

# Оглавление

<b>1</b>	<b>Введение</b>	<b>2</b>
1.1	Машинное обучение . . . . .	2
1.2	Виды задач . . . . .	3
1.3	Выбор модели, переобучение . . . . .	3
<b>2</b>	<b>Классическое обучение с учителем</b>	<b>4</b>
2.1	Линейные модели . . . . .	4

# Глава 1

## Введение

### 1.1 Машинное обучение

**Машинное обучение** – это наука, изучающая алгоритмы, автоматически улучшающиеся благодаря опыту.

Большинство решений задач можно представить в виде функции:

$$\text{Примеры (samples)} \rightarrow \text{Предсказания (targets)}.$$

Данная функция – **модель**, а набор примеров – **обучающая выборка (dataset)**.

Обучающая выборка = Объекты + Ответы.

Качество таких предсказаний измеряют **метриками** – функциями, которые показывают насколько полученные предсказания похожи на правильные ответы. Примером метрики является **среднее абсолютное отклонение**:

$$MAE(f, X, y) = L(f, X, y) = \frac{1}{N} \sum_{i=1}^N |f(x_i) - y_i|.$$

На сегодня достаточно знать два типа моделей – **градиентный бустинг на решающих деревьях** и **нейросетевые модели**.

Метрику, которую используют при поиске оптимальной модели – **функция потерь, лосс-функцией (loss)**.

## 1.2 Виды задач

Определенные выше задачи – **обучение с учителем (supervised learning)**, так как правильные ответы были даны заранее. Виды таких обучений:

- $\mathbb{Y} = \mathbb{R}$  или  $\mathbb{Y} = \mathbb{R}^M$  – **регрессия**;
- $\mathbb{Y} = \{0, 1\}$  – **бинарная классификация**;
- $\mathbb{Y} = \{1, \dots, K\}$  – **многоклассовая (multiclass) классификация**;
- $\mathbb{Y} = \{0, 1\}^K$  – **многоклассовая классификация с пересекающимися классами (multilabel classification)**;
- $\mathbb{Y}$  – конечное упорядоченное множество – **ранжирование**.

Имеется другой класс задач – **обучение без учителя (unsupervised learning)** – для которой известны только данные, а ответы отсутствуют. Одним из примеров является *кластеризация* – задача разделения объектов на группы, обладающие некоторыми свойствами.

## 1.3 Выбор модели, переобучение

**Обобщающаяся способность** модели – способность модели учить общие закономерности и давать адекватные предсказания. Выборку для этого делят на две части: **обучающая выборка** и **тестовая выборка (train и test)**.

Такой подход позволяет отделить модели, которые просто удачно подстроились к обучающим данным, от моделей, в которых произошла **генерализация (generalization)**, то есть от таких, которые на самом деле кое-что поняли о том, как устроены данные, и могут выдавать полезные предсказания для объектов, которые не видели.

**Переобученный алгоритм** – алгоритм, избыточно подстроившийся под данные.

## Глава 2

# Классическое обучение с учителем

### 2.1 Линейные модели