

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Московский Авиационный Институт

(Национальный исследовательский университет)

Институт №8

«Компьютерные науки и прикладная математика»

Кафедра 806

«Вычислительная математика и программирование»

Курсовая работа по дисциплине

«Системы интеллектуальной поддержки»

Тема «Анализ алгоритмов машинного обучения в контексте
прикладных задач»

Студент: Новиков С.С.

Группа: М8О-311Б-22

Преподаватель: Крылов С.С.

Оценка:

Дата:

Москва 2025

Содержание

Введение	3
Домашняя работа 1	5
Домашняя работа 2	6
Домашняя работа 3	7
Лабораторная 1	8
Лабораторная 2	9
Вывод	10

Введение

Машинное обучение (МО) является одной из наиболее динамично развивающихся областей искусственного интеллекта, находящей применение в широчайшем спектре задач – от анализа пользовательских предпочтений и медицинской диагностики до создания автономных транспортных средств и научных открытий. В основе МО лежит идея о том, что компьютерные системы могут обучаться на основе данных, выявлять в них закономерности и принимать решения с минимальным вмешательством человека. Данная курсовая работа посвящена исследованию и практическому применению фундаментальных методов машинного обучения, таких как классификация, кластеризация и снижение размерности.

Центральное место в работе занимает задача классификации – процесс отнесения объектов к одному из заранее определенных классов на основе их признаков. Мы углубимся в изучение одного из наиболее интуитивно понятных и широко используемых алгоритмов классификации – деревьев решений. Будет рассмотрена как стандартная реализация деревьев решений из популярной библиотеки `scikit-learn`, так и предпринята попытка разработки собственного алгоритма. Анализ будет проводиться на разнообразных наборах данных, что позволит оценить эффективность и ограничения деревьев решений в различных условиях.

Другим важным направлением исследования в данной работе является обучение без учителя, в частности, методы кластеризации и снижения размерности. В отличие от классификации, где метки классов известны заранее, кластеризация направлена на группировку объектов на основе их сходства, без предварительной информации о структуре данных. Мы исследуем, как можно выявить естественные группы или кластеры в данных.

Для анализа таких многомерных данных, как показания акселерометров и гироскопов из набора данных "Samsung Human Activity Recognition", часто необходимо предварительно снизить их размерность. Методы снижения размерности позволяют преобразовать данные из пространства высокой размерности в пространство более низкой размерности при сохранении наиболее значимой информации. Это не только упрощает последующий

анализ и визуализацию данных, но и может улучшить производительность моделей МО за счет удаления шума и избыточных признаков. В работе будут рассмотрены и применены соответствующие техники для предобработки данных перед их кластеризацией и последующей классификацией.

Практическая часть работы будет включать в себя применение этих методов к реальному набору данных о физической активности человека. Изначально мы подойдем к этой задаче с точки зрения обучения без учителя, попытаюсь кластеризовать виды активности на основе сенсорных данных. Затем мы решим ту же задачу как проблему классификации, используя известные метки активности для обучения моделей и оценки их точности. Такой двухэтапный подход позволит сравнить возможности различных парадигм машинного обучения на одной и той же предметной области.

В ходе выполнения курсовой работы будут решаться следующие ключевые задачи:

- Изучение теоретических основ деревьев решений, методов кластеризации и снижения размерности.
- Практическое применение деревьев решений из библиотеки `scikit-learn` для задач классификации, анализ влияния гиперпараметров на их поведение и качество.
- Реализация собственного алгоритма построения решающего дерева и его сравнение с готовыми решениями.
- Исследование различных подходов к кодированию категориальных признаков и их влияния на производительность моделей.
- Применение методов снижения размерности для предобработки данных и методов кластеризации для выявления структуры в данных о человеческой активности.
- Решение задачи классификации видов физической активности человека на основе данных с мобильных сенсоров.
- Анализ полученных результатов, оценка качества моделей и формулирование выводов о применимости рассмотренных методов.

Домашняя работа 1

Домашняя работа 2

Домашняя работа 3

Лабораторная 1

Лабораторная 2

Вывод

В ходе данной курсовой работы было проведено комплексное исследование и практическое применение фундаментальных методов машинного обучения, охватывающих как обучение с учителем, так и обучение без учителя. Основное внимание было уделено деревьям решений: их применению с использованием библиотеки `scikit-learn` для задач классификации, анализу влияния гиперпараметров на формируемые разделяющие поверхности, а также разработке и тестированию собственной реализации данного алгоритма. Кроме того, были рассмотрены и применены методы снижения размерности и кластеризации на примере набора данных "Samsung Human Activity Recognition", что позволило сначала исследовать структуру данных без использования меток классов, а затем перейти к задаче классификации видов физической активности.

Полученные результаты продемонстрировали эффективность деревьев решений в задачах классификации и важность тщательного подбора гиперпараметров. Сравнение собственной реализации с эталонной из `scikit-learn` позволило глубже понять внутренние механизмы алгоритма и выявить ключевые аспекты его производительности. Исследование методов кодирования категориальных признаков подчеркнуло их значимость для корректной работы моделей. Работа с методами снижения размерности и кластеризации подтвердила их полезность для предварительного анализа данных, выявления внутренней структуры и подготовки данных для последующего моделирования. В целом, выполненные задачи способствовали углублению понимания теоретических основ и развитию практических навыков применения рассмотренных методов машинного обучения для решения реальных прикладных задач.