**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ**

**Московский государственный технический университет**

**им. Н.Э. Баумана**

**(МГТУ им. Н.Э. Баумана)**

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

Лабораторная работа № 3.1

По дисциплине: «Машинное обучение»

Тема: «Модели классификации»

Выполнил: Веденеев А.А.,

Студент группы ИУ8-92

Проверила: Коннова Н.С.,

Преподаватель каф. ИУ8

г. Москва, 2024 г.

# Практическая часть лабораторной работы 3.1

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

**Ход работы:**

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated



A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

## Вывод

В лабораторной работе были рассмотрены несколько моделей классификации: логистическая регрессия, полиномиальный классификатор, SVM, дерево решений, случайный лес и многослойный перцептрон (MLP). Анализ показал, что каждая из моделей имеет свои особенности. Логистическая регрессия и SVM хорошо подходят для линейно разделимых данных, тогда как полиномиальный классификатор позволяет учитывать нелинейные зависимости. Дерево решений эффективно работает с данными сложной структуры, но подвержено переобучению, в то время как случайный лес улучшает устойчивость за счёт ансамблирования. MLP продемонстрировал высокую гибкость и способность справляться с нелинейными зависимостями, но требует тщательной настройки и больше ресурсов. Выбор оптимальной модели зависит от структуры данных, требований к точности и вычислительных ограничений.

**Контрольные вопросы**

1. Чем отличается применение разных моделей классификации в бибилиотеке sklearn?

Применение разных моделей классификации в sklearn отличается принципами работы моделей, их устойчивостью к переобучению и способностью обрабатывать различные типы данных. Линейные модели, такие как логистическая регрессия, хорошо работают с линейно разделимыми данными, тогда как деревья решений и случайные леса способны обрабатывать данные со сложной структурой, но первые склонны к переобучению. SVM эффективно разделяет данные с использованием гиперплоскостей и ядерных функций, что позволяет учитывать нелинейные зависимости. Нейронные сети (MLP) подходят для задач с более сложными паттернами, но требуют больших вычислительных ресурсов. Выбор модели зависит от характера данных, их объёма и требований к точности классификации.

1. Что показывает метрика точности регрессии?

Метрика точности регрессии показывает, насколько предсказанные значения модели близки к реальным значениям целевой переменной. Она оценивает способность модели объяснять вариацию данных и минимизировать ошибку предсказания. В зависимости от используемой метрики (MSE, MAE, R^2) точность может выражаться в виде среднего отклонения, суммы квадратов ошибок или доли объяснённой дисперсии.

1. Какое значение имеют коэффициенты логистической регрессии?

Коэффициенты логистической регрессии отражают влияние каждого признака на вероятность отнесения объекта к определённому классу. Они показывают, насколько изменение значения признака на одну единицу изменяет логарифм шансов принадлежности к классу, при условии, что остальные признаки остаются неизменными. Положительный коэффициент увеличивает шансы, отрицательный — уменьшает.

1. Что показывает матрица классификации?

Матрица классификации показывает, как модель предсказывает классы для каждого наблюдения по сравнению с их истинными значениями. Она содержит информацию о количествах верных и неверных предсказаний для каждого класса, отображая, сколько объектов было правильно классифицировано и сколько ошибочно. Это позволяет оценить качество модели, выявить смещения и анализировать, какие классы модель путает между собой.

1. Какие параметры имеет конструктор объекта логистической регрессии?

Конструктор объекта логистической регрессии в библиотеке sklearn имеет несколько ключевых параметров. Основные из них включают penalty, который определяет тип регуляризации (L1, L2, ElasticNet или None), и C, который настраивает силу регуляризации. Также присутствует параметр solver, определяющий метод оптимизации (например, ‘liblinear’, ‘lbfgs’, ‘newton-cg’). Параметр max\_iter задаёт максимальное количество итераций для сходимости алгоритма, а random\_state обеспечивает воспроизводимость результатов. Дополнительно можно настроить параметры fit\_intercept, чтобы определять, учитывать ли константу в модели, и class\_weight, чтобы учесть дисбаланс классов.

1. Какие атрибуты имеет объект логистической регрессии?

coef\_ содержит коэффициенты модели для каждого признака после обучения. intercept\_ представляет свободный член (смещение) модели. Атрибут classes\_ содержит список классов, которые модель может предсказывать. n\_iter\_ показывает количество итераций, выполненных алгоритмом для достижения сходимости. Также может присутствовать score, который позволяет оценивать качество модели на новых данных.

1. Какие параметры и атрибуты имеют объекты других моделей машинного обучения библиотеки sklearn?

Основные параметры включают гиперпараметры обучения, такие как регуляризация (C, penalty), метод оптимизации (solver), количество итераций (max\_iter) и вес классов (class\_weight). Атрибуты включают обученные коэффициенты (coef\_), смещение (intercept\_), список классов (classes\_) и количество итераций (n\_iter\_). Например, у деревьев решений есть feature\_importances\_ (влияние признаков), а у случайного леса — estimators\_ (обученные деревья).