

Тема 4. Регрессия

Задание 1. Написать (добавить в собственный класс/библиотеку) следующие функции:

1. Функцию, реализующую модель линейной регрессии и 3 видами регуляторов: L1, L2 и Стьюдента.
** Возможна самостоятельная декомпозиция задачи на несколько функций, решающих конкретные подзадачи. В таком случае, для удобства вызова пользователем, необходимо предусмотреть оберточную функцию.*

Формат входных данных:

Обязательные параметры:

- а) массив предикторов (переменные X);
- б) массив предсказываемой переменной (переменная y);

Необязательные параметры:

- в) вид регуляризации (по умолчанию None, регрессия выполняется без нее);
- г) построение графика (по умолчанию False), определяет, будет ли построен график.

* возможно добавление других обязательных или необязательных параметров

Формат выходных данных:

- а) получившееся функция в аналитическом виде;
- б) массив коэффициентов регрессии (веса каждого x);
- в) значение свободного члена регрессии;
- г) график регрессии, если стоит соответствующий параметр.

2. Добавить функцию, реализующую модель полиномиальной регрессии и 3 видами регуляторов: L1, L2 и Стьюдента. **(опционально)**
** Возможна самостоятельная декомпозиция задачи на несколько функций, решающих конкретные подзадачи. В таком случае, для удобства вызова пользователем, необходимо предусмотреть оберточную функцию.*

Формат входных данных:

Обязательные параметры:

- а) массив предикторов (переменные X);

б) массив предсказываемой переменной (переменная y).

Необязательные параметры:

в) вид регуляризации (по умолчанию None, регрессия выполняется без нее);

г) построение графика (по умолчанию False), определяет, будет ли построен график.

* возможно добавление других обязательных или необязательных параметров

Формат выходных данных:

а) получившееся функция в аналитическом виде;

б) массив коэффициентов регрессии (веса каждого x);

в) значение свободного члена регрессии;

г) график регрессии, если стоит соответствующий параметр.

3. Добавить функцию, реализующую модель экспоненциальной регрессии и 3 видами регуляторов: L1, L2 и Стьюдента. **(опционально)**

** Возможна самостоятельная декомпозиция задачи на несколько функций, решающих конкретные подзадачи. В таком случае, для удобства вызова пользователем, необходимо предусмотреть оберточную функцию.*

Формат входных данных:

Обязательные параметры:

а) массив предикторов (переменные X);

б) массив предсказываемой переменной (переменная y).

Необязательные параметры:

в) вид регуляризации (по умолчанию None, регрессия выполняется без нее);

г) построение графика (по умолчанию False), определяет, будет ли построен график.

* возможно добавление других обязательных или необязательных параметров

Формат выходных данных:

а) получившееся функция в аналитическом виде;

б) массив коэффициентов регрессии (веса каждого x);

- в) значение свободного члена регрессии;
- г) график регрессии, если стоит соответствующий параметр.

Задание 2. Для тестирования написанных функций а так же прототипирования различных методов проводится следующее приемо-сдаточное тестирование:

** Данные задания оформляются в отдельном пайплайне (в случае подключения собственной библиотеки), или ячейках, располагающихся ниже. Каждый тест пишется в своей ячейке и решает свой класс задач.*

1. В функцию передаются два массива данных, массив регрессоров может быть многомерным. Результатом работы программы является функция в аналитическом виде, отдельно значения коэффициентов регрессии для более удобной обработки и построенный график, если указан соответствующий параметр.

2. Визуализировать работу написанной функции.

Предлагается построить полученные регрессии с различными регуляризациями, показать отличия на одном наборе данных.

** Возможны другие варианты визуализации работы алгоритмов, на усмотрение студентов.*

Дополнительные задания (опционально)

4. Сравнить точность построенных регрессий на реальных данных

Предлагается для одного и того же набора данных построить все виды регрессий, указанных выше с различными регуляризациями, оценить точность предсказания.

5. Визуализировать работу написанных функций.

Предлагается построить полученные регрессии, построенные в п. 1 и 2 дополнительных заданий с различными регуляризациями (одна визуализация на каждый вид регрессии) и показать отличия на одном наборе данных.

** Возможны другие варианты визуализации работы алгоритмов, на усмотрение студентов.*

6. Необходимо оформить проектную документацию по проекту в формате файла «Описание структуры и разделов документации по проекту.docx»

Список литературы для подготовки:

- 1) <https://habr.com/ru/company/ods/blog/323890/>
- 2) <https://neerc.ifmo.ru/wiki/index.php?title=%D0%A0%D0%B5%D0%B3%D1%83%D0%BB%D1%8F%D1%80%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F>