

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Задорожная Людмила Ивановна
Должность: Проректор по учебной работе
Дата подписания: 29.07.2022 18:11:19
Уникальный программный ключ:
faa404d1aeb2a023b5f4a331eeb0d7a3e

АННОТАЦИЯ рабочей программы дисциплины

Математические методы анализа больших данных

1. Общая трудоёмкость

Трудоёмкость дисциплины составляет 7 зачётных единиц (252 часов), из них 34 часа лекционных занятий, 34 часа практических занятий.

2. Место дисциплины в структуре образовательной программы

Дисциплина относится к модулю обязательных профессиональных дисциплин обязательной части образовательной программы.

Для изучения данной дисциплины необходимы знания, умения и навыки, формируемые предшествующими элементами образовательной программы: Методы машинного обучения; Современные проблемы и методы прикладной информатики.

Результаты обучения, формируемые данной дисциплиной, потребуются при освоении следующих элементов образовательной программы: Технологии анализа больших данных; Интеллектуальные геоинформационные системы; Системы аналитики больших данных; Математические методы и модели поддержки принятия решений; Нейронные сети и глубокое обучение; производственная практика, проектно-технологическая практика; производственная практика, преддипломная практика.

3. Цель изучения дисциплины

Формировании у студентов способности применять математические методы и модели для обработки и анализа больших данных при решении профессиональных задач.

4. Содержание дисциплины

Модуль 1 «Математические методы машинного обучения для анализа больших данных»

1. Метрические алгоритмы. Байесовский подход к обучению

Обобщенный алгоритм. Примеры частных алгоритмов. Метод ближайших соседей. Метод окна Парзена. Понятие выступа объекта. Алгоритм отбора эталонов. Проклятие размерности. Выбор метрики. Применение формула Байеса к задаче машинного обучения. Построение приближенной плотности распределения: одномерный случай (дискретное, кусочно-постоянное, приближение с помощью ядра, приближение нормальным законом распределения), многомерный случай (наивный байесовский классификатор, приближение с помощью ядра, многомерное нормальное распределение). Логистическая регрессия. Бинаризация признаков. Скоринг. Смеси распределений. EM-алгоритм восстановления смеси.

2. Линейные алгоритмы классификации. Метод опорных векторов

Общая формула линейного классификатора. Метод стохастического градиента. Частные случаи. Обоснование метода СГ. Выступ объекта для линейного классификатора. ROC и AUC. Случай линейно разделимой и неразделимой выборки. Двойственная задача. Типы объектов. Нелинейное обобщение SVM. SVM-регрессия. L1 регуляризация

3. Методы восстановления регрессии. Логические методы классификации. Композиции алгоритмов

Метод наименьших квадратов. Геометрический смысл. Регуляризация. Сингулярное разложение. Непараметрическая регрессия. Понятие закономерности. Критерий качества закономерностей. Поиск закономерностей. Алгоритмы классификации на основе логических закономерностей. AdaBoost. AnyBoost. Градиентный бустинг. Бэггинг и метод случайных подпространств

4. Ранжирование и рекомендательные системы. Тематическое моделирование. Кластеризация

Оценки качества. Подходы к решению задачи: поточечный, попарный, списочный. Латентный семантический анализ (LSA). Вероятностный LSA (PLSA). Латентное размещение Дирихле (LDA). EM-алгоритм. Метод k- средних. DBSCAN

Модуль 2 «Искусственные нейронные сети»

1. Введение

Биологические предпосылки возникновения искусственных нейронных сетей. Структура человеческого мозга. Организация памяти в коре человеческого мозга. Ритмы колебаний больших нейронных ансамблей. Биологически правдоподобные модели нейронов. Модели визуального восприятия.

2. Искусственные нейронные сети

Типы функций активации нейронов. Представление нейронных сетей с помощью направленных графов. Архитектура сетей. Сети прямого распространения. Рекуррентные сети. Обучение нейронных сетей. Обучение, основанное на коррекции ошибок. Обучение на основе памяти. Обучение Хебба. Математические модели предложенного Хеббом механизма модификации синаптической связи. Конкурентное обучение. Обучение Больцмана. Обучение с учителем. Обучение с подкреплением. Обучение без учителя.

3. Однослойные и многослойные перцептроны

Однослойный перцептрон. Обучение перцептрона. Методы безусловной оптимизации. Метод наискорейшего спуска. Метод Ньютона. Метод Гаусса-Ньютона. Взаимосвязь перцептрона и байесовского классификатора. Многослойный перцептрон. Алгоритм обратного распространения ошибки. Извлечение признаков. Линейный дискриминант Фишера. Сети свертки.

4. Сети на основе радиальных базисных функций.

Теорема Ковера о разделимости множеств. Разделяющая способность поверхности. Задача интерполяции. Теория регуляризации. Функция Грина. Решение задачи регуляризации. Многомерные функции Гаусса. Обобщенные сети на основе радиальных базисных функций. Свойства аппроксимации сетей RBF. Сравнение сетей RBF и многослойных перцептронов.

5. Анализ главных компонент

Анализ признаков на основе самоорганизации. Структура анализа главных компонент. Представление данных. Сокращение размерности. Фильтр Хебба для выделения максимальных собственных значений. Анализ главных компонент на основе правила Хебба.

6. Карты самоорганизации Кохонена

Модели отображения признаков. Карты самоорганизации. Процессы конкуренции, кооперации и адаптации. Варианты самоорганизующихся карт. Адаптивные тензорные веса. Самоорганизующиеся карты для символьных строк. Самоорганизующиеся карты с эволюционным обучением. Пакеты программ, реализующие самоорганизующиеся карты.

7. Нейродинамические модели

Динамические системы. Пространство состояний. Условие Лившица. Теорема о дивергенции. Устойчивость состояний равновесия. Теоремы Ляпунова. Гиперболические аттракторы. Аддитивная и связанная нейродинамические модели. Модель Хопфилда. Теорема Коэна-Гроссберга.

5. Дополнительная полезная информация

Дисциплина предназначена для формирования элементов следующих компетенций образовательной программы:

ПК-1. Способен адаптировать и применять методы и алгоритмы машинного обучения для решения прикладных задач в различных предметных областях.

ПК-3. Способен руководить проектами со стороны заказчика по созданию, поддержке и использованию системы искусственного интеллекта на основе нейросетевых моделей и методов.

Форма промежуточной аттестации: экзамен.

Наименование оценочного средства: вопросы для собеседования; практикум; экзаменационные вопросы и билеты.