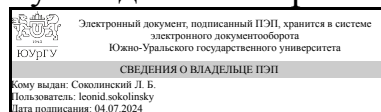


ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Руководитель направления



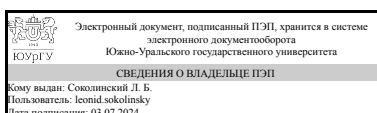
Л. Б. Соколинский

РАБОЧАЯ ПРОГРАММА

дисциплины 1.О.09 Машинное обучение
для направления 09.04.04 Программная инженерия
уровень Магистратура
форма обучения очная
кафедра-разработчик Системное программирование

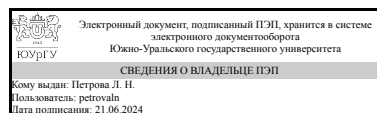
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 09.04.04 Программная инженерия, утверждённым приказом Минобрнауки от 19.09.2017 № 932

Зав.кафедрой разработчика,
д.физ.-мат.н., проф.



Л. Б. Соколинский

Разработчик программы,
старший преподаватель



Л. Н. Петрова

1. Цели и задачи дисциплины

Цель – сформировать у студентов навыки работы с данными и решения прикладных задач, дать представление об основных методах машинного обучения и видах задач, решаемых ими. Задачи: 1. Ознакомить студентов с основными задачами машинного обучения. 2. Дать представление об основных методах машинного обучения, выбора модели для конкретной задачи, оценке качества модели и ее настройке. 3.

Сформировать практические навыки решения задач машинного обучения, показать готовые реализации методов машинного обучения в современных библиотеках.

Краткое содержание дисциплины

Основные типы задач, решаемых с помощью методов машинного обучения, подготовка входных данных, оценка качества моделей, выбор модели для решения конкретной задачи, готовые реализации методов машинного обучения в современных библиотеках.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-3 Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач	Знает: классы методов и алгоритмов машинного обучения Умеет: ставить задачи и разрабатывать новые методы и алгоритмы машинного обучения Имеет практический опыт: решать основные классы задач методами и алгоритмами машинного обучения
ПК-4 Способен руководить проектами по созданию комплексных систем искусственного интеллекта	Знает: функциональность современных инструментальных средств и систем программирования в области создания моделей и методов машинного обучения

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Нет	1.О.06 Управление проектами в сфере искусственного интеллекта, 1.О.12 Интеллектуальный анализ данных

Требования к «входным» знаниям, умениям, навыкам студента, необходимым при освоении данной дисциплины и приобретенным в результате освоения предшествующих дисциплин:

Нет

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 5 з.е., 180 ч., 76,5 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам
		в часах
		Номер семестра
		1
Общая трудоёмкость дисциплины	180	180
<i>Аудиторные занятия:</i>	64	64
Лекции (Л)	32	32
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	32	32
Лабораторные работы (ЛР)	0	0
<i>Самостоятельная работа (СРС)</i>	103,5	103,5
Изучение тем и проблем, не выносимых на лекции и практические занятия	14	14
Подготовка к промежуточным тестам 1-7	21	21
Подготовка к защите практических работ	25	25
Подготовка к итоговому тесту	43,5	43,5
Консультации и промежуточная аттестация	12,5	12,5
Вид контроля (зачет, диф.зачет, экзамен)	-	экзамен

5. Содержание дисциплины

№ раздела	Наименование разделов дисциплины	Объем аудиторных занятий по видам в часах			
		Всего	Л	ПЗ	ЛР
1	Введение в машинное обучение. Задача линейной регрессии	14	8	6	0
2	Задача классификации. Логистическая регрессия. Проблема переобучения. Регуляризация	14	6	8	0
3	Метод kNN, деревья решений и ансамблевые методы.	12	6	6	0
4	Метод опорных векторов	10	4	6	0
5	Методы понижения размерности и визуализации данных	8	4	4	0
6	Дополнительные вопросы обучения моделей машинного обучения	6	4	2	0

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Вводная лекция. Примеры задач машинного обучения с учителем и без.	2
2	1	Одномерная линейная регрессия и метод максимального правдоподобия	2
3	1	Функция потерь, метод градиентного спуска	2
4	1	Множественная линейная регрессия. Нормализация признаков. Построение нелинейных моделей.	2
5	2	Задача бинарной классификации. Логистическая регрессия. Сигмоида и логлосс.	2
6	2	Задача множественной классификации.	2

7	2	Проблема переобучения. Регуляризация. Гребневая регрессия. Лассо.	2
8	3	Метод ближайших соседей kNN и его модификации.	2
9	3	Деревья решений.	2
10	3	Ансамбли деревьев решений.	2
11	4	Метод опорных векторов. Оптимальная разделяющая гиперплоскость. Зазор между классами. Функции ядра (kernel trick).	2
12	4	Нелинейный SVM.	2
13	5	Методы понижения размерности данных - метод главных компонент (PCA) и метод t-SNE.	2
14	5	Сжатие и визуализация данных.	2
15	6	Организация надежной валидации (dataset split, cross-validation), анализ learning curves.	2
16	6	Метрики качества моделей, несбалансированные классы, отбор признаков.	2

5.2. Практические занятия, семинары

№ занятия	№ раздела	Наименование или краткое содержание практического занятия, семинара	Кол-во часов
1	1	Реализация метода одномерной линейной регрессии для решения задачи прогнозирования прибыли при открытии нового филиала сети ресторанов	2
2	1	Реализация метода множественной линейной регрессии для решения задачи предсказания цены на дом при известных значениях площади дома и числе комнат.	2
3	1	Библиотека sklearn и ее методы решения задач линейной регрессии	2
4	2	Решение задач бинарной классификации методом логистической регрессии для случая линейно разделимых классов.	2
5	2	Решение задач бинарной классификации методом логистической регрессии для случая линейно не разделимых классов.	2
6	2	Использование логистической регрессии для решения задачи множественной классификации.	2
7	2	Распознавание рукописных цифр от 0 до 9.	2
8	3	Классификация ирисов Фишера с помощью метода kNN.	2
9	3	Использование ансамблевых моделей на основе деревьев решений.	2
10	3	Использование ансамблевых моделей на основе деревьев решений.	2
11	4	Настройка параметров SVM.	2
12	4	Использование SVM для решения задачи бинарной классификации.	2
13	4	Построение классификатора спама на основе SVM.	2
14	5	Использование PCA для уменьшения размерности данных.	2
15	5	Метод t-SNE	2
16	6	Изучение метрик качества в задаче с несбалансированными классами.	2

5.3. Лабораторные работы

Не предусмотрены

5.4. Самостоятельная работа студента

Выполнение СРС			
Подвид СРС	Список литературы (с указанием разделов, глав, страниц) / ссылка на	Семестр	Кол-во

	ресурс		часов
Изучение тем и проблем, не выносимых на лекции и практические занятия	Методические указания к дисциплине. Тема: "Отбор признаков" (Ибрияевой О.Л.)	1	14
Подготовка к промежуточным тестам 1-7	Презентации, выложенные в курсе Электронного ЮУрГУ	1	21
Подготовка к защите практических работ	Презентации в курсе Электронного ЮУрГУ	1	25
Подготовка к итоговому тесту	Презентации в курсе Электронного ЮУрГУ	1	43,5

6. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации

Контроль качества освоения образовательной программы осуществляется в соответствии с Положением о балльно-рейтинговой системе оценивания результатов учебной деятельности обучающихся.

6.1. Контрольные мероприятия (КМ)

№ КМ	Се-местр	Вид контроля	Название контрольного мероприятия	Вес	Макс. балл	Порядок начисления баллов	Учитывается в ПА
1	1	Текущий контроль	Тест 1. Линейная регрессия	1	4	Компьютерный тест содержит 4 равнозначных вопроса. Число баллов равно числу правильных ответов. Время на прохождение теста - 5 минут.	экзамен
2	1	Текущий контроль	Тест 2. Логистическая регрессия. Регуляризация	1	3	Компьютерный тест содержит 3 равнозначных вопроса. Число баллов равно числу правильных ответов. Время на прохождение теста - 4 минуты.	экзамен
3	1	Текущий контроль	Тест 3. Метод kNN	1	3	Компьютерный тест содержит 3 равнозначных вопроса. Число баллов равно числу правильных ответов. Время на прохождение теста - 4 минуты.	экзамен
4	1	Текущий контроль	Тест 4. Деревья решений и их ансамбли	1	4	Компьютерный тест содержит 4 равнозначных вопроса. Число баллов равно числу правильных ответов. Время на прохождение теста - 5 минут.	экзамен
5	1	Текущий контроль	Тест 5. SVM	1	4	Компьютерный тест содержит 4 равнозначных вопроса. Число баллов равно числу правильных ответов. Время на прохождение теста - 5 минут.	экзамен
6	1	Текущий контроль	Тест 6. PCA tSNE	1	2	Компьютерный тест содержит 2 равнозначных вопроса. Число баллов равно числу правильных ответов. Время на прохождение теста - 3 минуты.	экзамен
7	1	Текущий контроль	Тест 7. CV, GridSearch	1	3	Компьютерный тест содержит 3 равнозначных вопроса. Число баллов равно числу правильных ответов. Время на прохождение теста - 4 минуты.	экзамен
8	1	Текущий контроль	Практическая работа 1	1	12	Максимальное количество баллов - 12. 1. Правильно реализована функция	экзамен

					computeCost, вычисляющая значение целевой функции – 1 балл. 2. Правильно реализован алгоритм градиентного спуска в функции gradientDescent – 1 балл. 3. Получен график зависимости значений целевой функции от числа итераций – 1 балл. 4. Проведено исследование хода обучения при различных значениях параметра альфа и дано объяснение наблюдаемых явлений – 1 балл. 5. Найдено значение целевой функции в задаче множественной линейной регрессии – 1 балл. 6. Найдены новые параметры модели после 1000 итераций алгоритма градиентного спуска -1 балл. 7. Построен график изменения функции потерь в процессе обучения – 1 балл. 8. Решена задача одномерной линейной регрессии с помощью алгоритма из пакета sklearn -1 балл. 9. Решена задача множественной линейной регрессии с помощью алгоритма из пакета sklearn -1 балл. 10. Студент объясняет разницу между реализованным им алгоритмом градиентного спуска и алгоритмом Normal Equation, реализованным в sklearn -1 балл. 11. Студент демонстрирует свое умение применить полученную в первой задаче модель для предсказания выгоды от открытия ресторана в городе с населением 10 млн. - 1 балл. 12. Студент демонстрирует свое умение применить полученную во второй задаче модель для предсказания цены на дом площадью 3000 футов и числом комнат 4 – 1 балл.	
9	1	Текущий контроль	Практическая работа 2	1	Максимальное количество баллов - 12. 1. Правильно реализована функция сигмоида – 1 балл. 2. Правильно реализована функция costFunction – 1 балл. 3. Правильно реализована функция gradientFunc – 1 балл. 4. Построена граница решений в первой задаче, видно, что прямая отделяет два класса между собой – 1 балл. 5. Правильно предсказана вероятность поступления абитуриента с баллами 45, 85 – 1 балл. 6. Верно добавлены признаки в задаче классификации с линейно	экзамен

					неразделимыми классами -1 балл. 7. Верно добавлено слагаемое, отвечающее за регуляризацию в функции <code>costFunction</code> и <code>gradientFunc</code> – 1 балл. 8. Верно построена граница решений во второй задаче при трех разных значениях параметра регуляризации – 1 балл. 9. Студент объясняет, почему при разных значениях параметра регуляризации он наблюдает разные границы решений и может предложить другой вариант устранения явления переобучения в данной задаче, помимо изменения коэффициента регуляризации -1 балл. 10. Студент загрузил и верно подготовил данные для обучения в третьей задаче, добавил столбец из 1 - 1 балл 11. Верно построены 10 бинарных классификаторов в третьей задаче – 1 балл 12. Найдена доля правильных ответов модели – 1 балл.	
10	1	Текущий контроль	Практическая работа 3	1	9 Максимальное количество баллов - 9. 1. Правильно реализована функция, находящая евклидово расстояние между двумя векторами, выдает указанные ответы при заданном входе – 1 балл. 2. Верно реализована функция <code>predict_classification(train_set, labels, test_row, num_neighbors)</code> – 1 балл. 3. Построен график зависимости точности классификатора от числа соседей 1,..., 60 – 1 балл 4. Студент демонстрирует умение использовать алгоритм ближайших соседей из пакета <code>sklearn</code> – 1 балл. 5. Проведено сравнение с методом логистической регрессии – 1 балл. 6. Верно применен алгоритм дерева решений для решения задачи классификации ирисов Фишера -1 балл. 7. Верно применен алгоритм случайного леса для решения задачи классификации ирисов Фишера -1 балл. 8. Верно применен алгоритм градиентного бустинга для решения задачи классификации ирисов Фишера - 1 балл. 9. Для хотя бы одного из методов (дерева решений, случайный лес, градиентный бустинг) получена переобученная модель и путем настройки параметров модели устранено переобучение – 1 балл.	экзамен

11	1	Текущий контроль	Практическая работа 4	1	12	Максимальное количество баллов - 12. 1. В задаче с бинарной классификацией линейно разделимых классов изучено влияние параметра C на качество классификации, студент понимает, когда он наблюдает явление переобучения – 1 балл. 2. В задаче бинарной классификации линейно неразделимых классов построена граница решений, оценена точность классификатора, студент понимает, за счет чего он может ее увеличить – 1 балл. 3. Студент реализует метод решетчатого поиска параметров в задаче с нелинейным SVM, строит границу решений для лучших параметров, понимает, как образуется решетка значений параметров – 1 балл. 4. В задаче Breast Cancer студент производит разбиение на тренировочную и тестовую выборку, строит модель нелинейного SVM, оценивает ее точность на обеих выборках – 1 балл. 5. Студент верно определяет максимальное и минимальное значения признаков в задаче Breast Cancer, приводит их к одному масштабу – 1 балл. 6. Студент проводит обучение модели в задаче Breast Cancer на нормализованных данных, верно делает вывод о необходимости нормализации при использовании SVM -1 балл. 7. Студент верно объясняет, как происходит перевод письма в набор чисел в задаче классификации спама – 1 балл. 8. Студент верно объясняет, как сформирован словарь в задаче классификации спама– 1 балл. 9. Студент верно отвечает на вопросы, что будет с вектором чисел, соответствующим письму, если слово отсутствует в словаре, или если присутствует в письме более одного раза - 1 балл. 10. Студент демонстрирует способность решать задачу классификации спама с помощью метода линейного SVM, определяет точность модели на тренировочной и тестовой выборках - 1 балл 11. Студент демонстрирует способность решать задачу классификации спама с	экзамен
----	---	------------------	-----------------------	---	----	---	---------

						помощью метода нелинейного SVM, определяет точность модели на тренировочной и тестовой выборках – 1 балл 12. Студент демонстрирует способность проверить работу модели на конкретных письмах и определить класс письма – 1 балл.	
12	1	Текущий контроль	Практическая работа 5	1	12	Максимальное количество баллов - 12. 1. Студент верно объясняет, для чего используются методы PCA и tSNE, в чем их сходство и различие – 1 балл. 2. Студент демонстрирует способность применять PCA для визуализации данных, на примере MNIST – 1 балл. 3. Студент демонстрирует способность применять tSNE для визуализации данных, на примере MNIST – 1 балл. 4. Студент понимает задачу, которая досталась ему по номеру варианта, знает смысл признаков, тип задачи – 1 балл. 5, 6, 7. Студент верно применяет метод машинного обучения к своей задаче. Всего требуется использовать три разных метода. Число баллов равно числу правильных применений методов, максимально в этом пункте можно набрать – 3 балла. 8, 9, 10. Студент верно использует метод решетчатого поиска оптимальных параметров каждого из его трех методов. Число баллов равно числу верных применений метода решетчатого поиска в трех выбранных алгоритмах. Максимально в этом пункте можно набрать – 3 балла. 11. Студент понимает необходимость разделения на тестовую и тренировочную выборку, демонстрирует это в своей работе – 1 балл 12. Студент знает, для чего и как используется кросс-валидация, демонстрирует это в своей работе – 1 балл.	экзамен
13	1	Промежуточная аттестация	Итоговый тест	-	20	Тестирование проводится в системе edu.susu.ru. Тест содержит 20 вопросов. Количество баллов равно числу правильных ответов. На выполнение теста дается 60 минут.	экзамен

6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
------------------------------	----------------------	---------------------

экзамен	При оценивании результатов учебной деятельности обучающегося по дисциплине используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (Положение о БРС утверждено приказом ректора от 24.05.2019 г. № 179, в редакции приказа ректора от 10.03.2022 г. № 25-13/09). Процедура прохождения промежуточной аттестации осуществляется согласно Положению о текущем контроле успеваемости и промежуточной аттестации (приказ ректора от 27.02.2024 № 33-13/09). Оценка за дисциплину формируется на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля следующим образом: • Отлично: Величина рейтинга обучающегося по дисциплине 85...100 %. • Хорошо: Величина рейтинга обучающегося по дисциплине 75...84 %. • Удовлетворительно: Величина рейтинга обучающегося по дисциплине 60...74 %. • Неудовлетворительно: Величина рейтинга обучающегося по дисциплине 0...59 %. Если студент согласен с оценкой, полученной по результатам текущего контроля, то он может в день, предшествующий промежуточной аттестации дать свое согласие на автомат в личном кабинете. В случае явки студента на промежуточную аттестацию, давшего свое согласие на автомат в личном кабинете, студент имеет право пройти мероприятия текущего контроля по дисциплине на промежуточной аттестации для улучшения своего рейтинга в день ее проведения. Снижение оценки в этом случае запрещено. Если студент не дал согласия в личном кабинете, то он может согласиться с оценкой лично на промежуточной аттестации в день ее проведения. Если студент не согласен с оценкой, то он имеет право пройти мероприятия текущего контроля по дисциплине на промежуточной аттестации для улучшения своего рейтинга в день ее проведения. Фиксация результатов учебной деятельности по дисциплине проводится в день промежуточной аттестации на основе согласия студента, данного им в личном кабинете. При отсутствии согласия в журнале дисциплины фиксация результатов происходит при личном присутствии студента. Если студент не дал согласие в личном кабинете и не явился на промежуточную аттестацию – ему выставляется «неявка». Промежуточная аттестация проводится в форме тестирования. Тестирование проводится в системе edu.susu.ru. Тест содержит 20 вопросов. На выполнение теста дается 60 минут. В этом случае оценка за дисциплину рассчитывается на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля и промежуточной аттестации.	В соответствии с пп. 2.5, 2.6 Положения
---------	---	---

6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ КМ												
		1	2	3	4	5	6	7	8	9	10	11	12	13
ПК-3	Знает: классы методов и алгоритмов машинного обучения	+	+	+	+	+	+	+	+	+	+	+	+	+
ПК-3	Умеет: ставить задачи и разрабатывать новые методы и алгоритмы машинного обучения								+	+	+	+	+	+
ПК-3	Имеет практический опыт: решать основные классы задач методами и алгоритмами машинного обучения								+	+	+	+	+	+

1. Python Software Foundation-Python (бессрочно)
2. -Python(бессрочно)

Перечень используемых профессиональных баз данных и информационных справочных систем:

Нет

8. Материально-техническое обеспечение дисциплины

Вид занятий	№ ауд.	Основное оборудование, стенды, макеты, компьютерная техника, предустановленное программное обеспечение, используемое для различных видов занятий
Лекции	110 (3г)	Проектор
Практические занятия и семинары	110 (3г)	Компьютерный класс
Экзамен	110 (3г)	Компьютерный класс