

МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное
образовательное учреждение высшего образования
«Череповецкий государственный университет»
Институт информационных технологий
Кафедра Математического и программного обеспечения ЭВМ

Министерство сельского хозяйства Российской Федерации
федеральное государственное бюджетное образовательное учреждение
высшего образования «Вологодская государственная
молочнохозяйственная академия имени Н.В. Верещагина»

Инженерный факультет

Кафедра энергетических средств и технического сервиса

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

МАШИННОЕ ОБУЧЕНИЕ В АГРОИНЖЕНЕРИИ

Направление подготовки: 35.03.06 Агроинженерия

Профили подготовки: Искусственный интеллект

Квалификация (степень) выпускника: Бакалавр

Череповец
2023

Программа составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 35.03.06 «Агроинженерия»

Разработчик: канд. техн. наук, доцент Шабалов Виктор Александрович

Программа одобрена на заседании кафедры энергетических средств и технического сервиса Вологодской государственной молочно-хозяйственной академии имени Н.В. Верещагина 20 июня 2023 года, протокол №10.

Зав. кафедрой: канд. техн. наук, доцент Бирюков А.Л.

Рабочая программа дисциплины согласована на заседании методической комиссии инженерного факультета Вологодской государственной молочно-хозяйственной академии имени Н.В. Верещагина 22 июня 2023 года, протокол №10.

Председатель методической комиссии: канд. техн. наук, доцент Берденников Е.А.

Рабочая программа дисциплины рассмотрена и одобрена на заседании кафедры Математического и программного обеспечения ЭВМ Института информационных технологий Череповецкого государственного университета 25 сентября 2023, протокол № 2.

Зав. кафедрой: доктор техн. наук, профессор Ершов Е.В.

Рабочая программа дисциплины рассмотрена и утверждена Ученым советом Института информационных технологий Череповецкого государственного университета 26 сентября, протокол № 2.

Директор института: доктор техн. наук, профессор Ершов Е.В.

1 Цель и задачи учебной дисциплины

Цель – формирование компетенций эффективного применения методов машинного обучения для решения агротехнических задач.

Задачи дисциплины:

- изучение и практическое освоение технологий исследования данных, их визуализации и интерпретации;
- изучение и практическое освоение методов и программных средств классификации, числового прогнозирования и кластеризации;
- изучение и практическое освоение методов и программных средств обнаружения закономерностей на основе ассоциативных правил и понижения размерности данных;
- получение навыков программирования систем искусственного интеллекта.

2 Место дисциплины в структуре ООП

Дисциплина «Машинное обучение в агроинженерии» входит в модуль Б1.О.07 «Искусственный интеллект» и относится к обязательной части Блока 1. Дисциплины (модули) основной образовательной программы высшего образования (ООП ВО) по направлению подготовки 35.03.06 «Агроинженерия». Индекс по учебному плану – Б1.О.07.05.

Дисциплина «Технологии сбора и подготовки данных в агроинженерии» изучается в 4 семестре на основе знаний, полученных при изучении дисциплин «Введение в теорию искусственного интеллекта», «Технологии сбора и подготовки данных в агроинженерии», «Математика». Результаты изучения дисциплины востребованы в ходе изучения последующих дисциплин модуля «Искусственный интеллект», в ходе учебной и производственной практики, а также при подготовке выпускной квалификационной работы. Форма промежуточной аттестации – экзамен.

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

Дисциплина направлена на формирование следующих компетенций выпускника:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции
ПК-14. Способен использовать системы искусственного интеллекта в решении задач анализа, прогнозирования, планирования, синтеза и принятия решений	ИПК-14.1. Выбирает методы решения задач с использованием систем искусственного интеллекта ИПК-14.2. Решает задачи с использованием систем искусственного интеллекта
ПК-15. Способен разрабатывать и применять методы машинного обучения для решения задач	ИПК-15.1. Проводит анализ требований и определяет необходимые классы задач машинного обучения ИПК-15.2. Определяет метрики оценки результатов моделирования и критерии качества построенных моделей ИПК-15.3. Принимает участие в оценке и выборе

Код и наименование компетенции	Код и наименование индикатора достижения компетенции
	используемых методов машинного обучения

4 Структура и содержание дисциплины

Общий объем дисциплины составляет 4 зачетные единицы.

4.1 Структура дисциплины

Вид учебной нагрузки	Всего часов (очная форма)
	2 семестр
Аудиторные занятия (всего)	48
<i>В том числе:</i>	
Лекции	20
Лабораторные работы	28
Практические занятия	–
Самостоятельная работа (всего)	87
Вид промежуточной аттестации	экзамен
Контроль	9
Общая трудоёмкость, часы	144
Зачётные единицы	4

4.2 Содержание разделов дисциплины

Раздел 1. Введение в машинное обучение

Понятия «наука о данных», «машинное обучение» (далее англ. machine learning, *ML*), «интеллектуальный анализ данных». Составляющие *ML*: хранение данных: абстрагирование: обобщение: оценка. Этапы решения задач с использованием *ML*: сбор данных: исследование и подготовка данных: обучение модели: оценка модели: улучшение модели. Типы входных данных. Типы алгоритмов машинного обучения. Подбор алгоритмов по входным данным. Библиотеки *Python* для машинного обучения.

Раздел 2. Исследование данных, их визуализация и интерпретация

Преобразование данных, построение выводов по данным и оценка результатов. Структуры данных. Числовые переменные. Измерение средних значений: среднее арифметическое и медиана. Измерение разброса: квартили и пятичисловая сводка. Визуализация числовых переменных: диаграммы размаха; гистограммы (разбиения по интервалам и плотность). Интерпретация числовых данных: равномерное и нормальное распределение. Измерение разброса: дисперсия и стандартное отклонение. Категориальные переменные. Мода. Взаимосвязи между переменными. Визуализация отношений: диаграммы разброса. Исследование взаимосвязей: перекрестные таблицы.

Раздел 3. Методы классификации

Классификация с использованием метода ближайших соседей: что такое классификация методом ближайших соседей; алгоритм *k-NN*; измерение степени сходства с помощью расстояния; выбор подходящего *k* подготовка данных для использования в алгоритме *k-NN*; почему алгоритм *k-NN* называют ленивым. Вероятностное обучение, классификация с использованием наивного байесовского классификатора: наивный байесовский классификатор; основные понятия байесовских методов; наивный байесовский алгоритм: классификация по наивному байесовскому алгоритму; критерий Лапласа; использование числовых признаков в наивном байесовском алгоритме.

Классификация с использованием деревьев решений и правил: деревья решений; выбор лучшего разделения; сокращение дерева решений. Случайные леса. Градиентный бустинг.

Раздел 4. Методы числового прогнозирования

Прогнозирование числовых данных, регрессионные методы: понятие регрессии; простая линейная регрессия; оценка методом наименьших квадратов; корреляции; множественная линейная регрессия.

Раздел 5. Обнаружение закономерностей на основе ассоциативных правил

Ассоциативные правила. Типы задач, решаемых с использованием ассоциативных правил. Алгоритм Apriori для поиска ассоциативных правил, преимущества и недостатки алгоритма. Измерение интересности правила: поддержка и доверие. Построение набора правил по принципу Apriori. Выявление часто покупаемых продуктов в соответствии с ассоциативными правилами.

Раздел 6. Методы кластеризации

Что такое кластеризация. Кластеризация как задача машинного обучения. Алгоритм кластеризации методом k-средних: преимущества и недостатки метода; использование расстояния для разбиения на кластеры и внесения изменений; выбор количества кластеров. Сегментация рынка для подростков с использованием кластеризации методом k-средних.

Раздел 7. Методы понижения размерности данных

Для чего понижают размерность данных. Метод главных компонент, новая система координат, достоинства и ограничения метода. Использование метода главных компонент для понижения размерности данных в агроинженерных задачах машинного обучения.

4.3 Разделы дисциплины и вид занятий

№ п/п	Раздел дисциплины	Лекции	ЛЗ	СРС	Всего
1.	Раздел 1. Введение в машинное обучение	2	4	9	15
2.	Раздел 2. Исследование данных, их визуализация и интерпретация	4	4	12	20
3.	Раздел 3. Методы классификации	4	4	12	20
4.	Раздел 4. Методы числового прогнозирования	2	4	14	20
5.	Раздел 5. Обнаружение закономерностей на основе ассоциативных правил	2	4	14	20
6.	Раздел 6. Методы кластеризации	4	4	12	20
7.	Раздел 7. Методы понижения размерности данных	2	4	14	20
Всего		20	28	87	113

Примечание: ЛЗ – лабораторные занятия; СРС – самостоятельная работа студента

5 Матрица формирования компетенций по дисциплине

№ п/п	Разделы, темы дисциплины	Компетенции		Общее количество компетенций
		ПК-3	ПК-4	
1.	Введение в машинное обучение			

№ п/п	Разделы, темы дисциплины	Компетенции		Общее количество компетенций
		ПК-3	ПК-4	
2..	Исследование данных, их визуализация и интерпретация	+	+	2
3.	Методы классификации	+	+	2
4.	Методы числового прогнозирования	+	+	2
5.	Обнаружение закономерностей на основе ассоциативных правил	+	+	2
6.	Методы кластеризации	+	+	2
7.	Методы понижения размерности данных	+	+	2

6 Образовательные технологии

Объем аудиторных занятий всего – 48 часа, в том числе лекции – 20 часов, лабораторные работы – 28 часов.

Удельный вес занятий, проводимых в интерактивной форме, 100% от объема аудиторных занятий.

Семестр	Вид занятия	Наименование темы	Используемые интерактивные образовательные технологии	Количество часов
4	Лекция	Введение в машинное обучение	Лекция-визуализация	2
4	Лекция	Исследование данных, их визуализация и интерпретация	Лекция-визуализация	4
4	Лекция	Методы классификации	Лекция-визуализация	4
4	Лекция	Методы числового прогнозирования	Лекция-визуализация	2
4	Лекция	Обнаружение закономерностей на основе ассоциативных правил	Лекция-визуализация	2
4	Лекция	Методы кластеризации	Лекция-визуализация	4
4	Лекция	Методы понижения размерности данных	Лекция-визуализация	2
4	ЛЗ	Лабораторный практикум по дисциплине «Машинное обучение в агроинженерии»	Лабораторная работа	28
Итого:				48

7 Учебно-методическое обеспечение самостоятельной работы студентов. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины

7.1 Перечень оценочных средств

Компетенции	Индикаторы достижения компетенций	Оценочные средства
ПК-14. Способен использовать системы искусственного интеллекта в решении задач анализа, прогнозирования, планирования, синтеза и принятия решений	ИПК-14.1. Выбирает методы решения задач с использованием систем искусственного интеллекта ИПК-15.2. Решает задачи с использованием систем искусственного интеллекта	1. Задания для выполнения лабораторных работ. 2. Задания для самостоятельной работы. 3. Вопросы к зачету.

Компетенции	Индикаторы достижения компетенций	Оценочные средства
ПК-15. Способен разрабатывать и применять методы машинного обучения для решения задач	ИПК-15.1. Проводит анализ требований и определяет необходимые классы задач машинного обучения ИПК-15.2. Определяет метрики оценки результатов моделирования и критерии качества построенных моделей ИПК-15.3. Принимает участие в оценке и выборе используемых методов машинного обучения	1. Задания для выполнения лабораторных работ. 2. Задания для самостоятельной работы. 3. Вопросы к зачету.

7.2 Типовые контрольные задания и методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Образцы заданий для самостоятельной работы:

По итогам самостоятельной работы студент готовит отчет, включающий в себя ответы на вопросы и решение заданий, предполагавшихся к выполнению в ходе самостоятельной работы. Отчет сдается преподавателю в электронной форме.

Задания для самостоятельной работы по разделу «Введение в машинное обучение»:

1. Как Вы считаете, чем машинное обучение отличается от интеллектуального анализа данных (если эти понятия отличаются друг от друга)?
2. Приведите примеры использования методов машинного обучения.
3. Подготовьте интеллект-карту, включающую в себя представление составляющих машинного обучения: хранение данных: абстрагирование: обобщение: оценка.
4. Приведите описание этапов решения задач с использованием машинного обучения: сбор данных: исследование и подготовка данных: обучение модели: оценка модели: улучшение модели.
5. Дайте описание типов входных данных, используемых при решении задач с помощью методов машинного обучения.
6. Перечислите типы алгоритмов машинного обучения.
7. Как подбирается метод машинного обучения для решения конкретной прикладной задачи? Что влияет на выбор метода?
8. Каково назначение и возможности библиотек библиотеки Python для машинного обучения (дайте заключение на основе анализа документации разработчиков библиотек).
9. Приведите описание критериев эффективности и качества функционирования системы искусственного интеллекта: точность, релевантность, достоверность, целостность, быстрота решения задач, надежность, защищенность функционирования систем искусственного интеллекта.

10. Приведите описание методов постановки задач, проведения и анализа тестовых и экспериментальных испытаний работоспособности систем искусственного интеллекта, в том числе систем машинного обучения.
11. Перечислите методы и критерии оценки качества моделей машинного обучения.
12. Приведите содержание унифицированных и обновляемых методологии описания, сбора и разметки данных, а также механизмов контроля за соблюдением указанных методологий.

Задания для самостоятельной работы по разделу «Исследование данных, их визуализация и интерпретация»:

1. Для каких целей выполняется интерпретация данных?
2. Что такое структура данных?
3. Какие базовые наборы изменений обычно применяются в числовым данным?
4. Почему в ходе исследования данных запрашивают как средние, так и медианные значения числовых переменных?
5. Что такое «пятичисловая сводка»? Для каких целей она используется?
6. Что отображает диаграмма размаха?
7. Что отображает гистограмма?
8. Как выглядит гистограмма равномерного распределения?
9. Как выглядит кривая нормального распределения?
10. Что измеряется стандартным отклонением?
11. Что гласит правило «68-95-99.7»?
12. Что отображает таблица частотности?
13. Для каких целей строится диаграмма разброса?
14. Что показывают перекрестные таблицы (кросс-таблицы, таблицы сопряженности)?

Задания для самостоятельной работы по разделу «Методы классификации»:

1. В чем заключается суть метода k -NN?
2. Приведите примеры задач, решаемых с использованием метода k -NN.
3. Каковы преимущества метода k -NN?
4. Каковы недостатки метода k -NN?
5. Как измеряется степень сходства между экземплярами набора данных?
6. Каким образом выбирается подходящее k ?
7. Что такое «минимаксная» нормализация?
8. Каким образом выполняется стандартизация по z -оценке?
9. Что такое «фиктивное» кодирование?
10. Почему алгоритм k -NN называют ленивым?
11. Изучите документацию разработчиков библиотеки *Scikit-learn* (в части реализации метода k -NN).
12. Изучите пример использования метода k -NN для классификации данных.
13. Что такое «вероятностное обучение»?
14. В чем заключается суть работы наивного байесовского классификатора?
15. Приведите примеры задач, решаемых с использованием наивного байесовского классификатора.
16. Каковы преимущества наивного байесовского классификатора?
17. Каковы недостатки наивного байесовского классификатора?
18. Почему алгоритм называют наивным?
19. Изучите документацию разработчиков библиотеки *Scikit-learn* в части реализации наивного байесовского классификатора.

20. Изучите пример использования наивного байесовского алгоритма для классификации данных.
21. Для каких целей используются методы деревьев?
22. Почему группа методов получила такое название?
23. Приведите примеры задач, решаемых с использованием деревьев.
24. Что такое «рекурсивное сегментирование»?
25. Каким образом работает алгоритм дерева решений C5.0?
26. Каким образом выбирается лучшее разделение?
27. С какой целью выполняется «сокращение» дерева решений?
28. Изучите документацию разработчиков библиотеки в части реализации деревьев решений.
29. Изучите пример использования дерева решений для классификации данных.
30. В чем заключается суть работы со случайными лесами?
31. Как работает градиентный бустинг?

Задания для самостоятельной работы разделу дисциплины «Методы числового прогнозирования»:

1. Для каких целей используются методы регрессии?
2. Приведите понятие регрессии.
3. Приведите примеры задач, решаемых с использованием регрессии.
4. Как определяется простая линейная регрессия?
5. Приведите описание оценки методом наименьших квадратов.
6. Как рассчитывается коэффициент корреляции Пирсона?
7. Приведите описание множественной линейной регрессии. В чем заключаются преимущества и недостатки данного метода?
8. Изучите документацию разработчиков библиотеки *Scikit-learn* в части реализации линейной регрессии.
9. Изучите пример использования линейной регрессии для числового прогнозирования.

Задания для самостоятельной работы разделу дисциплины «Обнаружение закономерностей на основе ассоциативных правил»:

1. В чем заключается суть метода ассоциативных правил?
2. Какие задачи решаются с использованием данного метода?
3. К какому типу методов машинного обучения относится метод ассоциативных правил?
4. В чем заключается суть метода *Apriori*?
5. В каких библиотеках *Python* реализован метод ассоциативных правил?
6. Каким образом производится обучение модели?
7. Какие параметры необходимо указать для запуска обучения?
8. Как проверить эффективность модели?
9. Что необходимо сделать, чтобы повысить эффективность модели?
10. Как сохранить ассоциативные правила в файл или фрейм данных?

Задания для самостоятельной работы разделу дисциплины «Методы кластеризации»:

1. Что такое «кластеризация»? Чем кластеризация отличается от классификации?
2. Какие задачи решаются с использованием методов кластеризации?
3. Перечислите известные Вам методы кластеризации.

4. В чем заключаются суть метода k -средних?
5. Перечислите достоинства и недостатки метода k -средних
6. В каких библиотеках *Python* реализован метод k -средних?
7. Проанализируйте документацию разработчиков библиотек. Каким образом производится обучение модели? Какие параметры необходимо указать для запуска обучения? Как проверить эффективность модели?
8. Что необходимо сделать, чтобы повысить эффективность модели?
9. Изучите пример решения задачи с использованием метода k -средних (<https://coderlessons.com/tutorials/pytllion-teclmologies/iiznaite-inashiinoe-obuchenie-s-python/ml-algoritm-klasterizatsii-k-srednikh>).

Задания для самостоятельной работы разделу дисциплины «Методы понижения размерности данных»:

1. В чем заключается принцип работы алгоритма понижения размерности данных?
2. Какие задачи решаются с использованием данного алгоритма?
3. В каких библиотеках *Python* реализован данный алгоритм?
4. Проанализируйте информацию разработчиков средств визуализации.

Образцы заданий для лабораторных работ:

По итогам выполнения лабораторной работы студент демонстрирует результаты работы программы преподавателю, предварительно разработав тестовые случаи, а также сдает в электронном виде отчет, содержащий порядок выполнения работы.

Лабораторная работа «Программные инструменты, необходимые для выполнения лабораторных работ по машинному обучению»

Цель работы: установка и практическое изучение программных инструментов, необходимых для дальнейшего выполнения комплекса лабораторных работ по курсу «Машинное обучение в агроинженерии».

При использовании методов машинного обучения для решения различных практических задач, как правило, используют язык *Python* в интерактивной среде разработки (*Interactive Development Environment, IDE*), позволяющей быстро корректировать имеющуюся программу, и набор библиотек для расчетов и визуализации результатов. Для выполнения лабораторных работ понадобятся следующие инструменты:

- *Python* (версии 2.7 или 3.5, в зависимости от личных предпочтений);
- библиотеки языка *Python*:

NumPy – для работы с многомерными массивами данных, включая матрицы,

SciPy – для вычисления математических функций (интерполяция, интегрирование, дифференцирование, обработка сигналов и т.д.),

Pandas – для манипулирования числовыми таблицами и временными рядами,

Matplotlib – для визуализации данных,

Scikit-Learn – для использования методов машинного обучения (сокращения размерности, кластеризации, регрессии и т.д.).

Эти библиотеки можно установить с помощью *pip*-менеджера пакетов, написанных на *Python*, из терминала: *pip install numpy scipy pandas scikit-learn*;

- интерактивная среда разработки *Anaconda*. Можно также использовать среду *PyCharm*.

Лабораторная работа «Бинарная классификация объектов»

Цель работы: научиться вычислять коэффициенты разделяющей линии и величину отступа (margin) при бинарной классификации объектов.

Теория для выполнения лабораторной работы доступна на следующих страницах сайта: <https://proproprogs.ru/ml> в разделах:

Постановка задачи машинного обучения

Линейная модель. Понятие переобучения

Способы оценивания степени переобучения моделей

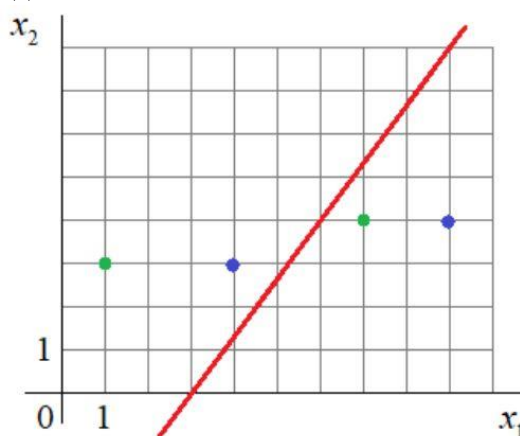
Уравнение гиперплоскости в задачах бинарной классификации

Решение простой задачи бинарной классификации

Задание на лабораторную работу (по вариантам):

1. Используя рисунок своего варианта, необходимо вычислить весовые коэффициенты для разделяющей линии.
2. Вычислить отступы.

Пример варианта задания



Варианты заданий приведены в учебном пособии: Методы машинного обучения: методические указания к лабораторным работам /сост. С. М. Наместников. – Ульяновск : УлГТУ, 2022. – 19 с.

Содержание отчета:

1. Титульный лист с названием лабораторной работы, номером варианта, фамилией студента и группы.
2. Расчеты для весов разделяющей линии.
3. Расчеты для отступов.
4. Выводы по полученным результатам.

Лабораторная работа «Обучение линейного алгоритма бинарной классификации образов с помощью градиентного алгоритма»

Цель работы: научиться реализовывать алгоритм градиентного спуска для задачи обучения линейной модели бинарной классификации образов.

Теория для выполнения лабораторной работы доступна на странице сайта: <https://proproprogs.ru/ml> в разделах:

1. Линейная модель. Понятие переобучения
2. Способы оценивания степени переобучения моделей
3. Уравнение гиперплоскости в задачах бинарной классификации
4. Решение простой задачи бинарной классификации
5. Функции потерь в задачах линейной бинарной классификации
6. Стохастический градиентный спуск SGD и алгоритм SAG

7. Пример использования SGD при бинарной классификации образов

Задание на лабораторную работу (по вариантам):

1. В файле `iris_data.py` даны обучающие выборки (по вариантам) для обучения линейного алгоритма бинарной классификации образов.
2. Выполнить обучение линейной модели (найти значения вектора весовых коэффициентов с помощью градиентного алгоритма (программы, написанной на языке Python), который должен минимизировать величину эмпирического риска.

Пример варианта задания

Функция потерь для реализации градиентного алгоритма $Q(M) = (1 - M)^2$ и её производная.

Содержание отчета

1. Титульный лист с названием лабораторной работы, номером своего варианта, фамилией студента и группы.
2. Математические выкладки, необходимые для реализации алгоритма обучения.
3. Текст программы обучения линейной модели с использованием градиентного алгоритма на языке Python.
4. Результаты работы программы в виде графика множества точек обучающей выборки (каждый класс точек должен быть представлен разными маркерами и цветами) и полученной разделяющей линии.
5. Выводы по полученным результатам.

Лабораторная работа «Исследование работы $L2$ -регуляризатора в задачах регрессии»

Цель работы: изучить особенности работы $L2$ -регуляризатора на примере задачи аппроксимации функции линейной моделью.

Теория для выполнения лабораторной работы доступна на следующих страницах сайта: <https://proproprogs.ru/ml> в разделах:

1. Функции потерь в задачах линейной бинарной классификации.
2. $L2$ -регуляризатор.
3. Математическое обоснование и пример работы.
4. $L1$ -регуляризатор.
5. Отличия между $L1$ - и $L2$ -регуляризаторами.
6. Вероятностный взгляд на $L1$ и $L2$ -регуляризаторы.

Задание на лабораторную работу (по вариантам):

1. Аппроксимировать (описать) функцию заданного варианта с помощью линейной модели.
2. Составить обучающую выборку следует из всех четных индексов сгенерированных значений функции.
3. Вычислить значения коэффициентов вектора весов для квадратической функции потерь (в задачах регрессии, обычно, используют именно такую функцию потерь), которые минимизируют эмпирический риск.
4. Вычислите прогнозы функции с помощью полученной модели для всего диапазона значений.
5. Вычислите коэффициенты вектора весов с $L2$ регуляризатором.
6. Для новой модели повторите вычисление прогнозов функции для всего диапазона значений.
7. Все программы реализовать на языке Python с использованием пакетов `NumPy` и `Matplotlib`.

Пример варианта задания: функция для исследования $L1$ и $L2$ -регуляризаторов
 $y(x) = \frac{1}{10 + x^2}$, $x \in [0; 10; 0,1]$.

Варианты заданий приведены в учебном пособии: Методы машинного обучения: методические указания к лабораторным работам /сост. С. М. Наместников. – Ульяновск : УлГТУ, 2022. – 19 с.

Содержание отчета

1. Титульный лист с названием лабораторной работы, номером своего
2. варианта, фамилией студента и группы.
3. Математические выкладки для реализации алгоритмов.
4. Тексты программ с результатами их работы.
5. Выводы по полученным результатам.

Лабораторная работа «Реализация наивного байесовского классификатора»

Цель работы: научиться строить наивный байесовский классификатор и с его помощью выполнять бинарную классификацию образов.

Теория для выполнения лабораторной работы доступна на странице сайта: <https://proproprogs.ru/ml> в разделах:

1. Логистическая регрессия. Вероятностный взгляд на машинное обучение
2. Вероятностный взгляд на $L1$ и $L2$ -регуляризаторы
3. Формула Байеса при решении конкретных задач
4. Байесовский вывод. Наивная байесовская классификация
5. Гауссовский байесовский классификатор
6. Линейный дискриминант Фишера

Задания на лабораторную работу (по вариантам)

1. Построить (реализовать на языке Python) наивный байесовский классификатор на основе, следующих данных обучающей выборки (для своего варианта).
2. Полагать, что признаки независимы и распределены по гауссовскому закону (нормальной плотности распределения вероятностей).
3. Для данной обучающей выборки подсчитать число и процент неверных классификаций.
4. Отобразить обучающую выборку в виде графика точек на плоскости (объекты разных классов должны быть иметь разные маркеры и цвет).

Содержание отчета

1. Титульный лист с названием лабораторной работы, номером своего варианта, фамилией студента и группы.
2. Расчеты, связанные с построением наивного байесовского классификатора.
3. Программа, реализующая наивный байесовский классификатор.
4. Графики и результаты работы программы.
5. Выводы по полученным результатам.

Лабораторная работа «Реализация алгоритма метода опорных векторов для задачи бинарной классификации»

Цель работы: реализовать метод опорных векторов (SVM) для задачи бинарной классификации.

Теория для выполнения лабораторной работы доступна на странице сайта: <https://proproprogs.ru/ml> в разделах:

1. Введение в метод опорных векторов (SVM)
2. Реализация метода опорных векторов (SVM)
3. Метод опорных векторов (SVM) с нелинейными ядрами

Задания на лабораторную работу (по вариантам)

1. Необходимо построить (реализовать на языке Python с применением пакета *Scikit-Learn*) линейный вариант метода опорных векторов, для данных обучающей выборки (для своего варианта).
2. Для данной обучающей выборки подсчитать число и процент неверных классификаций.
3. Отобразить обучающую выборку в виде графика точек на плоскости (объекты разных классов должны быть иметь разные маркеры и цвет), а также полученную (в результате обучения) разделяющую линию.

Содержание отчета

1. Титульный лист с названием лабораторной работы, номером своего варианта, фамилией студента и группы.
2. Программа, реализующая метод опорных векторов
3. Графики и результаты работы программы.
4. Выводы по полученным результатам.

Вопросы для экзамена

1. Понятия «наука о данных», «машинное обучение», «интеллектуальный анализ данных».
2. Составляющие машинного обучения: хранение данных: абстрагирование: обобщение: оценка.
3. Этапы решения задач с использованием машинного обучения:
4. Типы входных данных.
5. Типы алгоритмов машинного обучения.
6. Подбор алгоритмов по входным данным.
7. Библиотеки *Python* для машинного обучения.
8. Преобразование данных, построение выводов по данным и оценка результатов.
9. Структуры данных. Числовые переменные.
10. Измерение средних значений: среднее арифметическое и медиана.
11. Измерение разброса: квартили и пятичисловая сводка.
12. Визуализация числовых переменных: диаграммы размаха; гистограммы (разбиения по интервалам и плотность).
13. Интерпретация числовых данных: равномерное и нормальное распределение.
14. Измерение разброса: дисперсия и стандартное отклонение. Мода.
15. Взаимосвязи между переменными.
16. Визуализация отношений: диаграммы разброса.
17. Исследование взаимосвязей: перекрестные таблицы.
18. Классификация с использованием метода ближайших соседей.
19. Вероятностное обучение, классификация с использованием наивного байесовского классификатора.
20. Классификация с использованием деревьев решений и правил.
21. Прогнозирование числовых данных, регрессионные методы.
22. Ассоциативные правила. Типы задач, решаемых с использованием ассоциативных правил.
23. Алгоритм *Apriori* для поиска ассоциативных правил, преимущества и недостатки алгоритма.
24. Измерение интересности правила: поддержка и доверие.
25. Построение набора правил по принципу *Apriori*.
26. Кластеризация как задача машинного обучения.
27. Алгоритм кластеризации методом *k*-средних.
28. Понижение размерности данных.

29. Применение метода главных компонент для понижения размерности данных в агроинженерных задачах машинного обучения.

Уровни оценки компетенций:

- базовый 55-69 баллов,
- повышенный 70-100 баллов.

Преподаватель проводит систематический контроль знаний студентов, ориентируясь на перечень вопросов для проведения зачета.

Критерии оценивания ответа студента на зачете

Ответ на зачете оценивается исходя из 40 баллов (максимум). Билет содержит теоретический вопрос и практическое задание, преподаватель может задавать дополнительные вопросы. Полный ответ на основной вопрос оценивается максимум в 20 баллов, предполагает свободное изложение (не чтение) всего необходимого материала, ответы студента на уточняющие вопросы, если они есть. Правильный ответ на дополнительный вопрос оценивается максимум в 5 баллов. Правильное выполнение практического задания оценивается в 20 баллов.

Критерии оценки лабораторных работ и самостоятельной работы студента (от 0 до 10 баллов):

- **9-10 баллов** выставляется студенту, если работа выполнена самостоятельно и полностью верно; представлен отчет, содержащий результаты выполнения заданий работы и ответы на контрольные вопросы; студент анализирует результаты, полученные в ходе выполнения работы, делает выводы.
- **7-8 баллов** выставляется студенту, если работа выполнена самостоятельно, в целом правильно, но имеются некоторые неточности в выполнении заданий или ответах на контрольные вопросы; представлен отчет, содержащий результаты выполнения заданий и ответы на контрольные вопросы; студент анализирует результаты, полученные в ходе выполнения работы, делает выводы.
- **5-6 баллов** выставляется студенту, если работа выполнена самостоятельно, в целом правильно, но имеются некоторые неточности в выполнении заданий или ответах на контрольные вопросы; представлен отчет, содержащий результаты выполнения заданий лабораторной работы и ответы на контрольные вопросы; студент испытывает затруднения при проведении анализа результатов, полученных в ходе выполнения лабораторной работы, и формулировке выводов.
- **3-4 балла** выставляется студенту, если студент не до конца справился с заданием, не совсем верно ответил на контрольные вопросы, однако оформил отчет по результатам работы.
- **1-2 балла** выставляется студенту, если студент не до конца справился с заданием, не совсем верно ответил на контрольные вопросы, не оформил отчет по результатам работы.
- **0 баллов** выставляется студенту, если студент не справился с заданием, неверно ответил на представленные вопросы.

7.3 Шкала и критерии оценивания компетенций на различных этапах их формирования

Шкала оценивания компетенций:

Оценка в 100-балльной шкале	Оценка в 5-ти балльной шкале	Уровень сформированности компетенций
0-54 баллов	неудовлетворительно (не зачтено)	Недостаточный уровень
55-69 баллов	удовлетворительно (зачтено)	Пороговый уровень
70-85 баллов	хорошо (зачтено)	Продвинутый уровень
86-100 баллов	отлично (зачтено)	Высокий уровень

Критерии оценивания компетенций:

Индикаторы достижения компетенций	Критерии оценивания компетенций		
	Пороговый уровень	Продвинутый уровень	Высокий уровень
ПК-14. Способен использовать системы искусственного интеллекта в решении задач анализа, прогнозирования, планирования, синтеза и принятия решений	<p>Не знает основные методы решения задач с использованием систем искусственного интеллекта; классы задач, решаемых с использованием систем искусственного интеллекта, и критерии оценки получаемых результатов решения</p> <p>Не умеет выбирать методы решения задач с использованием систем искусственного интеллекта; применять методы обращения к целевым переменным и задания исходных данных и ограничений; оценивать полученные результаты решения задач с использованием систем искусственного интеллекта</p>	<p>Знает основные методы решения задач с использованием систем искусственного интеллекта; классы задач, решаемых с использованием систем искусственного интеллекта, и критерии оценки получаемых результатов решения</p> <p>Умеет выбирать методы решения задач с использованием систем искусственного интеллекта; применять методы обращения к целевым переменным и задания исходных данных и ограничений; оценивать полученные результаты решения задач с использованием систем искусственного интеллекта</p>	<p>Демонстрирует глубокое знание и понимание: основных методов решения задач с использованием систем искусственного интеллекта; классов задач, решаемых с использованием систем искусственного интеллекта, и критериев оценки получаемых результатов решения</p> <p>Полностью верно и самостоятельно умеет: выбирать методы решения задач с использованием систем искусственного интеллекта; применять методы обращения к целевым переменным и задания исходных данных и ограничений; оценивать полученные результаты решения задач с использованием систем искусственного интеллекта</p>
ПК-15. Способен разрабатывать и применять методы машинного обучения для решения задач	<p>Не знает принципы и методы машинного обучения, типы и классы задач машинного обучения; методы и критерии оценки качества моделей машинного обучения; классические методы и алгоритмы машинного обучения: предиктивные обучение с учителем, дескриптивные – обучение без учителя.</p> <p>Не умеет сопоставить задачам предметной</p>	<p>Знает принципы и методы машинного обучения, типы и классы задач машинного обучения; методы и критерии оценки качества моделей машинного обучения; классические методы и алгоритмы машинного обучения: предиктивные обучение с учителем, дескриптивные – обучение без учителя.</p> <p>Умеет сопоставить задачам предметной</p>	<p>Демонстрирует глубокое знание и понимание: принципов и методов машинного обучения, типы и классы задач машинного обучения; методы и критерии оценки качества моделей машинного обучения; классических методов и алгоритмов машинного обучения: предиктивные обучение с учителем, дескриптивные – обучение без учителя.</p>

Индикаторы достижения компетенций	Критерии оценивания компетенций		
	Пороговый уровень	Продвинутый уровень	Высокий уровень
	области классы задач машинного обучения; определять критерии и метрики оценки результатов моделирования при построении систем искусственного интеллекта в исследуемой области; проводить сравнительный анализ и осуществлять выбор методов и алгоритмов для решения задач машинного обучения.	области классы задач машинного обучения; определять критерии и метрики оценки результатов моделирования при построении систем искусственного интеллекта в исследуемой области; проводить сравнительный анализ и осуществлять выбор методов и алгоритмов для решения задач машинного обучения.	Полностью верно и самостоятельно умет: сопоставить задачам предметной области классы задач машинного обучения; определять критерии и метрики оценки результатов моделирования при построении систем искусственного интеллекта в исследуемой области; проводить сравнительный анализ и осуществлять выбор методов и алгоритмов для решения задач машинного обучения.

8 Учебно-методическое и информационное обеспечение дисциплины

8.1 Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

Основная литература:

1. Лимановская, О. В. Основы машинного обучения: учебное пособие / О. В. Лимановская, Т. И. Алферьева. - 2-е изд., стер. - Москва: ФЛИНТА Изд-во Урал. ун-та, 2022. - 88 с. - ISBN 978-5-9765-5006-3 (ФЛИНТА); - Текст: электронный. - URL: <https://znanium.com/catalog/product/189137>
2. Титов, А. Н. Python. Обработка данных: учебно-методическое пособие / А.Н. Титов, Р.Ф. Тагиева. – Казань: Изд-во КНИТУ, 2022. - 104 с. - ISBN 978-5-7882-3171-6. – Текст: электронный. – URL: <https://znanium.com/catalog/product/2069264> (дата обращения: 05.06.2023).

Дополнительная литература:

1. Коэльо, Луис Педро Построение систем машинного обучения на языке Python / Луис Педро Коэльо, Вилли Ричарт; пер. с англ. А. А. Слинкина. - 2-е изд. - Москва: ДМК Пресс, 2016. - 302 с. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1027824>
2. Никитина, Т. П. Программирование. Основы Python / Т. П. Никитина, Л. В. Королев. – Санкт-Петербург: Лань, 2023. – 156 с. – ISBN 978-5-507-45283-5. – Текст: электронный // Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/302714> (дата обращения: 05.06.2023).
3. Жуков, Р. А. Язык программирования Python: практикум: учебное пособие / Р.А. Жуков. – Москва: ИНФРА-М, 2023. – 216 с. + Доп. материалы [Электронный ресурс]. – (Высшее образование: Бакалавриат). – DOI 10.12737/textbook_5cb5ca35aaa7f5.89424805. - ISBN 978-5-16-016971-2. – Текст: электронный. – URL: <https://znanium.com/catalog/product/1915716>

8.2 Учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине (модулю)

Учебно-методические материалы для выполнения лабораторных работ.

8.3 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля), включая перечень информационных справочных систем (при необходимости)

1. Информационно-справочная система «КонсультантПлюс. – URL: <http://www.consultant.ru/>
2. Программирование на Python. URL: <https://stepik.org/course/67/syllabus> (дата доступа 04.06.2023).
3. Python – обучающий курс от Сергея Балакирева. URL: <https://stepik.org/course/100707/promo> (дата доступа 04.06.2023)
4. Основы статистики – URL: <https://stepik.org/course/76/syllabus> (дата доступа 04.06.2023)
5. Научная электронная библиотека eLIBRARY.RU – URL: <http://elibrary.ru>
6. ЭБС ЛАНЬ – URL: <https://e.lanbook.com/>,
7. ЭБС Znanium.com – URL: <https://znanium.com/>

8.4 Перечень информационных технологий, используемых при изучении дисциплины, включая программное обеспечение

1. ОС семейства Microsoft Windows.
2. MS Office 365.
3. Браузер.
4. Язык программирования Python – URL: <https://www.python.org/>
5. Среда программирования на языке Python, например, JupyterLab URL: <https://jupyter.org/>.
6. NumPy – пакет для научных вычислений с Python. – URL: <https://numpy.org/>
7. Программное обеспечение с открытым исходным кодом для математики, науки и техники – URL: <https://scipy.org/>
8. Библиотека на языке программирования Python для визуализации данных двумерной и трёхмерной графикой Matplotlib – URL: <https://matplotlib.org/>
9. Инструмент для анализа и обработки данных с открытым исходным кодом Pandas – URL: <https://pandas.pydata.org/>
10. Anaconda – платформа для быстрой разработки и развертывания безопасных решений Python – URL: <https://www.anaconda.com/>
11. Google Colab – URL: https://colab.research.google.com/#scrollTo=5fCEDCU_qrC0
12. API глубокого обучения Keras – URL: <https://keras.io/>
13. Комплексная платформа машинного обучения TensorFlow – URL: <https://www.tensorflow.org/>
14. Машинное обучение с открытым исходным кодом и визуализация данных Orange Data Mining – URL: <https://orangedatamining.com/>
15. Data Analytics Platform KNIME – URL: <https://www.knime.com/knime-analytics-platform>
16. Некоммерческий проект с открытым исходным кодом Project Jupyter – URL: <https://jupyter.org/>

9 Материально-техническое обеспечение дисциплины

Учебная аудитория 4205 для проведения занятий лекционного и семинарского типа (практические занятия); групповых и индивидуальных консультаций; текущего контроля и промежуточной аттестации.

Оснащенность:

Учебная мебель: столы – 35, стулья – 75, доска меловая, кафедра.

Основное оборудование: экран для проектора 1 шт., проектор – 1 шт., компьютер в комплекте - 1 шт.

Учебная аудитория 4202 Компьютерный класс, для проведения лабораторных занятий, самостоятельной работы.

Оснащенность:

Учебная мебель: стол преподавателя, компьютерные столы – 15, компьютерные кресла – 16.

Основное оборудование: 15 компьютеров с доступом в электронно-образовательную среду Академии, ЭБС и сети Интернет.

Учебная аудитория 4203 Компьютерный класс.

Оснащенность:

Учебная мебель: стол преподавателя, компьютерные столы – 15, компьютерные кресла – 16;

Основное оборудование: 15 компьютеров с доступом в электронно-образовательную среду Академии, ЭБС и сети Интернет.

Особенности реализации дисциплины (модуля) для инвалидов и лиц с ограниченными возможностями здоровья

Для обеспечения образования инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) реализация дисциплины может осуществляться в адаптированном виде, исходя из индивидуальных психофизических особенностей и по личному заявлению обучающегося, в части создания специальных условий.

В специальные условия могут входить: предоставление отдельной аудитории, необходимых технических средств, присутствие ассистента, оказывающего необходимую техническую помощь, выбор формы предоставления инструкции по порядку проведения текущего контроля и промежуточной аттестации, использование специальных технических средств, предоставление перерыва для приема пищи, лекарств и др.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

9 Карта компетенций дисциплины

Машинное обучение в агроинженерии (направление подготовки – 35.03.06 «Агроинженерия»)					
Цель дисциплины	формирование компетенций эффективного применения методов машинного обучения для решения агротехнических задач.				
Задачи дисциплины	<ul style="list-style-type: none"> • изучение и практическое освоение технологий исследования данных, их визуализации и интерпретации; • изучение и практическое освоение методов и программных средств классификации, числового прогнозирования и кластеризации; • изучение и практическое освоение методов и программных средств обнаружения закономерностей на основе ассоциативных правил и понижения размерности данных; • получение навыков программирования систем искусственного интеллекта. 				
В процессе освоения данной дисциплины студент формирует и демонстрирует следующие					
Компетенции		Планируемые результаты обучения (индикаторы достижения компетенции)	Технологии формирования	Форма оценочного средства	Ступени уровней освоения компетенции
Индекс	Формулировка				
Профессиональные компетенции					
ПК14	Способен использовать системы искусственного интеллекта в решении задач анализа, прогнозирования, планирования, синтеза и принятия решений	ИПК-14.1. Выбирает методы решения задач с использованием систем искусственного интеллекта ИПК-14.2. Решает задачи с использованием систем искусственного интеллекта	Лекции Лабораторные занятия Самостоятельная работа	Устный опрос Индивидуальное задание	<p>Пороговый уровень</p> <p>Частично знает основные методы решения задач с использованием систем искусственного интеллекта; классы задач, решаемых с использованием систем искусственного интеллекта, и критерии оценки получаемых результатов решения. В основном умеет выбирать методы решения задач с использованием систем искусственного интеллекта; применять методы обращения к целевым переменным и задания исходных данных и ограничений; оценивать полученные результаты решения задач с использованием систем искусственного интеллекта.</p> <p>Продвинутый уровень</p> <p>Знает основные методы решения задач с использованием систем искусственного интеллекта; классы задач, решаемых с использованием систем искусственного интеллекта, и критерии оценки получаемых результатов решения. Умеет выбирать методы решения задач с использованием систем искусственного интеллекта; применять методы обращения к целевым переменным и задания исходных</p>

					<p>данных и ограничений; оценивать полученные результаты решения задач с использованием систем искусственного интеллекта.</p> <p>Высокий уровень</p> <p>Демонстрирует глубокое знание и понимание: основных методов решения задач с использованием систем искусственного интеллекта; классов задач, решаемых с использованием систем искусственного интеллекта, и критериев оценки получаемых результатов решения. Полностью верно и самостоятельно умеет: выбирать методы решения задач с использованием систем искусственного интеллекта; применять методы обращения к целевым переменным и задания исходных данных и ограничений; оценивать полученные результаты решения задач с использованием систем искусственного интеллекта.</p>
ПК-15	Способен разрабатывать и применять методы машинного обучения для решения задач	<p>ИПК-15.1. Проводит анализ требований и определяет необходимые классы задач машинного обучения</p> <p>ИПК-15.2. Определяет метрики оценки результатов моделирования и критерии качества построенных моделей</p> <p>ИПК-15.3. Принимает участие в оценке и выборе используемых методов машинного обучения</p>	<p>Лекции</p> <p>Лабораторные занятия</p> <p>Самостоятельная работа</p>	<p>Устный опрос</p> <p>Индивидуальное задание</p>	<p>Пороговый уровень</p> <p>Частично знает принципы и методы машинного обучения, типы и классы задач машинного обучения; методы и критерии оценки качества моделей машинного обучения; классические методы и алгоритмы машинного обучения: предиктивное обучение с учителем, дескриптивные – обучение без учителя. В основном умеет сопоставить задачам предметной области классы задач машинного обучения; определять критерии и метрики оценки результатов моделирования при построении систем искусственного интеллекта в исследуемой области; проводить сравнительный анализ и осуществлять выбор методов и алгоритмов для решения задач машинного обучения.</p> <p>Продвинутый уровень</p> <p>Знает принципы и методы машинного обучения, типы и классы задач машинного обучения; методы и критерии оценки качества моделей машинного обучения; классические методы и алгоритмы машинного обучения: предиктивное обучение с учителем, дескриптивные – обучение без учителя. Умеет сопоставить задачам предметной области классы задач</p>

				<p>машинного обучения; определять критерии и метрики оценки результатов моделирования при построении систем искусственного интеллекта в исследуемой области; проводить сравнительный анализ и осуществлять выбор методов и алгоритмов для решения задач машинного обучения.</p> <p>Высокий уровень</p> <p>Демонстрирует глубокое знание и понимание: принципов и методов машинного обучения, типы и классы задач машинного обучения; методы и критерии оценки качества моделей машинного обучения; классических методов и алгоритмов машинного обучения: предиктивное обучение с учителем, дескриптивное –обучение без учителя. Полностью верно и самостоятельно умет: сопоставить задачам предметной области классы задач машинного обучения; определять критерии и метрики оценки результатов моделирования при построении систем искусственного интеллекта в исследуемой области; проводить сравнительный анализ и осуществлять выбор методов и алгоритмов для решения задач машинного обучения.</p>
--	--	--	--	--