

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО
ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ**
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ Н.Э. БАУМАНА
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»
(МГТУ им. Н.Э. Баумана)
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ НИЖЕГОРОДСКОЙ
ОБЛАСТИ
ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ
ИНЖЕНЕРНО-ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ»
(ГБОУ ВО НГИЭУ)
Институт: «Информационные технологии и системы
связи»
Кафедра: «Математика и вычислительная техника»

ОТЧЁТ ПО ПРОЕКТНО-ТЕХНОЛОГИЧЕСКОЙ ПРАКТИКЕ

Выполнил: студент 1-ого курса
института ИТиСС, группы 24ИВО-2
Лукьянов Алексей Львович

Руководитель:
К.э.н., доцент кафедры «Математика и вычислительная техника»
Черемухин Артём Дмитриевич

Работа защищена
«01» «июля» «2025» года
С оценкой «
»

Княгинино
2025 год

Содержание

ВВЕДЕНИЕ

Обработка и анализ данных играют ключевую роль в современных технологиях, особенно в областях маркетинга, финансов и научных исследований. Например, множественная регрессия используется для прогнозирования цен на недвижимость на основе таких факторов, как площадь, расположение и год постройки, что помогает принимать обоснованные решения в реальном времени. Актуальность данного метода обусловлена ростом объёма данных и необходимостью их эффективной интерпретации, что подчёркивается в источниках, доступных по ссылке <http://www.аспирантура.пф/aktualnost>.

Целью работы является разработка модульного приложения на языке C++ для реализации метода множественной регрессии и его тестирование на синтетических данных. Задачи включают: изучение теоретических основ метода, разработку псевдокода и блок-схемы, создание программного обеспечения, реализацию алгоритма с поддержкой различных модификаций (например, Ridge-регрессия), тестирование приложения и составление руководства пользователя, как указано в <http://www.аспирантура.пф/predmet>.

Объектом исследования является процесс анализа данных с использованием регрессионных методов. Предметом выступает программная реализация метода множественной регрессии на C++ с акцентом на предобработку данных, отбор значимых признаков и оценку качества модели, что соответствует материалам на <http://www.аспирантура.пф/predmet>.

ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ МЕТОДА МНОЖЕСТВЕННОЙ РЕГРЕССИИ

0.1 Обзор литературы по методу множественной регрессии

Метод множественной регрессии представляет собой статистический подход для моделирования зависимости одной целевой переменной от нескольких независимых переменных. Основная идея заключается в минимизации суммы квадратов ошибок между наблюдаемыми и предсказанными значениями, что реализовано через метод наименьших квадратов (МНК). Этот метод широко применяется в экономике для прогнозирования спроса, в медицине для анализа влияния факторов на здоровье и в инженерии для оптимизации процессов.

Математическая основа метода выражается через уравнение $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \epsilon$, где β_i — коэффициенты регрессии, x_i — независимые переменные, а ϵ — случайная ошибка. В матричной форме решение находится как $\beta = (X^T X)^{-1} X^T y$, где X — матрица признаков, y — вектор целевых значений. При мультиколлинеарности используются модификации, такие как гребневая регрессия.

Пример из практики: прогнозирование продаж в интернет-магазине на основе рекламы (x_1), цены (x_2) и сезона (x_3). Данные о 50 товарах с известными продажами (y) позволяют построить модель, например, $y = 100 + 5x_1 - 2x_2 + 10x_3$, где коэффициенты определяются МНК. Это демонстрирует практическую применимость метода.

Для обзора изучено 7 источников, включая научные статьи и учебные материалы, отобранных по критериям наличия формул, примеров и сравнения с другими методами. Полезным оказался анализ регуляризации из работ Hastie et al. (2009), который будет учтён при реализации Ridge-регрессии. Противоречия между источниками минимальны, но некоторые предлагают разные пороги корреляции для отбора признаков (0.1 против 0.05), что требует тестирования. Вывод: метод требует внимания к предобработке данных и выбору значимых переменных.

0.2 Составление псевдокода и блок-схемы реализации алгоритма на языке C++

Псевдокод алгоритма множественной регрессии описан ниже:

```
// 1:
-
-
xi
xi
-
-
(xi)
-
(xi)
-
25(xi)
-
75(xi)
-
- 1.5 *
+ 1.5 *
(xi < xi > )
-
(xi)
-
xi
xi_ ( ) / ( )
-
(xi)
-
(xi)
-
xi OneHot(xi)
-
-
```

```
// 2:
-
- ( _ )
xi -
(xi, _ )
( _ ) >= 0.1
xi -

(xi, xj) -
( _ [xi, xj]) > 0.8
(xi, xj) -

// 3:
( )
. beta_0 + (beta_i * xi xi _ )
. _ y = X * beta +

// 4:
- ( _ )
- ( _ )
_XTX _ * _
_XTX
_XTX ( _ _XTX)
_beta _XTX * _ * _
_beta QR SVD( _ , _ )

_beta

// 5:
( _ , _ ) ( _ , _ )
( , _ )
-
- (( _ - ( _ ))^2)
- ( ^2)
R_ 1 - ( _ _ / _ _ )
MAPE ( ( / _ )) * 100

// 6:
_lambda ( )
```

```

        _ridge      (      _XTX +      _lambda *      ) *      _
    _ridge      (      _ridge)
        (      _      )
        _      (      _      ,      _      ,      )
    _      (      _      )
    _huber      (      _      ,      _      ,      _delta)
    _huber      (      _huber)
        _      (      _      )
        _      (      _      )
    _      (      _      ,      _      )
//      7:
    , R_      , MAPE, (      _ridge,      _      ,      _huber,

```

Блок-схема алгоритма представлена в приложении А.