Математика для Data Science. Математический анализ. Шпаргалка

Содержание

Певрая неделя. Введение, множества и доказательства

Объекты и целевая функция

Целевая функция — то, что мы хотим научиться вычислять для объектов из некоторого множества.

Признаки (или *фичи*) — набор того, что описывает объекты.

Регрессия — тип задач, где значение целевой функции может быть произвольным числом из некоторого промежутка.

Классификация — тип задач, в которых нужно отнести объект к одному из классов. *Бинарная классификация* — частный случай классификации, в которой возможных классов всего два.

Объект в машинном обучении часто представляют в виде *вектора* или *матрицы*. Неформально говоря, *вектор* — это набор значений признаков, а *матрица* — это табличка, в которой стоят значения признаков.

Функция потерь и данные в машинном обучении

Функция потерь показывает, насколько предсказанный ответ далёк от реального.

Пусть a — предсказанный ответ, y — реальный ответ. Приведём **примеры функций потерь** $L(y, a)$ в рассмотренных нами типах задач.

1. Задача регрессии.
   - Модуль отклонения: $L(y, a) = |y - a|$.
   - Квадрат отклонения: $L(y, a) = (y - a)^2$.
2. Задача бинарной классификации для классов 0 и 1.
   - *Индикаторная функция потерь*: $L(0, 0) = L(1, 1) = 1$ и для всех остальных аргументов $L(y, a) = 0$. Обозначается $L(y, a)$ как $\mathbf{1}\{y = a\}$.
   - Функция потерь, предсказывающая не класс объекта, а *вероятность* принадлежности объекта к одному из классов.

Обучающая выборка — набор размеченных данных, то есть набор объектов, для которых известно значение целевой функции.

Пусть объекты пронумерованы числами от 1 до n, и для этих объектов значения целевой функции – $y_1, y_2, \ldots, y_n$ соответственно, а предсказание нашего алгоритма – $a_1, a_2, \ldots, a_n$ соответственно. Приведём **примеры функций потерь для нашей выборки:**

1. Задача регрессии.
   - *Mean absolute error (MAE)* или *среднее отклонение по модулю* – это среднее арифметическое модулей отклонений:
     $MAE(y_1, y_2, \ldots, y_n, a_1, a_2, \ldots, a_n) = \frac{1}{n}(|y_1 - a_1| + |y_2 - a_2| + \cdots + |y_n - a_n|)$.
   - *Mean squared error (MSE)* или *среднеквадратичная ошибка* — это среднее арифметическое квадратов отклонений:
     $MSE(y_1, y_2, \ldots, y_n, a_1, a_2, \ldots, a_n) = \frac{1}{n}((y_1 - a_1)^2 + (y_2 - a_2)^2 + \cdots + (y_n - a_n)^2)$.
2. Задача бинарной классификации.
   - *Точность (accuracy)* — доля правильных ответов: $Acc(y_1, y_2, \ldots, y_n, a_1, a_2, \ldots, a_n) = \frac{1}{n}(\mathbf{1}\{y_1 = a_1\} + \mathbf{1}\{y_2 = a_2\} + \cdots + \mathbf{1}\{y_n = a_n\})$.

Модель машинного обучения

Как правило, при решении задачи машинного обучения выбирается некоторый класс алгоритмов, где каждый конкретный алгоритм из класса задаётся *параметрами* или, иначе говоря, *весами*.

Алгоритм с фиксированными весами называется *моделью*.

Примеры классов алгоритмов:

1. Задача бинарной классификации.

   - Класс *константных функций*. Ему принадлежат алгоритмы, которые всегда выдают один и тот же ответ — постоянную величину c.
   - Класс *пороговых функций*. Ему принадлежат классификаторы вида $\mathbf{1}\{s \leq t\}$. Здесь $\mathbf{1}$ — индикаторная функция, которая возвращает 1, если условие внутри фигурных скобок выполнено, и 0 иначе. За s обозначено значение признака, а t — фиксированое число, называемое *порогом* (от слова *threshold*).

2. Задача регрессии.

   - Класс *линейных функций*. Ему принадлежат функции вида $\hat{f}(r, d, p) = w_r r + w_d d + w_p p + w_0$, где r, d, p — значения признаков (в общем случае их n, где n — количество признаков), а w_r, w_d, w_p, w_0 — коэффициенты (в общем случае их $n+1$). w_0 называется *свободным коэффициентом*, его ещё называют *сдвигом* (или *bias*).

Множества

Множество — математический объект, являющийся набором других объектов.

Объекты, из которых состоит множество, называют *элементами множества* или *точками множества*.

Множества обычно обозначают заглавными буквами латинского алфавита, а элементы множества — строчными.

Любой элемент содержится в множестве не больше одного раза. Множества, отличающиеся порядком элементов, считаются одинаковыми.

$x \in A$ читается как «x является элементом множества A» или «x принадлежит A».

$y \notin A$ читается как «y не принадлежит A».

Способы задания множества

- Перечислить его элементы внутри фигурных скобок.
- Задать описанием: «множество всех x, таких что для них выполнено условие P». Записывается это в форме $\{x \mid P(x)\}$.

Операции над множествами

- *Пересечение* множеств A и B — это множество $A \cap B := \{x \mid x \in A \text{ и } x \in B\}$. Здесь знак «:=» читается как «по определению равно».
- *Объединение* множеств A и B — это множество $A \cup B := \{x \mid x \in A \text{ или } x \in B\}$.
- *Разность* множеств A и B — это множество $A \setminus B := \{x \mid x \in A \text{ и } x \notin B\}$.

Множество A называется *подмножеством* множества B, если все элементы множества A также являются элементами множества B. Обозначение: $A \subset B$.

Пустое множество — это множество, в котором нет элементов.

Способы изобразить множества

1. На *диаграмме Эйлера* множества рисуются как круги, а внутри кругов располагаются элементы.
2. В общем случае, если про множества ничего не известно, рисуют диаграмму *Эйлера-Венна* — диаграмму Эйлера со всеми возможными пересечениями.

Некоторые часто встречающиеся множества

- $\mathbb{N} = \{1, 2, 3, \dots\}$ — множество *натуральных* чисел, то есть чисел, возникающих при счёте.
- $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$ — множество *целых* чисел.
- $\mathbb{Q}$ — множество *рациональных* чисел, то есть чисел, которые можно записать в виде дроби $\frac{m}{n}$, где $m \in \mathbb{Z}$ и $n \in \mathbb{N}$. Числа $x \notin \mathbb{Q}$ называются *иррациональными*.

- $\mathbb{R}$ — множество *действительных* (или *вещественных*) чисел. Действительное число — это бесконечная десятичная дробь, то есть выражение вида $\pm a_0.a_1a_2a_3\ldots$, где $\pm$ — это знак $+$ или знак $-$, a_0 — целое неотрицательное число, и $a_i \in \{0, 1, 2, \ldots, 9\}$ для всех $i \geq 1$.
- $\mathbb{R}^n$ — множество всех наборов из n действительных чисел.

Некоторые часто встречающиеся подмножества $\mathbb{R}$

- При $a < b$ *отрезком* называется множество $[a, b] := \{x \mid a \leq x \leq b\}$. Точки a и b называются *граничными* точками отрезка.
- При $a < b$ *интервалом* называется множество $(a, b) := \{x \mid a < x < b\}$.
- *Замкнутыми лучами* называются множества $[a, +\infty) := \{x \mid a \leq x\}$ и $(-\infty, a] := \{x \mid x \leq a\}$. Точка a называется *граничной* точкой замкнутого луча.
- *Открытыми лучами* называются множества $(a, +\infty) := \{x \mid a < x\}$ и $(-\infty, a) := \{x \mid x < a\}$.
- Интервал $(x_0 - \varepsilon, x_0 + \varepsilon)$ называется *ε-окрестностью* точки x_0, где ε (читается как «эпсилон») — положительное действительное число.
- *Проколотой ε-окрестностью* точки x_0 называется ε-окрестность точки x_0, в которую не входит сама точка x_0.

How-to по доказательствам

Общепринятые сокращения:

- $\implies$ — *следствие*. $A \implies B$ означает следующее: если выполнено утверждение A, то выполнено утверждение B.
- $\iff$ — *равносильность* (читается «тогда и только тогда»). Выражение $A \iff B$ означает, что $A \implies B$ и $B \implies A$. То есть: если выполнено утверждение A, то выполнено утверждение B, и наоборот — если верно утверждение B, то верно утверждение A.
- *Квантор* всегда идёт вместе с переменной или набором переменных, после чего идёт утверждение, в котором этот x фигурирует. Есть два вида кванторов:
  1. $\exists$ — *квантор существования.* $\exists x$: $A(x)$ означает, что существует значение x, при подстановке которого утверждение $A(x)$ становится истинным.
  2. $\forall$ — *квантор всеобщности.* $\forall x$: $A(x)$ означает, что для любого значения x утверждение $A(x)$ истинно.
- $\neg$ — отрицаниие. Отрицание к утверждению A записывается как $\neg A$.

Правила построения отрицаний

1. $\neg(A \text{ или } B) = \neg A\ \text{ и } \neg B$. То есть отрицанием к утверждению вида «A или B» будет утверждение: «(A неверно) и (B неверно)».
2. $\neg(A \text{ и } B) = \neg A\ \text{ или } \neg B$. То есть отрицанием к утверждению вида «A и B» будет утверждение: «(A неверно) или (B неверно)».
3. $\neg(\forall x:\ A(x)) \iff \exists x:\ \neg A(x)$. То есть отрицанием к утверждению вида «для всех x верно $A(x)$» будет утверждение вида «существует x такой, что неверно $A(x)$».
4. $\neg(\exists x : A(x)) \iff \forall x:\ \neg A(x)$. То есть отрицанием к утверждению вида «существует x такой, что верно $A(x)$» будет утверждение вида «для всех x неверно $A(x)$».

Закон контрапозиции гласит, что для утверждений X и Y выполнено

$$(X \implies Y) \iff (\neg Y \implies \neg X).$$

То есть утвеждения «если X, то Y» и «если неверно Y, то X тоже неверно» эквивалентны.

Закон контрапозиции используется при *доказательстве от противного*: мы предполагаем, что доказываемое утверждение неверно, после чего выводим противоречие.

Функции

Функция — это соответствие между элементами двух множеств, такое что каждому элементу первого множества соответствует ровно один элемент второго множества. Пусть первое множество обозначено через X, второе через Y, а функция через f. Тогда мы будем говорить, что «функция f отображает X в Y», «функция f из X в Y» или $f : X \to Y$.

Элемент $x \in X$, к которому мы применяем функцию f, называется *аргументом* функции, а элемент $f(x) \in Y$ называется *значением* функции.

Если функция f отображает X в Y, то X называется *областью определения* функции f. Множество всех значений, которые принимает функция f, называется *областью значений* функции f.

Точкой минимума функции $f : X \to \mathbb{R}$ называется такой $x_{min} \in X$, что $f(x_{min}) \le f(x)$ для всех $x \in X$.