АЛГОРИТМЫ И МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Тема 1.

Введение.

Постановка задач обучения.

Основные понятия.

Понятие машинного обучения

Машинное обучение (machine learning) — подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться.

Машинное обучение находится на стыке математической статистики, методов оптимизации и классических математических дисциплин.

Виды машинного обучения

Обучение по прецедентам (индуктивное обучение) основано на выявлении общих закономерностей по частным эмпирическим данным.

Дедуктивное обучение предполагает формализацию знаний экспертов и их перенос в компьютер в виде базы знаний.

Материалы по курсу

- **К. В. Воронцов** Машинное обучение – курс лекций
- **С. Осовский** Нейронные сети для обработки информации
- **С. Короткий** Нейронные сети: алгоритм обратного распространения ошибки

Microsoft Visual Studio 2015 Community Edition
https://www.visualstudio.com/ru-ru/products/visual-studio-community-vs.aspx

Задача обучения

X — множество *объектов*;
Y — множество *ответов*;
$y\colon X \to Y$ — неизвестная зависимость (target function).

Дано:
$\{x_1, \dots, x_\ell\} \subset X$ — *обучающая выборка* (training sample);
$y_i = y(x_i)$, $i = 1, \dots, \ell$ — известные ответы.

Найти:
$a\colon X \to Y$ — алгоритм, решающую функцию (decision function), приближающую y на всём множестве X.

Весь курс машинного обучения — это конкретизация:

- как задаются объекты и какими могут быть ответы;
- в каком смысле «a приближает y»;
- как строить функцию a.

Типы задач обучения

Задачи классификации (classification):

- $Y = \{-1, +1\}$ — *классификация* на 2 класса.
- $Y = \{1, \dots, M\}$ — на M непересекающихся классов.
- $Y = \{0, 1\}^M$ — на M классов, которые могут пересекаться.

Задачи восстановления регрессии (regression):

- $Y = \mathbb{R}$ или $Y = \mathbb{R}^m$.

Примеры задач обучения

Задачи классификации

- Задача медицинской диагностики
- Задача оценивания заемщиков
- Задача предсказания ухода клиентов

Задачи восстановления регрессии

- Задача прогнозирования потребительского спроса
- Задача предсказания рейтингов
- Задача аппроксимации функции

Модель алгоритмов

Модель (predictive model) — параметрическое семейство функций

$$A = \{g(x,\theta) \mid \theta \in \Theta\},$$

где $g\colon X \times \Theta \to Y$ — фиксированная функция,
Θ — множество допустимых значений параметра θ.

Пример.
Линейная модель с вектором параметров $\theta = (\theta_1, \dots, \theta_n)$, $\Theta = \mathbb{R}^n$:

$$g(x,\theta) = \sum_{j=1}^{n} \theta_j f_j(x)$$ — для регрессии и ранжирования, $Y = \mathbb{R}$;

$$g(x,\theta) = \operatorname{sign} \sum_{j=1}^{n} \theta_j f_j(x)$$ — для классификации, $Y = \{-1, +1\}$.

Метод обучения

Метод обучения (learning algorithm) — это отображение вида

$$\mu\colon (X \times Y)^\ell \to A,$$

которое произвольной выборке $X^\ell = (x_i, y_i)_{i=1}^\ell$ ставит в соответствие некоторый алгоритм $a \in A$.

В задачах обучения по прецедентам всегда есть два этапа:

- Этап *обучения* (training):
  метод μ по выборке X^ℓ строит алгоритм $a = \mu(X^\ell)$.
- Этап *применения* (testing):
  алгоритм a для новых объектов x выдаёт ответы $a(x)$.

Функционал качества обучения

$\mathscr{L}(a, x)$ — *функция потерь* (loss function) — величина ошибки алгоритма $a \in A$ на объекте $x \in X$.

Функции потерь для задач классификации:

- $\mathscr{L}(a, x) = [a(x) \neq y(x)]$ — индикатор ошибки;

Функции потерь для задач регрессии:

- $\mathscr{L}(a, x) = |a(x) - y(x)|$ — абсолютное значение ошибки;
- $\mathscr{L}(a, x) = (a(x) - y(x))^2$ — квадратичная ошибка.

Эмпирический риск — функционал качества алгоритма a на X^ℓ:

$$Q(a, X^\ell) = \frac{1}{\ell} \sum_{i=1}^{\ell} \mathscr{L}(a, x_i).$$

Сведение задачи обучения к задаче оптимизации

Метод минимизации эмпирического риска:

$$\mu(X^\ell) = \arg\min_{a\in A} Q(a, X^\ell).$$

Пример: *метод наименьших квадратов* ($Y = \mathbb{R}$, $\mathscr{L}$ квадратична):

$$\mu(X^\ell) = \arg\min_{\theta} \sum_{i=1}^{\ell} (g(x_i, \theta) - y_i)^2.$$

Проблема обобщающей способности:

- найдём ли мы «закон природы» или *переобучимся*, то есть подгоним функцию $g(x_i, \theta)$ под заданные точки?
- будет ли $a = \mu(X^\ell)$ приближать функцию y на всём X?
- будет ли $Q(a, X^k)$ мало́ на новых данных — *контрольной выборке* $X^k = (x_i', y_i')_{i=1}^k$, $y_i' = y(x_i)$?

Пример задачи оптимизации

Зависимость $y(x) = \dfrac{1}{1+25x^2}$ на отрезке $x \in [-2, 2]$.

Признаковое описание $x \mapsto (1, x^1, x^2, \dots, x^n)$.

Модель полиномиальной регрессии

$$a(x,\theta) = \theta_0 + \theta_1 x + \dots + \theta_n x^n \text{ — полином степени } n.$$

Обучение методом наименьших квадратов:

$$Q(\theta, X^\ell) = \sum_{i=1}^{\ell} (\theta_0 + \theta_1 x_i + \dots + \theta_n x_i^n - y_i)^2 \to \min_{\theta_0,\dots,\theta_n}.$$

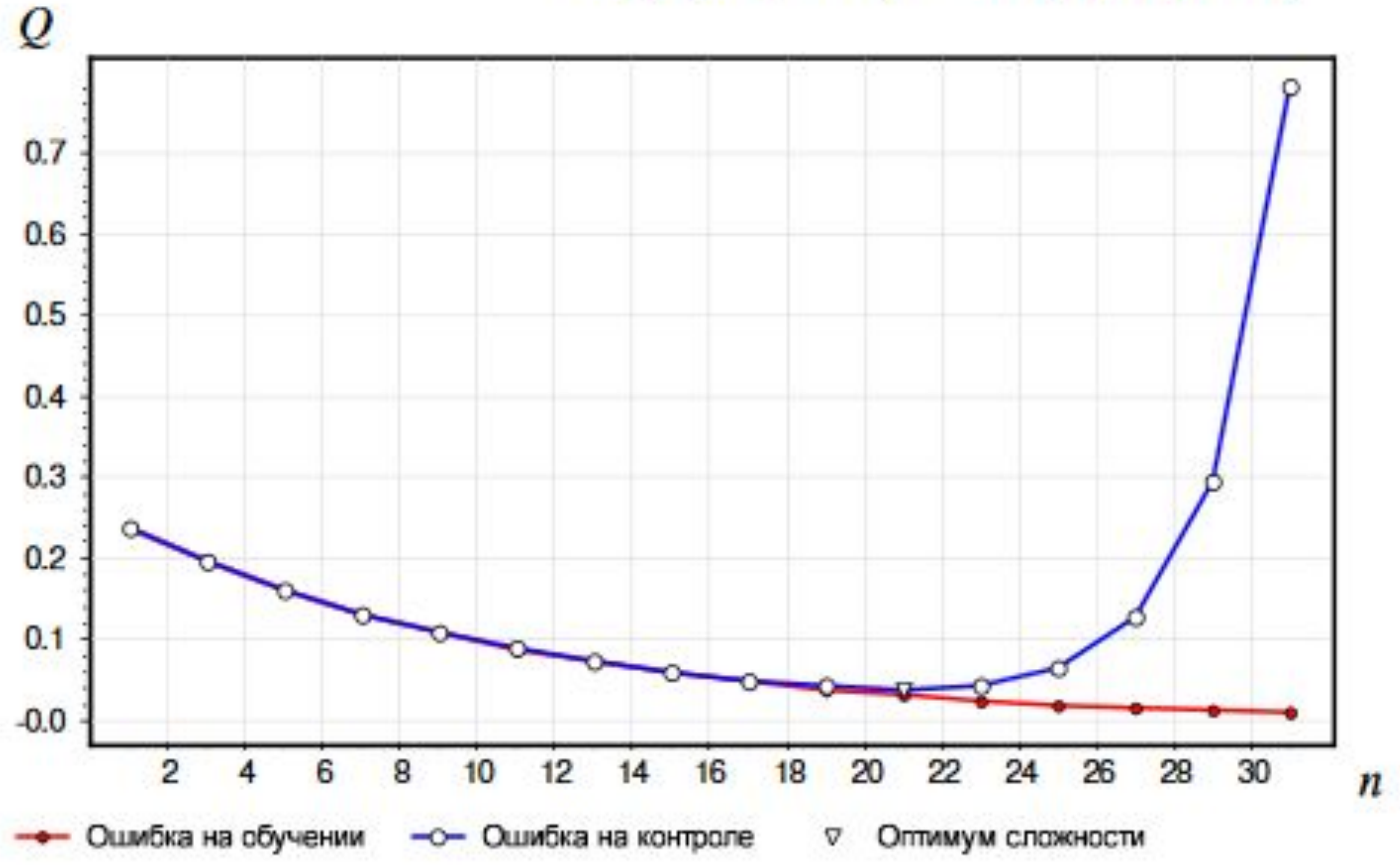
Обучающая выборка: $X^\ell = \{x_i = 4\frac{i-1}{\ell-1} - 2 \mid i = 1, \dots, \ell\}$.

Контрольная выборка: $X^k = \{x_i = 4\frac{i-0.5}{\ell-1} - 2 \mid i = 1, \dots, \ell-1\}$.

Что происходит с $Q(\theta, X^\ell)$ и $Q(\theta, X^k)$ при увеличении n?

Пример переобучения

Переобучение — это когда $Q(\mu(X^\ell), X^k) \gg Q(\mu(X^\ell), X^\ell)$:



Этапы решения задач машинного обучения

- Понимание задачи и данных
- Построение модели
- Сведение обучения к оптимизации
- Решение проблем оптимизации и переобучения
- Оценка качества
- Внедрение и эксплуатация