



Магистратура СамГТУ,
направление 02.04.03, профиль «Анализ данных»

Регрессоры

Контрольная работа.

ВАРИАНТ 1.

1. Что такое целевая функция регрессии? Как она выглядит в случае полиномиальной модели? Какие есть параметры полиномиальной регрессии, которые могут быть отрегулированы в процессе обучения модели?
2. Провести анализ данных из файла `data1.1.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. При помощи методов `figure` и `plot` библиотеки `matplotlib` визуализировать данные. Использовать ограничения по осям `xlim` и `ylim`, включить сетку методом `grid`. Визуализировать проекции данных по всем осям.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Реализовать 5 моделей, характерными отличительными особенностями которых являются следующие требования:

- (a) модель должна быть обучена с опцией `fit_intercept=False`;
 - (b) модель должна быть обучена на нормированных данных;
 - (c) модель должна быть оптимизирована с l_1 -регуляризацией;
 - (d) модель должна быть обучена с параметрами, предварительно отобранными на кросс-валидации;
 - (e) в модели в качестве оптимизатора нельзя брать метод наименьших квадратов, поэтому следует выбрать любой другой.
3. Провести анализа данных из файла `data1.2.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. Необходимо обучить минимум одну модель, найдя правильную функцию, описывающую и прогнозирующую данные.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Ответить на вопросы по результатам обучения:

- (a) Каким типом функции описываются данные?
- (b) Имеются ли в данных незначимые признаки?
- (c) Наблюдается ли в модели переобучение?

ВАРИАНТ 2.

1. Какую формулу может иметь экспоненциальная регрессия? Какими способами она может быть реализована в коде на языке python?
2. Провести анализ данных из файла data2.1.csv и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. При помощи методов `figure` и `plot` библиотеки `matplotlib` визуализировать данные. Использовать ограничения по осям `xlim` и `ylim`, включить сетку методом `grid`. Визуализировать проекции данных по всем осям.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Реализовать 5 моделей, характерными отличительными особенностями которых являются следующие требования:

- (a) модель должна быть обучена с опцией `fit_intercept=False`;
 - (b) модель должна быть обучена на нормированных данных;
 - (c) модель должна быть оптимизирована с l_2 -регуляризацией;
 - (d) модель должна быть обучена с параметрами, предварительно отобранными с помощью опции `GridSearchCV`;
 - (e) в модели в качестве оптимизатора нельзя брать метод наименьших квадратов, поэтому следует выбрать любой другой.
3. Провести анализа данных из файла data2.2.csv и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. Необходимо обучить минимум одну модель, найдя правильную функцию, описывающую и прогнозирующую данные.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Ответить на вопросы по результатам обучения:

- (a) Каким типом функции описываются данные?
- (b) Имеются ли в данных незначимые признаки?
- (c) Наблюдается ли в модели переобучение?

ВАРИАНТ 3.

1. Основные метрики, определяющие качество модели, решающей задачу регрессии. Написать формулы и пояснить все величины.
2. Провести анализ данных из файла `data3.1.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. При помощи методов `figure` и `plot` библиотеки `matplotlib` визуализировать данные. Использовать ограничения по осям `xlim` и `ylim`, включить сетку методом `grid`. Визуализировать проекции данных по всем осям.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Реализовать 5 моделей, характерными отличительными особенностями которых являются следующие требования:

- (a) модель должна быть обучена с опцией `fit_intercept=False`;
 - (b) модель должна быть обучена на нормированных данных;
 - (c) модель должна быть оптимизирована с l_1 -регуляризацией;
 - (d) модель должна быть обучена с параметрами, предварительно отобранными на кросс-валидации;
 - (e) в модели в качестве оптимизатора нельзя брать метод наименьших квадратов, поэтому следует выбрать любой другой.
3. Провести анализа данных из файла `data3.2.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. Необходимо обучить минимум одну модель, найдя правильную функцию, описывающую и прогнозирующую данные.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Ответить на вопросы по результатам обучения:

- (a) Каким типом функции описываются данные?
- (b) Имеются ли в данных незначимые признаки?
- (c) Наблюдается ли в модели переобучение?

ВАРИАНТ 4.

1. Что такое обучающая и тестовая выборка? Как в задаче регрессии используются обучающие и тестовые выборки? Как результаты обучения модели связаны с изначальными обучающими и тестовыми наборами данных и как это можно визуализировать на графике?
2. Провести анализ данных из файла `data4.1.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. При помощи методов `figure` и `plot` библиотеки `matplotlib` визуализировать данные. Использовать ограничения по осям `xlim` и `ylim`, включить сетку методом `grid`. Визуализировать проекции данных по всем осям.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Реализовать 5 моделей, характерными отличительными особенностями которых являются следующие требования:

- (a) модель должна быть обучена с опцией `fit_intercept=False`;
 - (b) модель должна быть обучена на нормированных данных;
 - (c) модель должна быть оптимизирована с l_2 -регуляризацией;
 - (d) модель должна быть обучена с параметрами, предварительно отобранными с помощью опции `GridSearchCV`;
 - (e) в модели в качестве оптимизатора нельзя брать метод наименьших квадратов, поэтому следует выбрать любой другой.
3. Провести анализа данных из файла `data4.2.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. Необходимо обучить минимум одну модель, найдя правильную функцию, описывающую и прогнозирующую данные.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Ответить на вопросы по результатам обучения:

- (a) Каким типом функции описываются данные?
- (b) Имеются ли в данных незначимые признаки?
- (c) Наблюдается ли в модели переобучение?

ВАРИАНТ 5.

1. Регуляризация в линейной модели. Зачем нужна и какими формулами определяется? Как реализуется регуляризация в коде на языке python?
2. Провести анализ данных из файла `data5.1.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. При помощи методов `figure` и `plot` библиотеки `matplotlib` визуализировать данные. Использовать ограничения по осям `xlim` и `ylim`, включить сетку методом `grid`. Визуализировать проекции данных по всем осям.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Реализовать 5 моделей, характерными отличительными особенностями которых являются следующие требования:

- (a) модель должна быть обучена с опцией `fit_intercept=False`;
 - (b) модель должна быть обучена на нормированных данных;
 - (c) модель должна быть оптимизирована с l_1 -регуляризацией;
 - (d) модель должна быть обучена с параметрами, предварительно отобранными на кросс-валидации;
 - (e) в модели в качестве оптимизатора нельзя брать метод наименьших квадратов, поэтому следует выбрать любой другой.
3. Провести анализа данных из файла `data5.2.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. Необходимо обучить минимум одну модель, найдя правильную функцию, описывающую и прогнозирующую данные.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Ответить на вопросы по результатам обучения:

- (a) Каким типом функции описываются данные?
- (b) Имеются ли в данных незначимые признаки?
- (c) Наблюдается ли в модели переобучение?

ВАРИАНТ 6.

1. Какую формулу может иметь регрессия для данных, описывающих колебательные процессы? Какими способами она может быть реализована в коде на языке python?
2. Провести анализ данных из файла `data6.1.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. При помощи методов `figure` и `plot` библиотеки `matplotlib` визуализировать данные. Использовать ограничения по осям `xlim` и `ylim`, включить сетку методом `grid`. Визуализировать проекции данных по всем осям.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Реализовать 5 моделей, характерными отличительными особенностями которых являются следующие требования:

- (a) модель должна быть обучена с опцией `fit_intercept=False`;
 - (b) модель должна быть обучена на нормированных данных;
 - (c) модель должна быть оптимизирована с l_2 -регуляризацией;
 - (d) модель должна быть обучена с параметрами, предварительно отобранными с помощью опции `RandomizedSearchCV`;
 - (e) в модели в качестве оптимизатора нельзя брать метод наименьших квадратов, поэтому следует выбрать любой другой.
3. Провести анализа данных из файла `data6.2.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. Необходимо обучить минимум одну модель, найдя правильную функцию, описывающую и прогнозирующую данные.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Ответить на вопросы по результатам обучения:

- (a) Каким типом функции описываются данные?
- (b) Имеются ли в данных незначимые признаки?
- (c) Наблюдается ли в модели переобучение?

ВАРИАНТ 7.

1. Каким образом производится настройка гиперпараметров модели? Какие есть для этого возможности в библиотеках языка python?
2. Провести анализ данных из файла `data7.1.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. При помощи методов `figure` и `plot` библиотеки `matplotlib` визуализировать данные. Использовать ограничения по осям `xlim` и `ylim`, включить сетку методом `grid`. Визуализировать проекции данных по всем осям.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Реализовать 5 моделей, характерными отличительными особенностями которых являются следующие требования:

- (a) модель должна быть обучена с опцией `fit_intercept=False`;
 - (b) модель должна быть обучена на нормированных данных;
 - (c) модель должна быть оптимизирована с l_1 -регуляризацией;
 - (d) модель должна быть обучена с параметрами, предварительно отобранными на кросс-валидации;
 - (e) в модели в качестве оптимизатора нельзя брать метод наименьших квадратов, поэтому следует выбрать любой другой.
3. Провести анализа данных из файла `data7.2.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. Необходимо обучить минимум одну модель, найдя правильную функцию, описывающую и прогнозирующую данные.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Ответить на вопросы по результатам обучения:

- (a) Каким типом функции описываются данные?
- (b) Имеются ли в данных незначимые признаки?
- (c) Наблюдается ли в модели переобучение?

ВАРИАНТ 8.

1. Какую формулу может иметь логарифмическая регрессия? С помощью каких инструментов она может быть реализована на языке python?
2. Провести анализ данных из файла data8.1.csv и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. При помощи методов `figure` и `plot` библиотеки `matplotlib` визуализировать данные. Использовать ограничения по осям `xlim` и `ylim`, включить сетку методом `grid`. Визуализировать проекции данных по всем осям.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Реализовать 5 моделей, характерными отличительными особенностями которых являются следующие требования:

- (a) модель должна быть обучена с опцией `fit_intercept=False`;
 - (b) модель должна быть обучена на нормированных данных;
 - (c) модель должна быть оптимизирована с l_2 -регуляризацией;
 - (d) модель должна быть обучена с параметрами, предварительно отобранными с помощью опции `RandomizedSearchCV`;
 - (e) в модели в качестве оптимизатора нельзя брать метод наименьших квадратов, поэтому следует выбрать любой другой.
3. Провести анализа данных из файла data8.2.csv и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. Необходимо обучить минимум одну модель, найдя правильную функцию, описывающую и прогнозирующую данные.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Ответить на вопросы по результатам обучения:

- (a) Каким типом функции описываются данные?
- (b) Имеются ли в данных незначимые признаки?
- (c) Наблюдается ли в модели переобучение?

ВАРИАНТ 9.

1. Что такое функция ошибки в задачах регрессии? Какая функция выстраивается при реализации метода наименьшего квадрата? Какие еще могут быть функции? Как параметры функции ошибки связаны в искомой функции в задаче регрессии?
2. Провести анализ данных из файла `data9.1.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. При помощи методов `figure` и `plot` библиотеки `matplotlib` визуализировать данные. Использовать ограничения по осям `xlim` и `ylim`, включить сетку методом `grid`. Визуализировать проекции данных по всем осям.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Реализовать 5 моделей, характерными отличительными особенностями которых являются следующие требования:

- (a) модель должна быть обучена с опцией `fit_intercept=False`;
 - (b) модель должна быть обучена на нормированных данных;
 - (c) модель должна быть оптимизирована с l_1 -регуляризацией;
 - (d) модель должна быть обучена с параметрами, предварительно отобранными на кросс-валидации;
 - (e) в модели в качестве оптимизатора нельзя брать метод наименьших квадратов, поэтому следует выбрать любой другой.
3. Провести анализа данных из файла `data9.2.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. Необходимо обучить минимум одну модель, найдя правильную функцию, описывающую и прогнозирующую данные.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Ответить на вопросы по результатам обучения:

- (a) Каким типом функции описываются данные?
- (b) Имеются ли в данных незначимые признаки?
- (c) Наблюдается ли в модели переобучение?

ВАРИАНТ 10.

1. Перечислить типы регрессии в зависимости от типа функции, которой могут быть описаны данные. Как должны быть обработаны исходные признаки, чтобы они могли быть использованы для построения на данных определенного типа функции? Имеются ли для этого инструменты в библиотеках на языке python?
2. Провести анализ данных из файла data10.1.csv и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. При помощи методов `figure` и `plot` библиотеки `matplotlib` визуализировать данные. Использовать ограничения по осям `xlim` и `ylim`, включить сетку методом `grid`. Визуализировать проекции данных по всем осям.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Реализовать 5 моделей, характерными отличительными особенностями которых являются следующие требования:

- (a) модель должна быть обучена с опцией `fit_intercept=False`;
 - (b) модель должна быть обучена на нормированных данных;
 - (c) модель должна быть оптимизирована с l_2 -регуляризацией;
 - (d) модель должна быть обучена с параметрами, предварительно отобранными с помощью опции `GridSearchCV`;
 - (e) в модели в качестве оптимизатора нельзя брать метод наименьших квадратов, поэтому следует выбрать любой другой.
3. Провести анализа данных из файла data10.2.csv и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. Необходимо обучить минимум одну модель, найдя правильную функцию, описывающую и прогнозирующую данные.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Ответить на вопросы по результатам обучения:

- (a) Каким типом функции описываются данные?
- (b) Имеются ли в данных незначимые признаки?
- (c) Наблюдается ли в модели переобучение?

ВАРИАНТ 11.

1. В чем отличие логистической регрессии? Какой формулой она описывается? Зачем нужна и какими формулами определяется? Как она реализуется реализуется в коде на языке python?
2. Провести анализ данных из файла data11.1.csv и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. При помощи методов `figure` и `plot` библиотеки `matplotlib` визуализировать данные. Использовать ограничения по осям `xlim` и `ylim`, включить сетку методом `grid`. Визуализировать проекции данных по всем осям.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Реализовать 5 моделей, характерными отличительными особенностями которых являются следующие требования:

- (a) модель должна быть обучена с опцией `fit_intercept=False`;
 - (b) модель должна быть обучена на нормированных данных;
 - (c) модель должна быть оптимизирована с l_1 -регуляризацией;
 - (d) модель должна быть обучена с параметрами, предварительно отобранными на кросс-валидации;
 - (e) в модели в качестве оптимизатора нельзя брать метод наименьших квадратов, поэтому следует выбрать любой другой.
3. Провести анализа данных из файла data11.2.csv и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. Необходимо обучить минимум одну модель, найдя правильную функцию, описывающую и прогнозирующую данные.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Ответить на вопросы по результатам обучения:

- (a) Каким типом функции описываются данные?
- (b) Имеются ли в данных незначимые признаки?
- (c) Наблюдается ли в модели переобучение?

ВАРИАНТ 12.

1. Что такое переобучение? Как оно может быть предотвращено?
2. Провести анализ данных из файла `data12.1.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. При помощи методов `figure` и `plot` библиотеки `matplotlib` визуализировать данные. Использовать ограничения по осям `xlim` и `ylim`, включить сетку методом `grid`. Визуализировать проекции данных по всем осям.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Реализовать 5 моделей, характерными отличительными особенностями которых являются следующие требования:

- (a) модель должна быть обучена с опцией `fit_intercept=False`;
 - (b) модель должна быть обучена на нормированных данных;
 - (c) модель должна быть оптимизирована с l_2 -регуляризацией;
 - (d) модель должна быть обучена с параметрами, предварительно отобранными с помощью опции `RandomizedSearchCV`;
 - (e) в модели в качестве оптимизатора нельзя брать метод наименьших квадратов, поэтому следует выбрать любой другой.
3. Провести анализа данных из файла `data12.2.csv` и решить задачу регрессии, выбрав параметр y в качестве целевой переменной. Необходимо обучить минимум одну модель, найдя правильную функцию, описывающую и прогнозирующую данные.

В качестве метрик выбрать коэффициент детерминации R^2 , абсолютное отклонение и среднеквадратичное отклонение. Вывести коэффициенты модели при помощи метода `coef_` и свободный член модели при помощи метода `intercept_`.

Ответить на вопросы по результатам обучения:

- (a) Каким типом функции описываются данные?
- (b) Имеются ли в данных незначимые признаки?
- (c) Наблюдается ли в модели переобучение?