

АННОТАЦИЯ

учебной дисциплины Б1.В.ДВ.01.01 «Анализ и обработка больших данных» по образовательной программе направления подготовки 09.06.01 – Информатика и вычислительная техника, направленность «Системный анализ, управление и обработка информации» квалификация (степень) выпускника: исследователь. преподаватель-исследователь.

Цель дисциплины – формирование у аспирантов теоретических знаний и практических навыков обработки и анализа большого объема данных.

Задачи дисциплины – изучение алгоритмов методов обработки и анализа большого объема данных, их программной реализации в R.

Объем дисциплины: в 3 зачетных единицах и 108 часах

Семестр: 3

Краткое содержание дисциплины:

1. Об анализе больших данных в целом.

Сведения из истории развития и связь анализа больших данных с другими дисциплинами: базами данных и базами знаний, машинным обучением, искусственным интеллектом и пр. Основные понятия анализа больших данных, терминология. Основные классы задач анализа больших данных: прогнозирование, классификация, кластеризация, поиск ассоциативных правил и последовательных ассоциаций, поиск отклонений. Визуализация данных. Особенности больших данных. Предварительная обработка больших данных. Установка пакета R и интерфейса RStudio. Основные сведения о структуре пакета.

2. Задача прогнозирования и регрессионный анализ

Постановка задачи. Линейная регрессионная модель (ЛРМ). Метод наименьших квадратов (МНК). Применение МНК в однофакторной задаче. Визуальная оценка качества регрессионной модели. Применение МНК в многофакторной задаче. Проблема мультиколлинеарности. Исследование качества модели. Применение ЛРМ в прогнозировании. Решение задач для реальных данных в среде R.

3. Задачи классификации и кластеризации.

Постановка задач классификации и кластеризации, сходство и различие постановок. Примеры прикладных задач. Понятие об обучении с учителем и без учителя. Методы решения задачи классификации: метод k ближайших соседей, метод наивного Байеса, метод деревьев решений, метод дискриминантного анализа, эвристические алгоритмы. Методы решения задачи кластеризации: метод k средних, метод главных компонент, эвристические алгоритмы. Иерархическая кластеризация. Числовые характеристики решения задач классификации и кластеризации. Решение задач для реальных данных в среде R.

Аудиторный курс включает в себя лекции и практические занятия.

Форма промежуточной аттестации – зачет с оценкой