**Федеральное государственное образовательное**

**бюджетное учреждение**

**высшего образования**

**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ**

**ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ**

**ФЕДЕРАЦИИ»**

**(Финансовый университет)**

**Факультет**

**информационных технологий и анализа больших данных**

**Кафедра «Прикладная математика и информатика»**

**Домашнее задание № 4**

«Регрессия»

Студенты группы ПМ19-2:

Коротенко В.Р

Пономаренко А.П

Васильева А.Н

Морозов М.

Жигулина Ю.А

Брашич И

Аракелян Р.

Руководитель:

Аксенов Дмитрий Андреевич

**Москва 2022**

Оглавление.

1. Постановка задачи (физическая модель)
2. Математическая модель
3. Алгоритмы
   1. Алгоритм 1
      1. Описание входных данных
      2. Описание алгоритма решения
      3. Описание выходных данных
   2. Алгоритм 2
      1. Описание входных данных
      2. Описание алгоритма решения
      3. Описание выходных данных
   3. Алгоритм 3
      1. Описание входных данных
      2. Описание алгоритма решения
      3. Описание выходных данных
4. Варианты использования системы
   1. ВИ 1
   2. ВИ 2
5. Архитектура решения
   1. Функции считывания информации
   2. Функции обработки информации
   3. Функции вывода информации
6. Тестирование
7. Заключение
8. **Постановка задачи (физическая модель)**

Спрогнозировать какой из критериев будет больше влиять на стоимость автомобиля (год выпуска автомобиля, пробег(в км), тип кузова (седан, хетчбэк, универсал и т.д.), коробка передач (автомат, механика), объем двигателя(в л), тип двигателя (бензиновый, дизельный, гибридный), привод (передний, задний, полный), руль (левый, правый), цвет, состояние (не битый, битый)).

1. **Математическая модель**

В разделе описывается формульные зависимости в общем виде необходимые для решения класса подобных задач.

1. **Алгоритмы**
   1. **Алгоритм 1**

Линейная регрессия

* + 1. **Описание входных данных**
       1. Массив предсказываемых данных
       2. Массив предикатов размерностью
       3. – параметр, отвечающий за вид регуляризации (по умолчанию (без регуляризации), может принимать значения
          1. Если , то вводим коэффициент регуляризации (; чем больше, тем сильнее регуляризация)
          2. Если , то вводим предполагаемое стандартное отклонение остатков (; чем больше, тем слабее регуляризация)
    2. **Описание алгоритма решения**
       1. Добавить к матрице колонку единиц слева
       2. Убедиться, что (если нет, то введены некорректные данные – имеются линейно зависимые предикаты – ошибка, выход из алгоритма)
       3. Составляем функцию потерь
          1. Если , то , где
          2. Если , то
          3. Если , то
          4. Если , то
       4. С помощью метода сопряжённых градиентов находим минимум функции ( – начальная точка)
       5. Вычисляем вектор модельных предсказанных данных
    3. **Описание выходных данных**

.На выходе получаем вектор весов.

* 1. **Алгоритм 2**

Полиномиальная регрессия.

* + 1. **Описание входных данных**
       1. Массив предсказываемых данных
       2. Массив предикатов размерностью
       3. степень полинома (чем больше, тем лучше аппроксимация, но при высокой степени будет переобучение)
       4. – параметр, отвечающий за вид регуляризации (по умолчанию (без регуляризации), может принимать значения
          1. Если , то вводим коэффициент регуляризации (; чем больше, тем сильнее регуляризация)
          2. Если , то вводим предполагаемое стандартное отклонение остатков (; чем больше, тем слабее регуляризация)
    2. **Описание алгоритма решения**
       1. Обновить матрицу с помощью Python:

from sklearn.preprocessing import PolynomialFeatures

X = PolynomialFeatures(deg).fit\_transform(X)

* + - 1. Составляем функцию потерь , где len(X[0])- 1
         1. Если , то , где
         2. Если , то
         3. Если , то
         4. Если , то
      2. С помощью метода сопряжённых градиентов находим минимум функции ( – начальная точка)
      3. Вычисляем вектор модельных предсказанных данных
    1. **Описание выходных данных**

На выходе получаем вектор весов.

* 1. **Алгоритм 3**

Экспоненциальная регрессия.

* + 1. **Описание входных данных**
       1. Массив предсказываемых данных
       2. Массив предикатов размерностью
       3. – параметр, отвечающий за вид регуляризации (по умолчанию (без регуляризации), может принимать значения
          1. Если , то вводим коэффициент регуляризации (; чем больше, тем сильнее регуляризация)
          2. Если , то вводим предполагаемое стандартное отклонение остатков (; чем больше, тем слабее регуляризация)
    2. **Описание алгоритма решения**
       1. Вычислим массив :
       2. Составляем функцию потерь
          1. Если , то , где
          2. Если , то
          3. Если , то
          4. Если , то
       3. С помощью метода сопряжённых градиентов находим минимум функции ( – начальная точка)
       4. Вычисляем вектор параметров :
       5. Вычисляем вектор модельных предсказанных данных
    3. **Описание выходных данных**

На выходе получаем вектор весов.

1. **Тестирование**

Таблица 1. Результаты тестирования программы

**Входные данные:**   
Массив предсказанных данных y = (1, 2, 3, 2, 1)

Массив предикатов X размерностью :

Параметр отвечающий за вид регуляции – None(без регуляции)

Для полиномиальной регрессии степень полинома 1

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Линейная регрессия** | **Полиномиальная регрессия** | **Экспоненциальная регрессия** |
| **Полученное решение** | Вектор модельных предсказанных данных: [1.97167800099167, 2.09431330914668, 1.84904269283666, 1.72640738468165, 1.35850146021662]  Массив коэффициентов регрессии: [-0.122635308155011]  Свободный член: 1.97167800099167  Функция в аналитическом виде: y^=1.97167800099167 + -0.122635308155011 \* x1 | Вектор модельных предсказанных данных: [1.97167800099167, 2.09431330914668, 1.84904269283666, 1.72640738468165, 1.35850146021662]  Массив коэффициентов регрессии: [-0.122635308155011]  Свободный член: 1.97167800099167  Функция в аналитическом виде: y^=1.97167800099167 + -0.122635308155011 \* x1 | Вектор модельных предсказанных данных: [7.18271899144544, 8.11986321953359, 6.35373413507259, 5.62042556687334, 3.89035907363203]  Массив коэффициентов регрессии: [0.884586205118123]  Свободный член: 7.18271899144544  Функция в аналитическом виде: y^=7.18271899144544 \* 0.884586205118123\*\*x1 |
| **Время исполнения (в секундах)** | 3.6043612957000732 | 3.7390010356903076 | 3.3779664039611816 |

Таблица 2. Результаты тестирования программы

**Входные данные:**   
Массив предсказанных данных y = (1, 2, 3, 2, 1)

Массив предикатов X размерностью :

Параметр отвечающий за вид регуляции – L1. Лямбда = 0.95

Для полиномиальной регрессии степень полинома 1

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Линейная регрессия** | **Полиномиальная регрессия** | **Экспоненциальная регрессия** |
| **Полученное решение** | Вектор модельных предсказанных данных: [1.06705899373755, 0.476305266704051, 1.28228026851944, 1.49750154330134, 1.39210046314384]  Массив коэффициентов регрессии: [-0.160311177469701, 0.375532452251598]  Свободный член: 0.691526541485948  Функция в аналитическом виде: y^=0.691526541485948 + -0.160311177469701 \* x1 + 0.375532452251598 \* x2 | Вектор модельных предсказанных данных: [1.06705899373755, 0.476305266704051, 1.28228026851944, 1.49750154330134, 1.39210046314384]  Массив коэффициентов регрессии: [-0.160311177469701, 0.375532452251598]  Свободный член: 0.691526541485948  Функция в аналитическом виде: y^=0.691526541485948 + -0.160311177469701 \* x1 + 0.375532452251598 \* x2 | Вектор модельных предсказанных данных: [2.90681794657233, 1.61011445527750, 3.60485038761067, 4.47050574060897, 4.02329196041247]  Массив коэффициентов регрессии: [0.851878662271006, 1.45576633435854]  Свободный член: 1.99676134690473  Функция в аналитическом виде: y^=1.99676134690473 \* 0.851878662271006\*\*x1 \* 1.45576633435854\*\*x2 |
| **Время исполнения (в секундах)** | 1.199791669845581 | 1.2217321395874023 | 1.1918120384216309 |

Таблица 3. Результаты тестирования программы

**Входные данные:**   
Массив предсказанных данных y = (1, 2, 3, 2, 1)

Массив предикатов X размерностью :

Параметр отвечающий за вид регуляции – L2. Лямбда = 0.95

Для полиномиальной регрессии степень полинома 1

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Линейная регрессия** | **Полиномиальная регрессия** | **Экспоненциальная регрессия** |
| **Полученное решение** | Вектор модельных предсказанных данных: [1.56046057072504, -0.0637871558360776, 1.90283759723199, 2.24521462373894, 0.708604303151448]  Массив коэффициентов регрессии: [-0.939493673547219, 1.28187070005417]  Свободный член: 0.278589870670871  Функция в аналитическом виде: y^=0.278589870670871 + -0.939493673547219 \* x1 + 1.28187070005417 \* x2 | Вектор модельных предсказанных данных: [1.56046057072504, -0.0637871558360776, 1.90283759723199, 2.24521462373894, 0.708604303151448]  Массив коэффициентов регрессии: [-0.939493673547219 1.28187070005417]  Свободный член: 0.278589870670871  Функция в аналитическом виде: y^=0.278589870670871 + -0.939493673547219 \* x1 + 1.28187070005417 \* x2 | Вектор модельных предсказанных данных: [4.76101352369997, 0.938204669669168, 6.70489326059104, 9.44244191118834, 2.03115440361240]  Массив коэффициентов регрессии: [0.390825670645149, 3.60337425687880]  Свободный член: 1.32126534306317  Функция в аналитическом виде: y^=1.32126534306317 \* 0.390825670645149\*\*x1 \* 3.60337425687880\*\*x2 |
| **Время исполнения (в секундах)** | 3.6043612957000732 | 3.3330869674682617 | 3.7001049518585205 |

Таблица 4. Результаты тестирования программы

**Входные данные:**   
Массив предсказанных данных y = (1, 2, 3, 2, 1)

Массив предикатов X размерностью :

Параметр отвечающий за вид регуляции – norm. = 0.1

Для полиномиальной регрессии степень полинома 1

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Линейная регрессия** | **Полиномиальная регрессия** | **Экспоненциальная регрессия** |
| **Полученное решение** | Вектор модельных предсказанных данных: [1.48069444353561, -0.339329399871366, 1.88107952216139, 2.28146460078716, 0.643342307102093]  Массив коэффициентов регрессии: [-1.01925368615542, 1.41963876478120]  Свободный член: 0.0610556787544107  Функция в аналитическом виде: y^=0.0610556787544107 + -1.01925368615542 \* x1 + 1.41963876478120 \* x2 | Вектор модельных предсказанных данных: [1.48069444353561, -0.339329399871366, 1.88107952216139, 2.28146460078716, 0.643342307102093]  Массив коэффициентов регрессии: [-1.01925368615542, 1.41963876478120]  Свободный член: 0.0610556787544107  Функция в аналитическом виде: y^=0.0610556787544107 + -1.01925368615542 \* x1 + 1.41963876478120 \* x2 | Вектор модельных предсказанных данных: [4.39599739314743, 0.712247796111563, 6.56058333592186, 9.79100983423445, 1.90283010543803]  Массив коэффициентов регрессии: [0.360864157617071, 4.13562623653971]  Свободный член: 1.06295809672239  Функция в аналитическом виде: y^=1.06295809672239 \* 0.360864157617071\*\*x1 \* 4.13562623653971\*\*x2 |
| **Время исполнения (в секундах)** | 5.206078290939331 | 5.011598348617554 | 4.981677293777466 |