

Документ подписан простой электронной подписью.
Информация о владельце:

ФИО: Кандрашина Елена Александровна

Должность: И.о. ректора ФГАОУ ВО «Самарский государственный экономический университет»

Дата подписания: 07.08.2025 15:02:37

Уникальный программный ключ:

2db64eb9605ce27edd3b8e8fdd32c70e0674ddd2

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Самарский государственный экономический университет»

Институт Институт национальной и мировой экономики
Кафедра Статистики и эконометрики

УТВЕРЖДЕНО

Ученым советом Университета
(протокол № 10 от 22 мая 2025 г.)

РАБОЧАЯ ПРОГРАММА

Наименование дисциплины	Б1.В.01 Методы машинного обучения
Основная профессиональная образовательная программа	09.04.03 Прикладная информатика программа Цифровая экономика: анализ и управление данными

Квалификация (степень) выпускника магистр

Содержание (рабочая программа)

Стр.

- 1 Место дисциплины в структуре ОП
- 2 Планируемые результаты обучения по дисциплине, обеспечивающие достижение планируемых результатов обучения по программе
- 3 Объем и виды учебной работы
- 4 Содержание дисциплины
- 5 Материально-техническое и учебно-методическое обеспечение дисциплины
- 6 Фонд оценочных средств по дисциплине

Целью изучения дисциплины является формирование результатов обучения, обеспечивающих достижение планируемых результатов освоения образовательной программы.

1. Место дисциплины в структуре ОП

Дисциплина Методы машинного обучения входит в часть, формируемая участниками образовательных отношений блока Б1.Дисциплины (модули)

Последующие дисциплины по связям компетенций: Современные методы проектирования систем искусственного интеллекта, Технологии разработки программного обеспечения, Инструменты управления бизнес-процессами: программные решения, Моделирование бизнес-процессов и оптимизация системы управления

2. Планируемые результаты обучения по дисциплине, обеспечивающие достижение планируемых результатов обучения по программе

Изучение дисциплины Методы машинного обучения в образовательной программе направлено на формирование у обучающихся следующих компетенций:

Профессиональные компетенции (ПК):

ПК-2 - Способен разрабатывать инструменты и методы сбора исходных данных в рамках управления работами по сопровождению и проектами создания (модификации) ИС

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
ПК-2	ПК-2.1: Знать:	ПК-2.2: Уметь:	ПК-2.3: Владеть (иметь навыки):
	основы теории систем и системного анализа, источники информации, необходимой для профессиональной деятельности при управлении работами по сопровождению и проектами создания (модификации) ИС	разрабатывать и анализировать исходную документацию в рамках управления работами по сопровождению и проектами создания (модификации) ИС, использовать инструменты и методы определения финансовых и производственных показателей деятельности организаций	навыками ведения документооборота в организациях, инструментами и методами моделирования бизнес-процессов в ИС

3. Объем и виды учебной работы

Учебным планом предусматриваются следующие виды учебной работы по дисциплине:

Очная форма обучения

Виды учебной работы	Всего час/ з.е.
	Сем 2
Контактная работа, в том числе:	12.15/0.34
Занятия семинарского типа	12/0.33
Индивидуальная контактная работа (ИКР)	0.15/0
Самостоятельная работа:	113.85/3.16
Промежуточная аттестация	18/0.5
Вид промежуточной аттестации:	
Зачет	Зач
Общая трудоемкость (объем части образовательной программы): Часы	144
Зачетные единицы	4

заочная форма

Виды учебной работы	Всего час/ з.е.
	Сем 4
Контактная работа, в том числе:	6.15/0.17
Занятия семинарского типа	6/0.17
Индивидуальная контактная работа (ИКР)	0.15/0
Самостоятельная работа:	119.85/3.33
Промежуточная аттестация	18/0.5
Вид промежуточной аттестации: Зачет	Зач
Общая трудоемкость (объем части образовательной программы): Часы	144
Зачетные единицы	4

4. Содержание дисциплины

4.1. Разделы, темы дисциплины и виды занятий:

Тематический план дисциплины Методы машинного обучения представлен в таблице.

**Разделы, темы дисциплины и виды занятий
Очная форма обучения**

№ п/п	Наименование темы (раздела) дисциплины	Контактная работа			Самостоятельная работа	Планируемые результаты обучения в соотношении с результатами обучения по образовательной программе
		Занятия семинарского типа	ИКР	ГКР		
		Практич. занятия				
1.	Математические основы машинного обучения. Задачи классификации	6	0.1		60	ПК-2.1, ПК-2.2, ПК-2.3
2.	Деревья в машинном обучении. Случайный лес. Линейные классификаторы: основные понятия	6	0.05		53.85	ПК-2.1, ПК-2.2, ПК-2.3
	Контроль	18				
	Итого	12	0.15		113.85	

заочная форма

№ п/п	Наименование темы (раздела) дисциплины	Контактная работа			Самостоятельная работа	Планируемые результаты обучения в соотношении с результатами обучения по образовательной программе
		Занятия семинарского типа	ИКР	ГКР		
		Практич. занятия				
1.	Математические основы машинного обучения. Задачи классификации	4	0.1		60	ПК-2.1, ПК-2.2, ПК-2.3
2.	Деревья в машинном обучении. Случайный лес. Линейные классификаторы: основные понятия	2	0.05		59.85	ПК-2.1, ПК-2.2, ПК-2.3
	Контроль	18				
	Итого	6	0.15		119.85	

4.2 Содержание разделов и тем

4.2.1 Контактная работа

Тематика занятий семинарского типа

№п/п	Наименование темы (раздела) дисциплины	Вид занятия семинарского типа**	Тематика занятия семинарского типа
1.	Математические основы машинного обучения. Задачи классификации	практическое занятие	Выполнение упражнений на kaggle.com в интерактивном датасете
		практическое занятие	Выполнение упражнений на kaggle.com в интерактивном датасете
		практическое занятие	Выполнение упражнений на kaggle.com в интерактивном датасете
2.	Деревья в машинном обучении. Случайный лес. Линейные классификаторы: основные понятия	практическое занятие	Выполнение упражнений на kaggle.com в интерактивном датасете
		практическое занятие	Выполнение упражнений на kaggle.com в интерактивном датасете
		практическое занятие	Выполнение упражнений на kaggle.com в интерактивном датасете

** семинары, практические занятия, практикумы, лабораторные работы, коллоквиумы и иные аналогичные занятия

Иная контактная работа

При проведении учебных занятий СГЭУ обеспечивает развитие у обучающихся навыков командной работы, межличностной коммуникации, принятия решений, лидерских качеств (включая при необходимости проведение интерактивных лекций, групповых дискуссий, ролевых игр, тренингов, анализ ситуаций и имитационных моделей, преподавание дисциплин (модулей) в форме курсов, составленных на основе результатов научных исследований, проводимых организацией, в том числе с учетом региональных особенностей профессиональной деятельности выпускников и потребностей работодателей).

Формы и методы проведения иной контактной работы приведены в Методических указаниях по основной профессиональной образовательной программе.

4.2.2 Самостоятельная работа

№п/п	Наименование темы (раздела) дисциплины	Вид самостоятельной работы ***
1.	Математические основы машинного обучения. Задачи классификации	- подготовка доклада - подготовка электронной презентации - тестирование
2.	Деревья в машинном обучении. Случайный лес. Линейные классификаторы: основные понятия	- подготовка доклада - подготовка электронной презентации - тестирование

*** самостоятельная работа в семестре, написание курсовых работ, докладов, выполнение контрольных работ

5. Материально-техническое и учебно-методическое обеспечение дисциплины

5.1 Литература:

Основная литература

1. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — 2-е изд. — Москва : Издательство Юрайт, 2025. — 89 с. — (Высшее образование). — ISBN 978-5-534-20732-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/558662>

2. Кудрявцев, В. Б. Интеллектуальные системы : учебник и практикум для вузов / В. Б. Кудрявцев, Э. Э. Гасанов, А. С. Подколзин. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2025. — 165 с. — (Высшее образование). — ISBN 978-5-534-07779-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/561954>

3. Новиков, Ф. А. Символический искусственный интеллект: математические основы представления знаний : учебник для вузов / Ф. А. Новиков. — Москва : Издательство Юрайт, 2025.

— 278 с. — (Высшее образование). — ISBN 978-5-534-00734-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/561410>

Дополнительная литература

1. Нугуманова, А. Б. Автоматизированная обработка текстовых массивов : учебник и практикум для вузов / А. Б. Нугуманова. — 2-е изд. — Москва : Издательство Юрайт, 2024. — 82 с. — (Высшее образование). — ISBN 978-5-534-20738-5. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/558668>

2. Бессмертный, И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — 2-е изд. — Москва : Издательство Юрайт, 2025. — 250 с. — (Высшее образование). — ISBN 978-5-534-20734-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/558664>

3. Бессмертный, И. А. Системы искусственного интеллекта : учебник для вузов / И. А. Бессмертный. — 3-е изд., испр. и доп. — Москва : Издательство Юрайт, 2025. — 164 с. — (Высшее образование). — ISBN 978-5-534-18416-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/561602>

5.2. Перечень лицензионного программного обеспечения

1. Astra Linux Special Edition «Смоленск», «Орел»; РедОС ; ОС "Альт Рабочая станция" 10; ОС "Альт Образование" 10

2. МойОфис Стандартный 2, МойОфис Образование, Р7-Офис Профессиональный, МойОфис Стандартный 3, МойОфис Профессиональный 3

5.3 Современные профессиональные базы данных, к которым обеспечивается доступ обучающихся

1. Профессиональная база данных «Информационные системы Министерства экономического развития Российской Федерации в сети Интернет» (Портал «Официальная Россия» - <http://www.gov.ru/>)

2. Государственная система правовой информации «Официальный интернет-портал правовой информации» (<http://pravo.gov.ru/>)

3. Профессиональная база данных «Финансово-экономические показатели Российской Федерации» (Официальный сайт Министерства финансов РФ - <https://www.minfin.ru/ru/>)

4. Профессиональная база данных «Официальная статистика» (Официальный сайт Федеральной службы государственной статистики - <http://www.gks.ru/>)

5.4. Информационно-справочные системы, к которым обеспечивается доступ обучающихся

1. Справочно-правовая система «Консультант Плюс»

2. Справочно-правовая система «ГАРАНТ-Максимум»

5.5. Специальные помещения

Учебные аудитории для проведения занятий лекционного типа	Комплекты ученической мебели Мультимедийный проектор Доска Экран
Учебные аудитории для проведения практических занятий (занятий семинарского типа)	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Учебные аудитории для групповых и индивидуальных консультаций	Комплекты ученической мебели Мультимедийный проектор Доска

	Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Учебные аудитории для текущего контроля и промежуточной аттестации	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Помещения для самостоятельной работы	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Помещения для хранения и профилактического обслуживания оборудования	Комплекты специализированной мебели для хранения оборудования

5.6 Лаборатории и лабораторное оборудование

6. Фонд оценочных средств по дисциплине Методы машинного обучения:

6.1. Контрольные мероприятия по дисциплине

Вид контроля	Форма контроля	Отметить нужное знаком « + »
Текущий контроль	Оценка докладов	+
	Устный/письменный опрос	+
	Тестирование	+
	Практические задачи	+
Промежуточный контроль	Зачет	+

Порядок проведения мероприятий текущего и промежуточного контроля определяется Методическими указаниями по основной профессиональной образовательной программе высшего образования; Положением о балльно-рейтинговой системе оценки успеваемости обучающихся по основным образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры в федеральном государственном автономном образовательном учреждении высшего образования «Самарский государственный экономический университет».

6.2. Планируемые результаты обучения по дисциплине, обеспечивающие достижение планируемых результатов обучения по программе

Профессиональные компетенции (ПК):

ПК-2 - Способен разрабатывать инструменты и методы сбора исходных данных в рамках управления работами по сопровождению и проектами создания (модификации) ИС

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
	ПК-2.1: Знать: основы теории систем и	ПК-2.2: Уметь: разрабатывать и	ПК-2.3: Владеть (иметь навыки): навыками ведения

	системного анализа, источники информации, необходимой для профессиональной деятельности при управлении работами по сопровождению и проектами создания (модификации) ИС	анализировать исходную документацию в рамках управления работами по сопровождению и проектами создания (модификации) ИС, использовать инструменты и методы определения финансовых и производственных показателей деятельности организаций	документооборота в организациях, инструментами и методами моделирования бизнес-процессов в ИС
Пороговый	основы теории систем, источники информации, необходимой для профессиональной деятельности	анализировать исходную документацию в рамках управления работами по сопровождению и проектами создания (модификации) ИС	навыками ведения документооборота в организациях
Стандартный (в дополнение к пороговому)	основы теории систем и системного анализа, источники информации, необходимой для профессиональной деятельности	анализировать исходную документацию в рамках управления работами по сопровождению и проектами создания (модификации) ИС, использовать инструменты и методы определения финансовых и производственных показателей деятельности организаций	навыками анализа исходной документацию в рамках управления работами по сопровождению и проектами создания (модификации) ИС, ведения документооборота в организациях
Повышенный (в дополнение к пороговому, стандартному)	основы теории систем и системного анализа, источники информации, необходимой для профессиональной деятельности при управлении работами по сопровождению и проектами создания (модификации) ИС	разрабатывать и анализировать исходную документацию в рамках управления работами по сопровождению и проектами создания (модификации) ИС, использовать и осваивать новые инструменты и методы определения финансовых и производственных показателей деятельности организаций	навыками разработки и анализа исходной документации в рамках управления работами по сопровождению и проектами создания (модификации) ИС, ведения документооборота в организациях, инструментами и методами моделирования бизнес-процессов в ИС

6.3. Паспорт оценочных материалов

№ п/п	Наименование темы (раздела) дисциплины	Контролируемые планируемые результаты обучения в соотношении с результатами обучения по программе	Вид контроля/используемые оценочные средства	
			Текущий	Промежуточный
1.	Математические основы машинного обучения. Задачи классификации	ПК-2.1, ПК-2.2, ПК-2.3	Устный/письменный опрос Практические	Зачёт

			задачи Тестирование	
2.	Деревья в машинном обучении. Случайный лес. Линейные классификаторы: основные понятия	ПК-2.1, ПК-2.2, ПК-2.3	Устный/письменный опрос Практические задачи Тестирование	Зачёт

6.4.Оценочные материалы для текущего контроля

Примерная тематика докладов

Раздел дисциплины	Темы

Вопросы для устного/письменного опроса

Раздел дисциплины	Вопросы
Математические основы машинного обучения. Задачи классификации	<ol style="list-style-type: none"> 1. Постановка задач обучения по прецедентам. Объекты и признаки. Типы шкал: бинарные, номинальные, порядковые, количественные. 2. Типы задач: классификация, регрессия, прогнозирование, ранжирование. 3. Основные понятия: модель алгоритмов, метод обучения, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль. 4. Линейные модели регрессии и классификации. Метод наименьших квадратов. Полиномиальная регрессия. 5. Примеры прикладных задач. 6. Методика экспериментального исследования и сравнения алгоритмов на модельных и реальных данных. 7. Конкурсы по анализу данных kaggle.com. Полигон алгоритмов классификации. 8. CRISP-DM — межотраслевой стандарт ведения проектов интеллектуального анализа данных. 9. Линейные модели регрессии и классификации. Метод наименьших квадратов. Полиномиальная регрессия. 10. Примеры прикладных задач. 11. Методика экспериментального исследования и сравнения алгоритмов на модельных и реальных данных. 12. Конкурсы по анализу данных kaggle.com. Полигон алгоритмов классификации.
Деревья в машинном обучении. Случайный лес. Линейные классификаторы: основные понятия	<ol style="list-style-type: none"> 1. Решающее дерево. Жадная нисходящая стратегия «разделяй и властвуй». Алгоритм ID3. Недостатки жадной стратегии и способы их устранения. Проблема переобучения. 2. Вывод критериев ветвления. Мера нечистоты (impurity) распределения. Энтропийный критерий, критерий Джини. 3. Редукция решающих деревьев: предредукция и постредукция. Алгоритм C4.5. 4. Деревья регрессии. Алгоритм CART. 5. Небрежные решающие деревья (oblivious decision tree) 6. Решающее дерево. Жадная нисходящая стратегия «разделяй и властвуй». Алгоритм ID3. Недостатки жадной стратегии и способы их устранения. Проблема переобучения. 7. Вывод критериев ветвления. Мера нечистоты (impurity) распределения. Энтропийный критерий, критерий Джини.

- | | |
|--|--|
| | <ol style="list-style-type: none">8. Деревья регрессии. Алгоритм CART.9. Решающий лес. Случайный лес (Random Forest).10. Решающий пень. Бинаризация признаков. Алгоритм разбиения области значений признака на информативные зоны.11. Решающий список. Жадный алгоритм синтеза списка.
Преобразование решающего дерева в решающий список. |
|--|--|

Задания для тестирования по дисциплине для оценки сформированности компетенций (min 20, max 50 + ссылку на ЭИОС с тестами)

укажите задания

1. Что, из нижеперечисленного, относится к обучающей выборке?
 - 1) классификация данных;
 - 2) объекты с известными ответами;
 - 3) алгоритм, решающий функцию;
 - 4) множество предикатов.

2. Какие алгоритмы лучше работают на больших обучающих выборках?
 - 1) Бустинг;
 - 2) Беггинг;
 - 3) Реслинг;
 - 4) RSM.

3. С чем, из нижеперечисленного сравнивают линейный классификатор?
 - 1) с аксоном;
 - 2) с правилом Хебба;
 - 3) с генетическим алгоритмом;
 - 4) с нейроном.

4. Как называется алгоритм, который добавляет к набору G по одному признаку, каждый раз выбирая тот признак, который приводит к наибольшему уменьшению внешнего критерия?
 - 1) Add;
 - 2) Bootstrap;
 - 3) FullSearch;
 - 4) Поиск в ширину.

5. Что поступает на вход рекурсивного алгоритма синтеза бинарного решающего дерева ID3?
 - 1) обучающая выборка;
 - 2) корневая вершина дерева;
 - 3) максимальный ранг конъюнкций;
 - 4) ограничение на число конъюнкций.

6. Что называют обучением с подкреплением?
 - 1) обучение основанное на собственном опыте;
 - 2) обучение с n-количеством учителей;
 - 3) обучение с репетитором;
 - 4) обучение к контролирующими параметрами.

7. Выберите неверные утверждения:
 - 1) Наивный байесовский классификатор может быть только параметрическим.
 - 2) Наивный байесовский классификатор будет оптимальным, если признаки действительно независимы.
 - 3) При классификации объекта заодно оцениваются априорные вероятности его принадлежности каждому из классов.
 - 4) Наивный байесовский классификатор может быть только непараметрическим.

8. В каком методе базовые алгоритмы обучаются на различных подмножествах признакового

описания, которые выделяются случайным образом?

- 1) RSM;
- 2) Беггинг;
- 3) BGT;
- 4) CSEL.

9. Какие задачи из нижеперечисленных относятся к задачам классификации?

- 1) определение наиболее целесообразного способа лечения;
- 2) определение длительности и исхода заболевания;
- 3) оценивание кредитоспособности заёмщика;
- 4) задачи поискового вывода

10. К какому методу обучения относится метод главных компонент?

- 1) обучение с учителем;
- 2) обучение без учителя;
- 3) обучение по Хеббу;
- 4) обучение функционала.

11. Что означает, если веса w_{jh} между входными и скрытым слоем будут обнулены?

- 1) диагональные элементы доминируют в гессиане;
- 2) после стабилизации функционала ошибки Q вектор весов w находится в локальном минимуме;
- 3) информация, накопленная в сети, является полезной и не теряется при добавлении новых нейронов;
- 4) h -ый нейрон скрытого слоя не будет учитывать j -й признак.

12. Какой тип экспериментального исследования имеет цель - понимание, на что влияют параметры метода обучения?

- 1) исследование задач ранжирования
- 2) исследование задач классификации
- 3) исследование на модельных данных
- 4) исследование на сходимость

13. Какой алгоритм использует только операцию добавления термов?

- 1) жадный алгоритм синтеза конъюнкции;
- 2) стохастический локальный поиск;
- 3) процедура стабилизации;
- 4) процедура редукции.

14. Самоорганизующиеся карты Кохонена применяются для:

- 1) визуализации многомерных данных;
- 2) обучения задач классификации;
- 3) выявления ошибки обучения;
- 4) исследование на сходимость.

15. Какие алгоритмы лучше работают на коротких обучающих выборках?

- 1) Бустинг;
- 2) Беггинг;
- 3) RSM;
- 4) Поиск в глубину.

16. Как называется алгоритм, который осуществляет полный перебор всевозможных наборов признаков G в порядке возрастания сложности?

- 1) Add;
- 2) Bootstrap;
- 3) FullSearch;
- 4) Поиск в ширину.

17. Выберите правильный ответ. Задача регрессии - это:

- 1) множество объектов, разделенных на классы;
- 2) исследование влияние одного или нескольких признаков на объект;
- 3) определение порядка признака согласно рангу;
- 4) определение сонаправленности действия.

18. Какие слои в нейронной сети называются скрытыми?

- 1) все, кроме последнего;
- 2) только, те слои, которые находятся между 1-ым и 2-ым слоем;
- 3) только, первый и последний;
- 4) все, кроме первого.

19. Какая, из перечисленных ниже функций, соответствует алгоритму бустинга Ada Boost?

- 1) квадратичная;
- 2) сигмоидная;
- 3) кусочно-линейная;
- 4) экспоненциальная.

20. Что называют в теории нейронных сетей сокращением весов?

- 1) квадратичную регуляризацию;
- 2) нормализацию признаков;
- 3) стохастический градиент;
- 4) распределение Лапласа.

21. Что объясняет эффективность бустинга?

- 1) по мере добавления базовых алгоритмов увеличиваются отступы обучающих объектов $M_i = y_i \alpha(x_i)$;
- 2) выбирается тот класс, в котором осталось больше непокрытых объектов;
- 3) увеличение l_1 повышает качество базовых алгоритмов;
- 4) композиции можно периодически возвращаться к ранее построенным алгоритмам и обучать их заново, что приводит к улучшению.

22. Что из нижеперечисленного относится к недостаткам алгоритма AdaBoost?

- 1) По мере увеличения числа базовых алгоритмов обобщающая способность может улучшаться;
- 2) Склонен к переобучению при наличии значительного уровня шума в данных;
- 3) Требуется достаточно длинных обучающих выборок;
- 4) Простота реализации;
- 5) Возможность идентифицировать объекты, являющиеся шумовыми выбросами;
- 6) Бустинг может приводить к построению громоздких композиций, состоящих из сотен алгоритмов

23. Какой входной набор данных характерен для жадного алгоритма построения решающего списка?

- 1) Максимальное допустимое число отказов.
- 2) Максимальная допустимая доля ошибок на обучающей выборке.
- 3) Минимальная допустимая информативность правил в списке.
- 4) Множество элементарных предикатов.
- 5) Параметр критерия останова.
- 6) Максимальное число итераций.

24. Что из нижеперечисленного является достоинством алгоритма AdaBoost?

- 1) По мере увеличения числа базовых алгоритмов обобщающая способность может улучшаться;
- 2) Склонен к переобучению при наличии значительного уровня шума в данных;
- 3) Требуется достаточно длинных обучающих выборок;
- 4) Простота реализации;
- 5) Возможность идентифицировать объекты, являющиеся шумовыми выбросами;
- 6) Бустинг может приводить к построению громоздких композиций, состоящих из сотен алгоритмов

25. Какие параметры участвуют в алгоритме Беггинга?

- 1) максимальное число поколений;
- 2) порог качества базовых алгоритмов на контроле;
- 3) размер основной популяции;
- 4) порог качества базовых алгоритмов на обучении;
- 5) размер промежуточной популяции;
- 6) длина признакового подописания;
- 7) размер элиты, переходящей в следующее поколение без изменений;
- 8) длина обучающих подвыборок.

Практические задачи (min 20, max 50 + ссылку на ЭИОС с электронным изданием, если имеется)

Раздел дисциплины	Задачи
Математические основы машинного обучения. Задачи классификации	<ol style="list-style-type: none">1. Банку необходимо сделать пять рекламных кампаний, чтобы привлечь новых клиентов именно в свой банк. Для этого, на основании статистических данных о кредитных рисках и портретах клиентов немецких банков, размещённых в свободном доступе на странице Естественно-научного колледжа Эберли (https://newonlinecourses.science.psu.edu/stat508/resource/analysis/gcd), нужно составить пять портретов, определяющих целевую аудиторию каждого рекламного продукта. Разбейте клиентов на пять кластеров, а затем дайте описание типичного представителя каждого кластера – его центра.2. На основе данные о пассажирах «Титаника» (https://www.kaggle.com/prkukunoor/TitanicDataset). решите задачу классификации. По различным характеристикам пассажиров найдите у выживших пассажиров два наиболее важных признака (из четырех рассматриваемых: пол, класс, возраст, цена билета).3. Примените метод SVM (например, из библиотеки sklearn) для датасета blobs2. Визуализируйте результат (разбиение плоскости и опорные вектора) при разных вариантах ядер (линейное; полиномиальное степеней 2,3,5; RBF)4. Учет пропусков. В реальных задачах данные могут содержать пропуски. Это может быть вызвано тем, что клиенты не заполняли все поля в анкете или аккаунте, или тем, что не все параметры были оцифрованы за все время работы системы. Таким образом, встает вопрос о том, как интерпретировать пропуски.5. Кодирование нечисловых признаков. Очевидно, что далеко не все признаки объектов естественно описываются численным значением. Если говорить о размерах объекта или стоимости какого-то товара, то эти признаки, несомненно, будут числовыми. Если же речь идет о цвете, типе товара (категории) или вообще о текстовом описании некоторого объекта, то подобные признаки, как правило, поступают неоцифрованными. Нечисловые признаки с неупорядоченными значениями (в которых между значениями не определена дистанция, то есть нельзя сказать, что больше или меньше) называют категориальными, или номинальными. Типичный подход к их обработке – кодирование категориального признака с m возможными значениями с помощью m бинарных признаков.6. Приведение данных к единому масштабу и стандартизация «Сырые» данные имеют разный масштаб и разное распределение по каждому признаку. Например, какой-то химический показатель смеси может иметь значения в диапазоне от 0.0001 до 0.2, а другой показатель от -100 до 100. Или, скажем, возраст клиентов может быть от 16 до 40, причем гораздо

	<p>больше клиентов имеет возраст от 18 до 25, иными словами, математическое ожидание смещено относительно центра распределения. Подобные различия в признаках могут вносить существенную ошибку для множества моделей (например, для регрессии, нейронных сетей), и потому требуется привести все признаки к единому виду.</p> <p>7. Разметка данных. . Если речь идет о том, чтобы обучить модель прогнозировать будущие показатели по прошлому опыту (например, прогноз средней выручки по количеству посетителей магазинов), то такие данные, как правило, приходят с известной целевой переменной Y (то есть средней выручкой). Необходимо будет самостоятельно размечать выборку.</p> <p>8. Использование библиотеки Scikit-learn для классификации линейно разделимых объектов.</p> <p>9. Классификация линейно разделимых объектов (три признака, четыре класса) с использованием библиотеки Scikit-learn.</p> <p>10. Классификация на основе модели логистической регрессии с использованием библиотеки Scikit-learn.</p> <p>11. Классификация методом опорных векторов с использованием библиотеки Scikit-learn.</p>
<p>Деревья в машинном обучении. Случайный лес. Линейные классификаторы: основные понятия</p>	<ol style="list-style-type: none"> 1. Реализуйте алгоритмы построения дерева с критерием информационного выигрыша и критерием Джини и определению класса по мажоритарному классу в листе. Найдите оптимальную глубину дерева в обоих случаях (в отрезке 2-10) 2. Реализуйте алгоритм логистической регрессии со стохастическим градиентным спуском, обучите его на датасете <code>srambase_old (train)</code> и проверьте на датасете <code>srambase_new (val)</code>. Получите ROC кривые для вариантов без нормировки и с нормировкой признаков 3. Реализуйте алгоритм линейной регрессии, и полиномиальной регрессии (для датасета <code>noisysine</code> – степеней от 2 до 5, для датасета <code>hydrodynamics</code> – степени 2) без регуляризации. 4. Найдите максимум функции с помощью алгоритма кросс-энтропийного поиска, изображая распределение на каждом шаге. 5. Найдите лучший путь в задаче коммивояжера с помощью алгоритма отжига. 6. Реализация алгоритмов машинного обучения на Python. Модули NumPy, Pandas, Matplotlib, Scikit-Learn. 7. Модель логистической регрессии. Тренировка логистической регрессионной модели в Scikit-learn. 8. Разработка модели линейного адаптивного нейрона (ADALINE) на основе базового класса перцептрона 9. Настройка машинно-обучаемых моделей методом сеточного поиска

Тематика контрольных работ

Раздел дисциплины	Темы

6.5. Оценочные материалы для промежуточной аттестации

Фонд вопросов для проведения промежуточного контроля в форме зачета

Раздел дисциплины	Вопросы
<p>Математические основы машинного обучения. Задачи</p>	<ol style="list-style-type: none"> 1. Постановка задач обучения по прецедентам. Объекты и признаки. Типы шкал: бинарные, номинальные, порядковые, количественные. 2. Типы задач: классификация, регрессия, прогнозирование,

классификации	<p>ранжирование.</p> <p>3. Основные понятия: модель алгоритмов, метод обучения, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль.</p> <p>4. Линейные модели регрессии и классификации. Метод наименьших квадратов. Полиномиальная регрессия.</p> <p>5. Примеры прикладных задач.</p> <p>6. Методика экспериментального исследования и сравнения алгоритмов на модельных и реальных данных.</p> <p>7. Конкурсы по анализу данных kaggle.com. Полигон алгоритмов классификации.</p> <p>8. CRISP-DM — межотраслевой стандарт ведения проектов интеллектуального анализа данных.</p> <p>9. Линейные модели регрессии и классификации. Метод наименьших квадратов. Полиномиальная регрессия.</p> <p>10. Примеры прикладных задач.</p> <p>11. Методика экспериментального исследования и сравнения алгоритмов на модельных и реальных данных.</p> <p>12. Конкурсы по анализу данных kaggle.com. Полигон алгоритмов классификации.</p>
Деревья в машинном обучении. Случайный лес. Линейные классификаторы: основные понятия	<p>1. Решающее дерево. Жадная нисходящая стратегия «разделяй и властвуй». Алгоритм ID3. Недостатки жадной стратегии и способы их устранения. Проблема переобучения.</p> <p>2. Вывод критериев ветвления. Мера нечистоты (impurity) распределения. Энтропийный критерий, критерий Джини.</p> <p>3. Редукция решающих деревьев: предредукция и постредукция. Алгоритм C4.5.</p> <p>4. Деревья регрессии. Алгоритм CART.</p> <p>5. Небрежные решающие деревья (oblivious decision tree)</p> <p>6. Решающее дерево. Жадная нисходящая стратегия «разделяй и властвуй». Алгоритм ID3. Недостатки жадной стратегии и способы их устранения. Проблема переобучения.</p> <p>7. Вывод критериев ветвления. Мера нечистоты (impurity) распределения. Энтропийный критерий, критерий Джини.</p> <p>8. Деревья регрессии. Алгоритм CART.</p> <p>9. Решающий лес. Случайный лес (Random Forest).</p> <p>10. Решающий пень. Бинаризация признаков. Алгоритм разбиения области значений признака на информативные зоны.</p> <p>11. Решающий список. Жадный алгоритм синтеза списка. Преобразование решающего дерева в решающий список.</p>

6.6. Шкалы и критерии оценивания по формам текущего контроля и промежуточной аттестации

Шкала и критерии оценивания

Оценка	Критерии оценивания для мероприятий контроля с применением 2-х балльной системы
«зачтено»	ПК-2
«не зачтено»	Результаты обучения не сформированы на пороговом уровне