**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ**

**Московский государственный технический университет**

**им. Н.Э. Баумана**

**(МГТУ им. Н.Э. Баумана)**

**Кафедра «Информационная безопасность» (ИУ8)**

Лабораторная работа № 3.3

По дисциплине: «Машинное обучение»

Тема: «Модели регрессии»

Выполнил: Веденеев А.А.,

Студент группы ИУ8-92

Проверила: Коннова Н.С.,

Преподаватель каф. ИУ8

г. Москва, 2024 г.

# Практическая часть лабораторной работы 3.3

**Цель работы:** Познакомиться с основными приемами работы с моделями регрессии в scikit-learn.

**Ход работы:**

**A screenshot of a computer

Description automatically generated**

**A white rectangular object with a black border

Description automatically generated**

A screenshot of a computer

Description automatically generated

A screen shot of a graph

Description automatically generated

A screen shot of a graph

Description automatically generated

A screen shot of a graph

Description automatically generated

A screen shot of a graph

Description automatically generated

A screen shot of a graph

Description automatically generated

A screenshot of a computer

Description automatically generated

## Вывод

В ходе лабораторной работы были рассмотрены различные модели регрессии: логистическая регрессия, полиномиальная регрессия, SVM, дерево решений, случайный лес и многослойный перцептрон (MLP). Анализ показал, что каждая модель обладает уникальными характеристиками и показателями производительности в зависимости от природы данных. Логистическая регрессия и SVM хорошо справляются с линейными зависимостями, в то время как полиномиальная регрессия эффективна для нелинейных зависимостей. Дерево решений и случайный лес демонстрируют устойчивость к шуму и сложным паттернам в данных, но случайный лес снижает риск переобучения. MLP показывает высокую гибкость и способность моделировать сложные нелинейные зависимости, но требует больших вычислительных ресурсов и внимательной настройки. Выбор модели зависит от характеристик данных и требований к задаче.

**Контрольные вопросы**

1. Почему SVM отработал так плохо (вопрос от преподавателя)?

SVM плохо работает на датасете California Housing из-за высокой размерности данных, большого количества выбросов и сложных нелинейных зависимостей между признаками и целевой переменной. Также модель чувствительна к масштабированию данных, и если их не нормализировать, производительность значительно снижается. В некоторых случаях из-за большого количества признаков и корреляций между ними SVM теряет способность обобщать закономерности, что приводит к низкой точности.

Оптимальные настройки гиперпараметров, такие как параметры C и gamma, могут существенно повлиять на результат, но из-за вычислительной сложности и особенностей данных SVM часто показывает низкую производительность по сравнению с более устойчивыми методами, такими как случайный лес или деревья решений. При подборе гиперпараметров C и gamma с ядром rbf можно получить более приемлемый резльутат:

A screen shot of a graph

Description automatically generated

1. Чем отличается применение разных моделей регрессии в бибилиотеке *sklearn* от моделей классфикации?

В моделях регрессии целевая переменная представляет собой непрерывные числовые значения, и задача заключается в прогнозировании конкретного числового значения. В моделях классификации целевая переменная представляет собой категориальные данные, и задача заключается в предсказании принадлежности объекта к определённому классу. Соответственно, в регрессии используются метрики, такие как MSE или R², а в классификации — точность, матрица ошибок и другие метрики. Также алгоритмы могут отличаться подходами к разделению данных и оптимизации.

1. Что показывает коэффициент детерминации для модели регрессии?

Коэффициент детерминации (R^2) показывает, какую долю вариации целевой переменной модель способна объяснить на основе входных данных. Значение R^2 варьируется от 0 до 1:

• R^2 = 1: модель идеально объясняет все вариации данных.

• R^2 = 0: модель не объясняет вариаций данных.

• R^2 < 0: модель работает хуже, чем просто предсказание среднего значения целевой переменной.

Он позволяет оценить качество модели и её способность обобщать закономерности в данных.

1. Какое значение имеют коэффициенты линейной регрессии?

Коэффициенты линейной регрессии показывают степень влияния каждого признака на целевую переменную. Они определяют, как изменяется целевая переменная при изменении соответствующего признака на одну единицу, при условии, что остальные признаки остаются неизменными.

Положительный коэффициент означает, что увеличение признака приводит к увеличению значения целевой переменной, а отрицательный — что увеличение признака снижает значение целевой переменной. Коэффициенты позволяют интерпретировать связь между признаками и целевой переменной.

1. Какие атрибуты имеет объект линейной регрессии?

coef\_ содержит коэффициенты модели для каждого признака после обучения. intercept\_ представляет свободный член (смещение) модели. Атрибут classes\_ отсутствует, так как линейная регрессия используется для предсказания числовых значений, а не классов. n\_iter\_ показывает количество итераций, выполнённых для сходимости. Метод score используется для оценки качества модели на основе R^2.