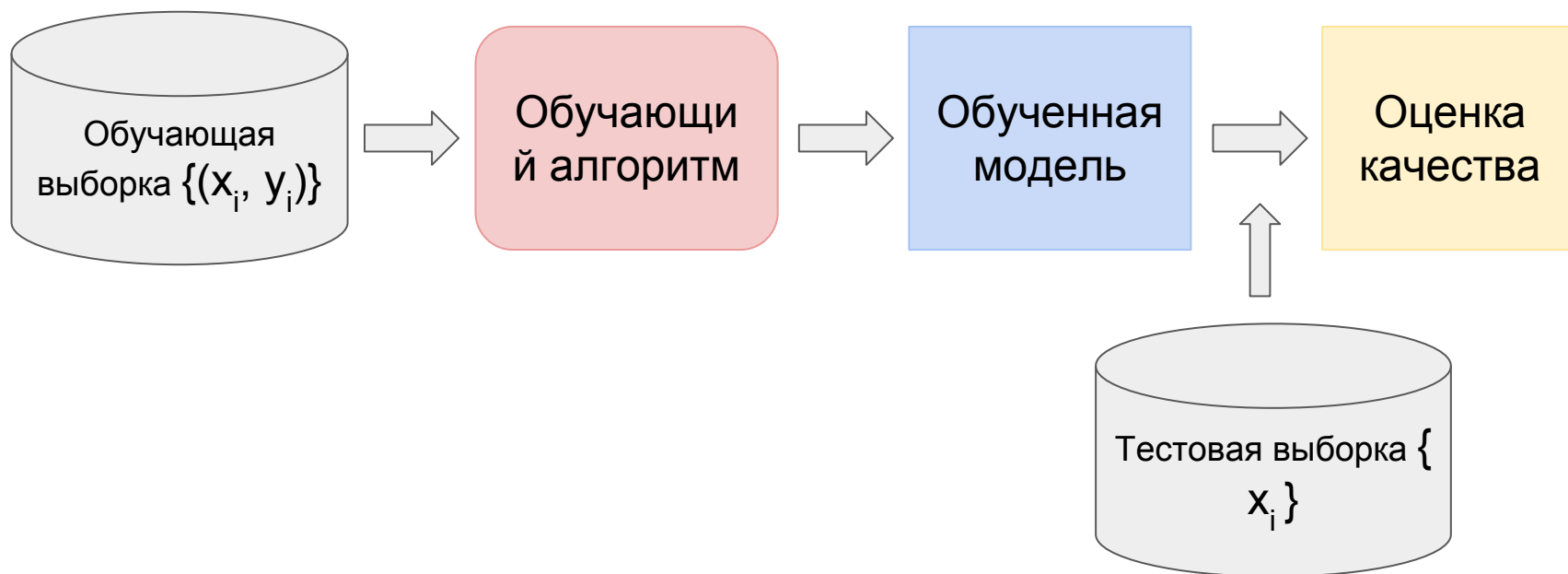
Стандартные задачи машинного обучения

- Обучение с учителем (supervised learning) - регрессия, классификация.
- Обучение без учителя (unsupervised learning) - кластеризация, редукция размерности.

Задача обучения с учителем

- X - множество объектов;
- Y - подмножество \mathbb{R} или конечное множество классов;
- $L = \{(x_1, y_1), \dots (x_n, y_n)\} \subset X \times Y$ - обучающая выборка;
- $t: X \rightarrow Y$ - целевая зависимость;
- Нужно построить $a: X \rightarrow Y$ по L , приближающую t .

Обучение с учителем



Проблема

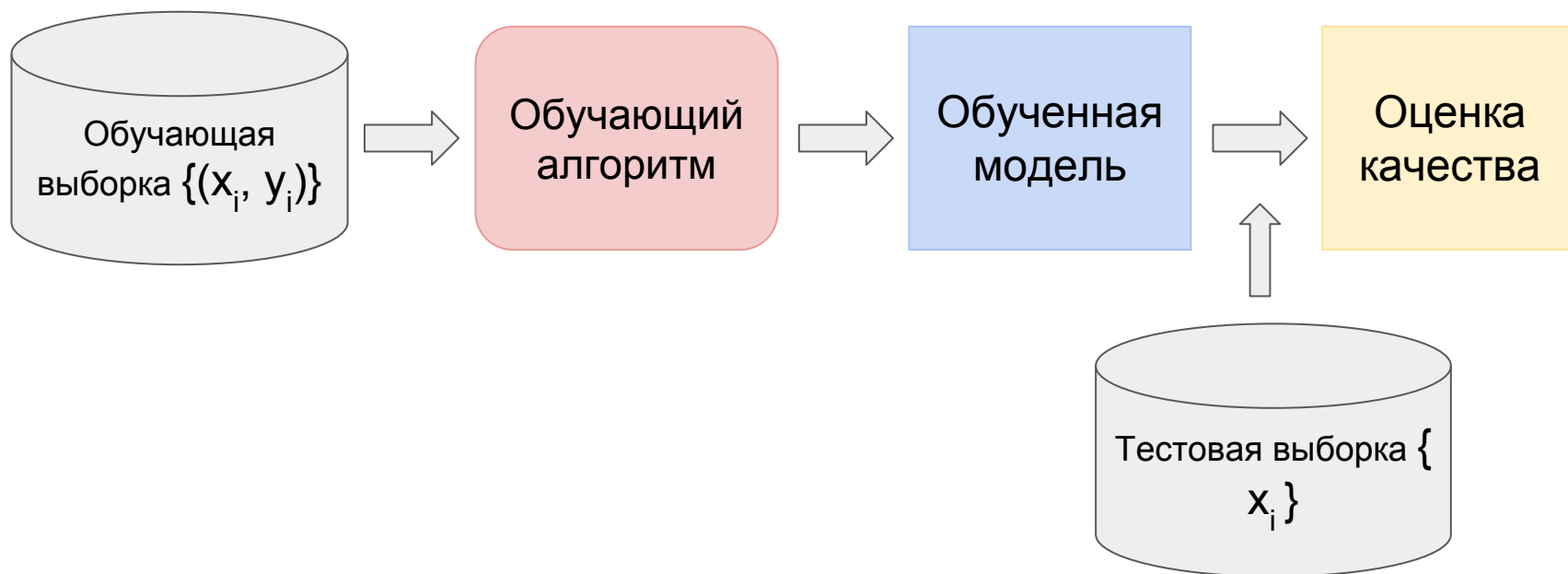
- Рассмотрим задачу классификации документов (изображений etc).
- В реальной жизни составить выборку из документов достаточно просто, а вот разметить их дорого и затратно по времени.

Проблема

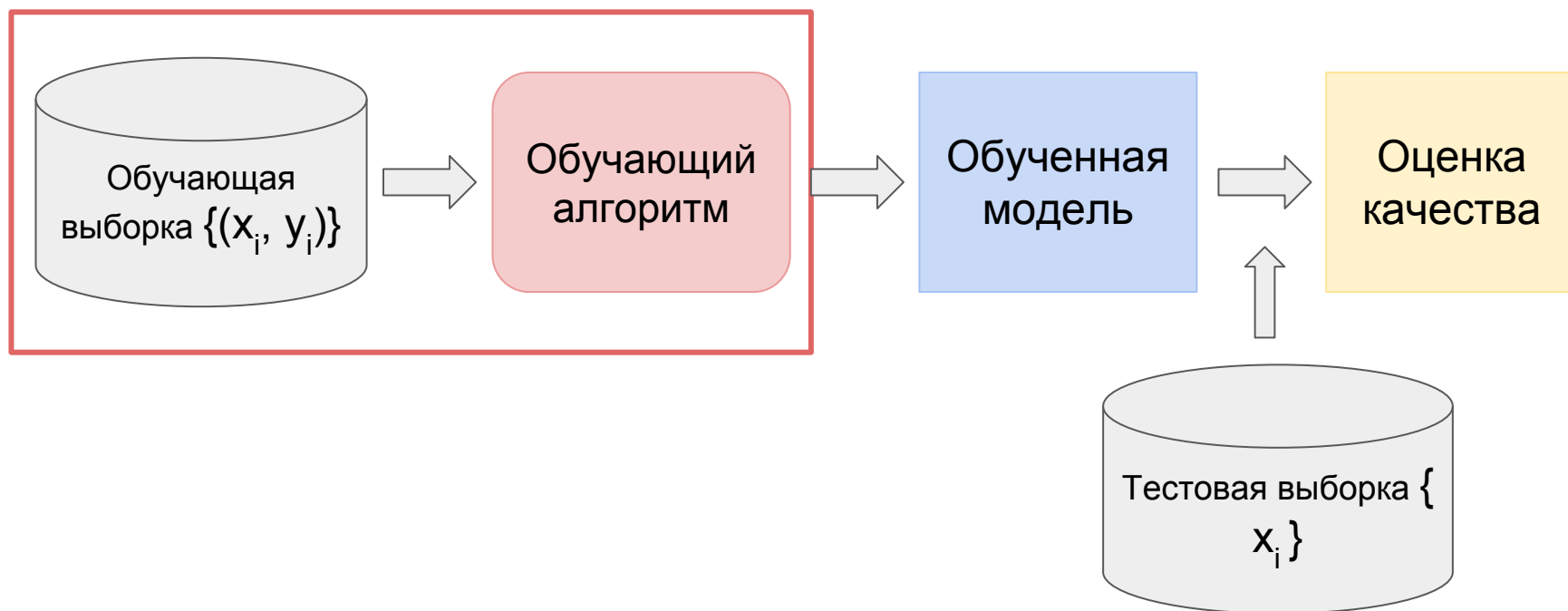
- Рассмотрим задачу классификации документов (изображений etc).
- В реальной жизни составить выборку из документов достаточно просто, а вот разметить их дорого и затратно по времени.

Будем тщательно выбирать данные, которые хотим разметить.

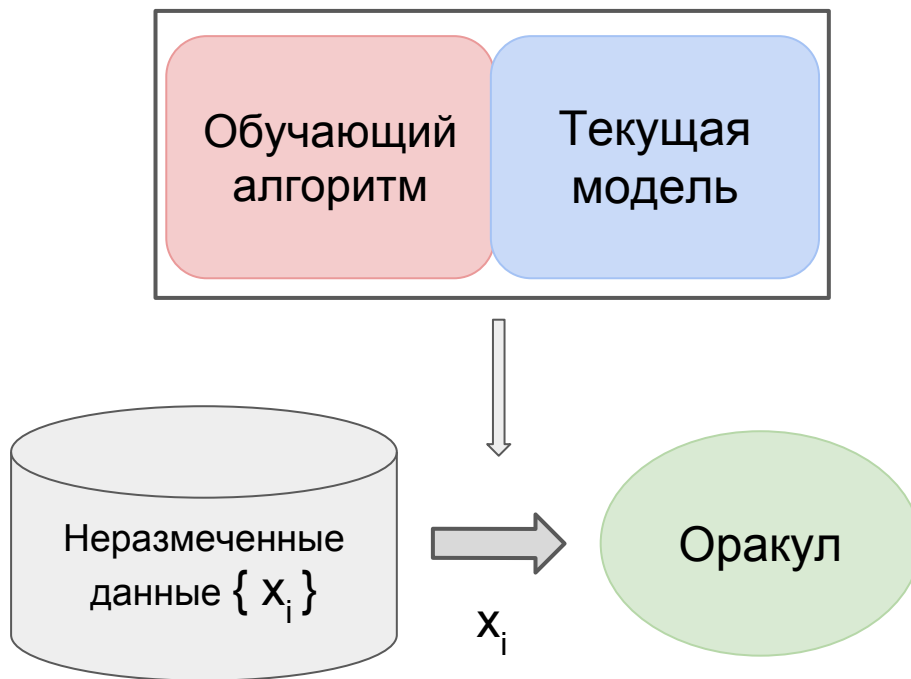
Обучение с учителем



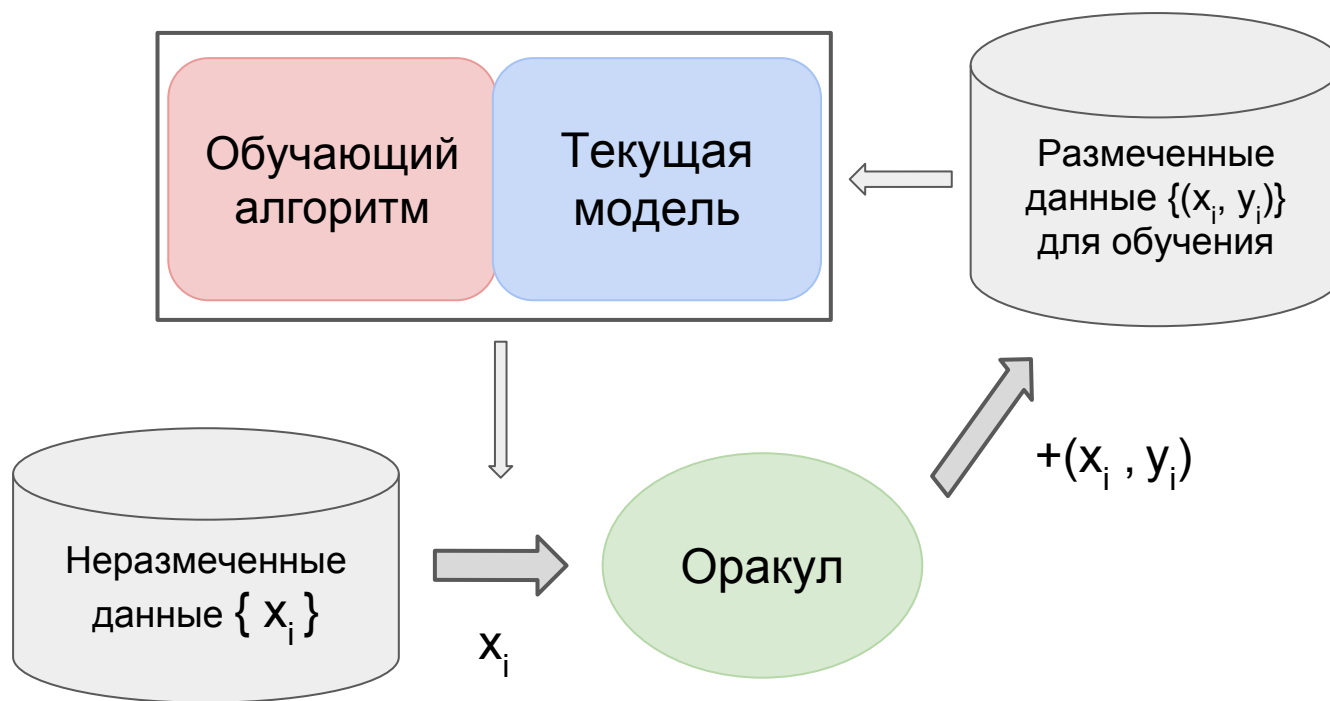
Обучение с учителем



Активное обучение (active learning)



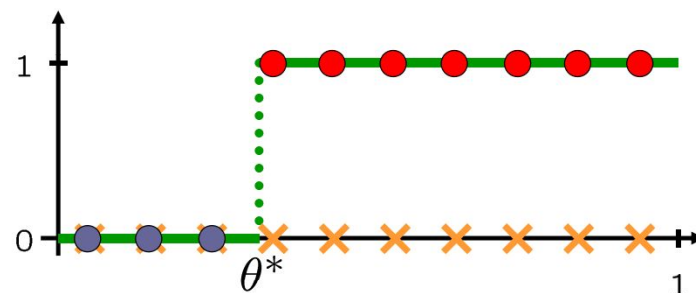
Активное обучение (active learning)



Пример

Задача линейного разделения:

- $g(x, \theta) = \mathbb{I}_{x > \theta}(x)$, $x \in [0, 1]$, $\theta \in [0, 1]$
- Нужно оценить θ



$$|\hat{\theta}_n - \theta^*| \sim \frac{1}{n}$$

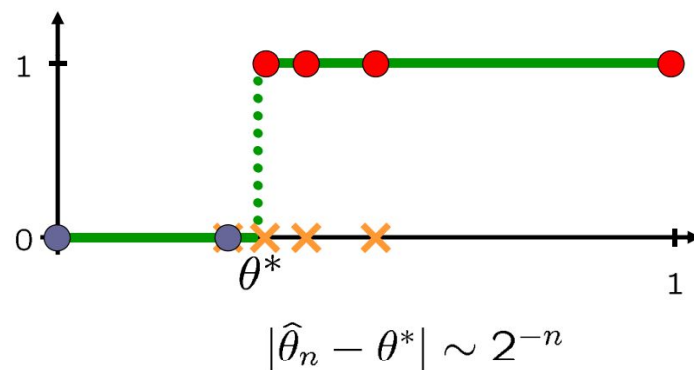
Стандартный подход: $L = \{ i/n \}_{i=1..n-1}$ - находим решение за $O(n)$ измерений.

Пример

Задача линейного разделения:

- $g(x, \theta) = \mathbb{I}_{x > \theta}(x)$, $x \in [0, 1]$, $\theta \in [0, 1]$
- Нужно оценить θ

Двоичный поиск: $O(\log n)$ измерений.

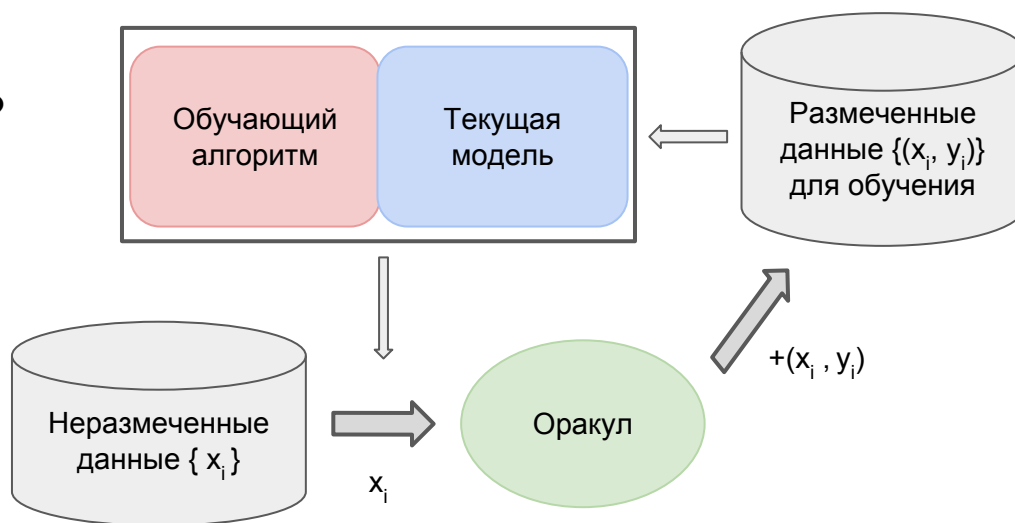


Типы активного обучения

- Выбор объекта из коллекции (pool-based sampling)
- Выбор объекта из потока (stream-based selective sampling)
- Генерация запросов (query synthesis)

Естественные вопросы

- Как выбрать начальное множество $\{x_i\}$ для первой итерации обучения модели?
- Как выбрать следующий объект для разметки?



Выбор начального множества

- Равномерно из неразмеченных данных.
- Исходя из задачи. Например, для задач классификации предварительно сделать кластеризацию и взять центры кластеров.

Выбор следующего объекта

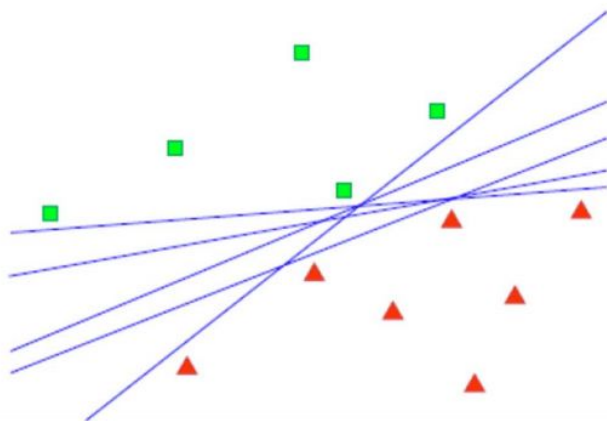
- Выбор по степени неуверенности (uncertainty sampling)
- Отбор комитетом (query by committee)
- Ожидаемое влияние на модель (expected model change)
- Ожидаемое уменьшение ошибки (expected error reduction)

Выбор по степени неуверенности

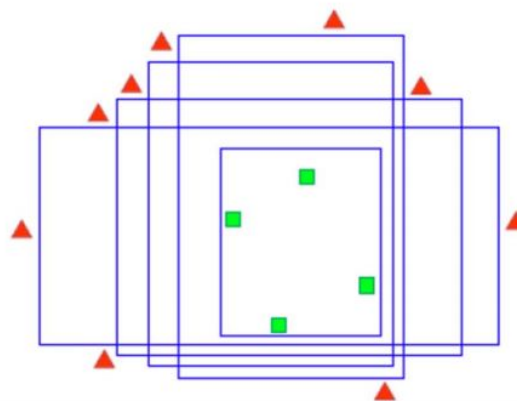
- U - множество неразмеченных x ;
- В задаче бинарной классификации $x^* = \operatorname{argmin}_{x \in U} | P_{\theta}(y|x) - 0.5 |$;
- В задаче мультиклассификации $x^* = \operatorname{argmax}_{x \in U} \varphi_{\theta}(x)$:
 - $\varphi_{\theta}(x) = 1 - P_{\theta}(y^*|x)$, $y^* = y^*(x)$ - наиболее вероятный класс для x ;
 - $\varphi_{\theta}(x) = P_{\theta}(y^*_2|x) - P_{\theta}(y^*_1|x)$, $y^*_i = y^*_i(x)$ - i -й вероятный класс для x ;
 - $\varphi_{\theta}(x) = - \sum_{y \in Y} P_{\theta}(y|x) \log P_{\theta}(y|x)$.

Отбор комитетом

- $C_L = \{\theta_1, \dots, \theta_m\}$ - комитет моделей, обученных на одном и том же множестве $L \subset X \times Y$.



(a)



(b)

Отбор комитетом

- $V(y, x) = |\{\theta \in C: y_\theta(x) = y\}|$ - количество моделей из комитета C , выбравших y ;
- $P^*(y|x) = V(y, x) / |C|$;
- $\varphi_\theta(x) = - \sum_{y \in Y} P^*(y|x) \log P^*(y|x)$ - энтропия голосующей вероятности;
- $x^* = \operatorname{argmax}_{x \in U} \varphi_\theta(x)$ - следующая точка.

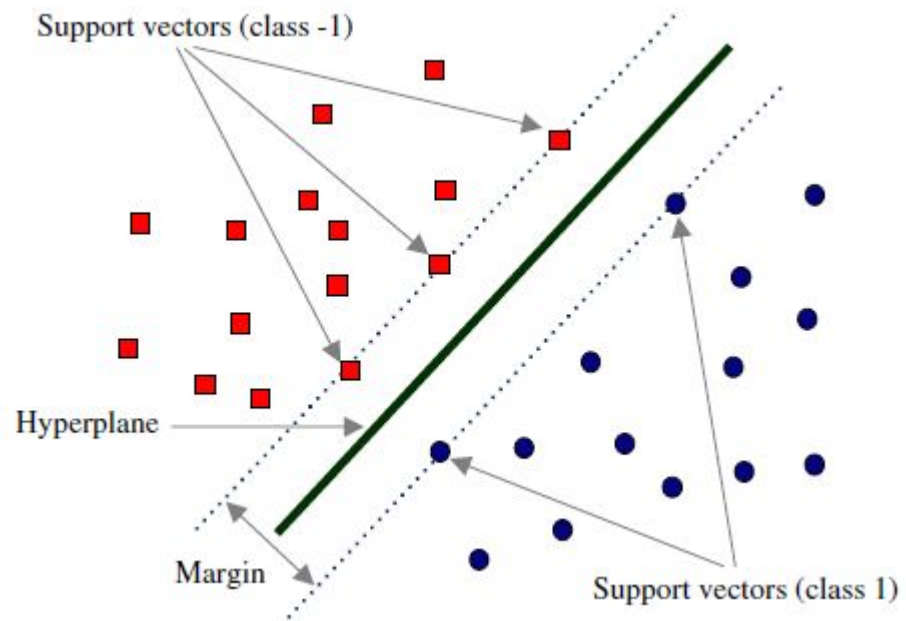
Ожидаемое влияние на модель

- U - множество размеченных x ;
- $\nabla l_{\theta}(L)$ - градиент функционала обучения;
- $\varphi_{\theta}(x) = \sum_{y \in Y} P_{\theta}(y|x) \|\nabla l_{\theta}(L \cup (x, y))\|$;
- $x^* = \operatorname{argmax}_{x \in U} \varphi_{\theta}(x)$ - следующая точка.

Ожидаемое уменьшение ошибки

- $\theta^+(x, y)$ - оптимальный вектор параметров после дообучения модели на $L \cup (x, y)$;
- $y^* = y^*(z)$ - наиболее вероятный класс для z в модели, обученной на L ;
- $\varphi_\theta(x) = - \sum_{y \in Y} P_\theta(y|x) \sum_{z \in U} (1 - P_{\theta^+(x, y)}(y^*|z))$ - expected 0-1 loss (максимизация уверенности на остальных объектах);
- $x^* = \operatorname{argmax}_{x \in U} \varphi_\theta(x)$ - следующая точка.

SVM



Активное обучение в SVM

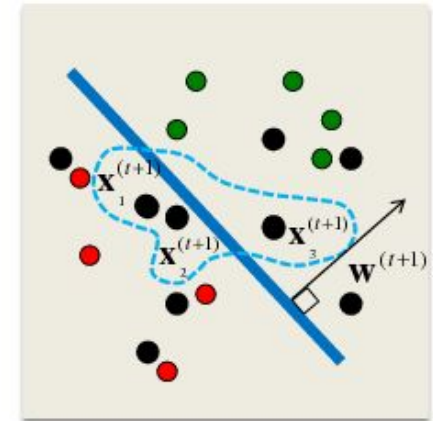
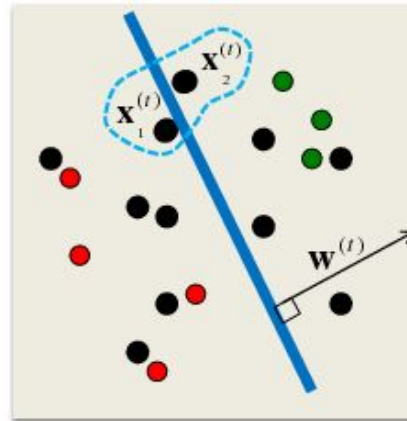
Tong & Koller (2001):

- Пространство возможных решений - множество разделяющих гиперплоскостей.
- Теоретически доказано, что оптимальнее выбирать следующую точку для запроса так, чтобы она разделяла пространство возможных моделей на два подпространства с примерно одинаковым зазором.

Активное обучение в SVM

Методы выбора точки для запроса:

- **Simple Margin:** точка, ближайшая к разделяющей гиперплоскости;



Активное обучение в SVM

Методы выбора точки для запроса:

- **MaxMin Margin:** $\operatorname{argmax}_x \min\{m_+(x), m_-(x)\};$
- **Ratio Margin:** $\operatorname{argmax}_x \min\{m_+(x)/m_-(x), m_-(x)/m_+(x)\};$

где m_+ , m_- - расстояние от x до опорной гиперплоскости после назначения x к классу $+$ и $-$ соответственно.