

JeML 

Машинное обучение и с чем его едят

Варвус Артем

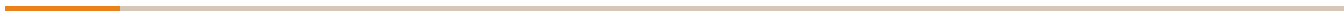
2024-04-22

Финансовый Университет при Правительстве РФ

Table of contents

1. О курсе
2. Что такое Машинное обучение
3. Какие задачи решаем
4. Регрессия
5. Классификация
6. Ранжирование

О курсе



В рамках данного курса мы рассмотрим основы машинного обучения для **уверенного** старта в данной сфере

В рамках данного курса мы рассмотрим основы машинного обучения для **уверенного** старта в данной сфере

Для прохождения курса вам понадобятся навыки:

- Python
- Numpy
- Pandas
- *Математика*

Каждое занятие будет проходить в формате: 1 час теории, 10 минут ~~чиллим~~ отдыхаем, 50 минут практики

Авторы курса

Что такое Машинное обучение

Что такое Машинное обучение

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

Что такое Машинное обучение

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

Машинное обучение это когда...:

- перевести текст с одного языка на другой
- отличить бобика от мурзика
- оценить, по какой цене удастся продать квартиру

В 1956 году американский математик **Артур Сэмюэл** предложил первую формальную концепцию машинного обучения и создал программу, способную играть в шашки на уровне человека. Это стало одним из первых примеров применения алгоритмов обучения для достижения высокой производительности в задачах, требующих интеллектуальных усилий.

Что такое Машинное обучение

Модель машинного обучения — это

Модель машинного обучения — это объект, который был **обучен** для распознавания определенных типов и дальнейшего **предсказания**. Вы обучаете модель на основе набора данных, предоставляя ей алгоритм, который она может использовать для анализа и обучения на основе этих данных.

Тогда пусть задана выборка: (\mathbb{X}, \mathbb{Y}) , где \mathbb{X} - множество признаков, \mathbb{Y} - “ответы” к ним

Тогда пусть задана выборка: (\mathbb{X}, \mathbb{Y}) , где \mathbb{X} - множество признаков, \mathbb{Y} - “ответы” к ним
Пусть задана: $g : \mathbb{X} \times \Theta \rightarrow \mathbb{Y}$, где Θ - веса модели, которые подбираются при обучении

Тогда пусть задана выборка: (\mathbb{X}, \mathbb{Y}) , где \mathbb{X} - множество признаков, \mathbb{Y} - “ответы” к ним
Пусть задана: $g : \mathbb{X} \times \Theta \rightarrow \mathbb{Y}$, где Θ - веса модели, которые подбираются при обучении
То модель: $A = \{g(x, \theta) \mid \theta \in \Theta\}$

Что значит обучается?

У модели есть:

- параметры
-

У модели есть:

- параметры
- гиперпараметры

Параметры модели — параметры, которые изменяются и оптимизируются в процессе обучения модели и итоговые значения этих параметров являются результатом обучения модели.

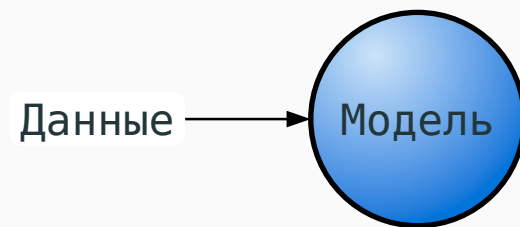
Гиперпараметры модели — параметры, значения которых задается до начала обучения модели и не изменяется в процессе обучения. У модели может не быть гиперпараметров.

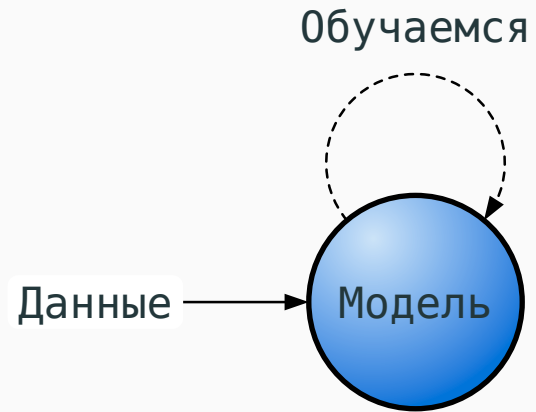
Что из этого является весами модели? 🤔

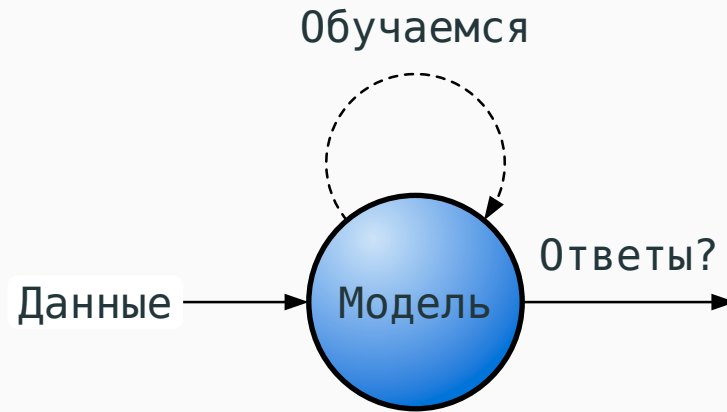
Что из этого является весами модели? 🤔

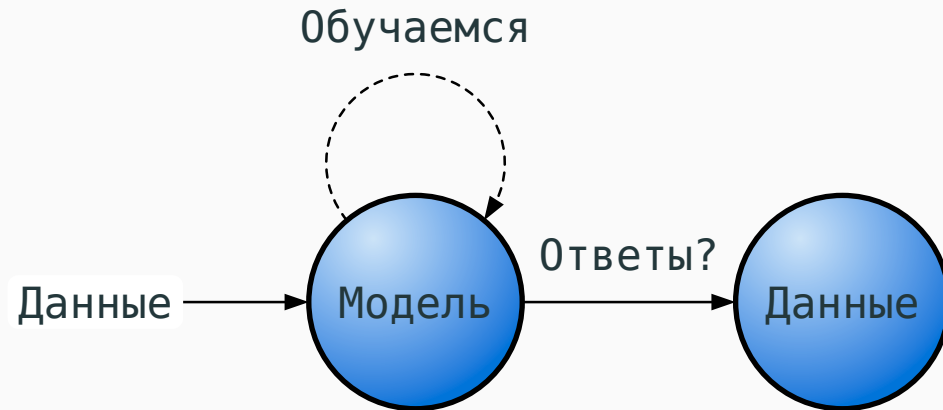
Ответ: Параметры

Данные →









Какие задачи решаем

Какие задачи решаем

Рассмотрим задачи **обучения с учителем**

Рассмотрим задачи **обучения с учителем** , то есть те, в которых у нас есть \mathbb{Y}

Рассмотрим задачи **обучения с учителем** , то есть те, в которых у нас есть \mathbb{Y}

- Регрессия
-
-

Рассмотрим задачи **обучения с учителем** , то есть те, в которых у нас есть \mathbb{Y}

- Регрессия
- Классификация
-

Рассмотрим задачи **обучения с учителем** , то есть те, в которых у нас есть \mathbb{Y}

- Регрессия
- Классификация
- Ранжирование

Регрессия

Целевая переменная представлена: $Y = \mathbb{R}$ или $Y = \mathbb{R}^M$

Целевая переменная представлена: $Y = \mathbb{R}$ или $Y = \mathbb{R}^M$

- продолжительности поездки на каршеринге
-
-

Целевая переменная представлена: $Y = \mathbb{R}$ или $Y = \mathbb{R}^M$

- продолжительности поездки на каршеринге
- спрос на конкретный товар
-

Целевая переменная представлена: $Y = \mathbb{R}$ или $Y = \mathbb{R}^M$

- продолжительности поездки на каршеринге
- спрос на конкретный товар
- погода на Рязанском проспекте завтра

Классификация

Можно разбить на несколько подклассов:

- Бинарная
-

Можно разбить на несколько подклассов:

- Бинарная
- Многоклассовая

Таргет представлен: $\mathbb{Y} = \{0, 1\}$

- сдаст ли студент сессию
-
-

Таргет представлен: $\mathbb{Y} = \{0, 1\}$

- сдаст ли студент сессию
- упадет метеорит
-

Таргет представлен: $\mathbb{Y} = \{0, 1\}$

- сдаст ли студент сессию
- упадет метеорит
- перейдет ли клиент в другой банк

Ранжирование



Y-ки можно представить: $\mathbb{Y} = \{1, \dots, M\}$

- сегодня выбираете на какую пару пойти
-
-

Y-ки можно представить: $\mathbb{Y} = \{1, \dots, M\}$

- сегодня выбираете на какую пару пойти
- предсказать язык в переводчике
-

Y-ки можно представить: $\mathbb{Y} = \{1, \dots, M\}$

- сегодня выбираете на какую пару пойти
- предсказать язык в переводчике
- какая порода представлена перед нами

Как оценить модель?

Как оценить модель?

По обучающей выборке мы хотим построить модель, предсказания которой достаточно хороши. Что вообще значит «достаточно хороши»? Не понимая, чего мы хотим добиться, мы не предложим хорошего решения, поэтому нужно внимательно относиться к выбору метрик качества.

Метрика – внешний, объективный критерий качества, обычно зависящий не от параметров модели, а только от предсказанных меток.

MAE(mean absolute error):

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

MAE(mean absolute error):

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

LogLoss:

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i))$$

Отдых!

Переходим к практике

Что сегодня узнали?

Что сегодня узнали?

- Какие задачи решает **Машинное обучение**
- Основные термины
- Поигрались с **scikit-learn**

Спасибо за внимание!