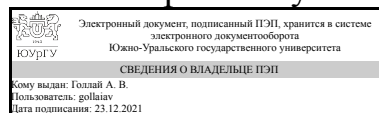


ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Директор института
Высшая школа электроники и
компьютерных наук



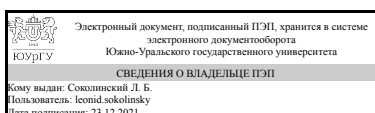
А. В. Голлой

РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.П2.20 Основы машинного обучения
для направления 11.03.02 Инфокоммуникационные технологии и системы связи
уровень Бакалавриат
профиль подготовки Коммуникационные технологии и интеллектуальная обработка данных
форма обучения очная
кафедра-разработчик Системное программирование

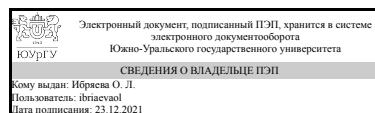
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 11.03.02 Инфокоммуникационные технологии и системы связи, утверждённым приказом Минобрнауки от 19.09.2017 № 930

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



Л. Б. Соколинский

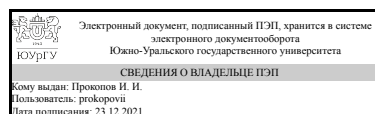
Разработчик программы,
к.физ.-мат.н., доцент



О. Л. Ибряева

СОГЛАСОВАНО

Руководитель образовательной
программы



И. И. Прокопов

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-7 Способен оценить параметры безопасности и защиты программного обеспечения и сетевых устройств администрируемой сети с помощью специальных средств управления безопасностью	Знает: основные алгоритмы машинного обучения и технологию их создания, валидации, оптимальной настройки Умеет: реализовывать алгоритмы машинного обучения, производить их оптимальную настройку, с использованием библиотек языка Python Имеет практический опыт: решения задач машинного обучения, их анализа и оптимизации

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Теория, методы и средства параллельной обработки информации, Основы цифровых устройств и математическая логика	Не предусмотрены

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

Дисциплина	Требования
Теория, методы и средства параллельной обработки информации	Знает: Способы организации современных многопроцессорных вычислительных систем. Технологию проектирования параллельных алгоритмов. Методы и средства разработки параллельных программ. Умеет: Применять на практике методы и средства разработки

	параллельных программ. Имеет практический опыт: Разработки параллельных программ с использованием стандарта OpenMP.
Основы цифровых устройств и математическая логика	Знает: Теоретические основы математической логики и теории алгоритмов. Алгоритмические системы и их характеристики. Методы и приемы формализации задач; методы построения рассуждений и логических конструкций; методы формального представления и построения алгоритмов. Умеет: Строить формальные доказательства и выводы; переводить на формальный язык содержательные математические утверждения; проверять истинность утверждений, записанных на формальном языке. Вырабатывать варианты реализации алгоритмов решения задач. Имеет практический опыт: решения проблемных задач, требующих применение логико-математического аппарата

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч., 56,5 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах	
		Номер семестра	
		7	
Общая трудоёмкость дисциплины	108	108	
<i>Аудиторные занятия:</i>	48	48	
Лекции (Л)	16	16	
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	32	32	
Лабораторные работы (ЛР)	0	0	
<i>Самостоятельная работа (СРС)</i>	51,5	51,5	
с применением дистанционных образовательных технологий	0		
Изучение тем и проблем, не выносимых на лекции и практические занятия	9	9	
Подготовка к защите практических работ	20	20	
Подготовка к итоговому тесту	8,5	8,5	
Подготовка к промежуточным тестам 1-7	14	14	
Консультации и промежуточная аттестация	8,5	8,5	
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	12	4	8	0
2	Задача классификации. Логистическая регрессия. Проблема переобучения. Регуляризация	12	4	8	0
3	Метод kNN, деревья решений и ансамблевые методы.	6	2	4	0
4	Метод опорных векторов	6	2	4	0
5	Методы понижения размерности и визуализации данных	6	2	4	0
6	Дополнительные вопросы обучения моделей машинного обучения	6	2	4	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. Одномерная линейная регрессия и метод максимального правдоподобия	2
2	1	Функция потерь, метод градиентного спуска. Множественная линейная регрессия. Нормализация признаков. Построение нелинейных моделей.	2
3	2	Задача бинарной классификации. Логистическая регрессия. Сигмоида и логлосс. Задача множественной классификации.	2
4	2	Проблема переобучения. Регуляризация. Гребневая регрессия. Лассо.	2
5	3	Метод ближайших соседей kNN и его модификации. Деревья решений. Ансамбли деревьев решений.	2
6	4	Метод опорных векторов. Оптимальная разделяющая гиперплоскость. Зазор между классами. Функции ядра (kernel trick). Нелинейный SVM.	2
7	5	Методы понижения размерности данных - метод главных компонент (PCA) и метод t-SNE. Сжатие и визуализация данных.	2
8	6	Организация надежной валидации (dataset split, cross-validation), анализ learning curves. Метрики качества моделей, 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	4
2	1	Реализация метода множественной линейной регрессии для решения задачи предсказания цены на дом при известных значениях площади дома и числе комнат.	4
3	2	Решение задач бинарной классификации методом логистической регрессии для случая линейно разделимых и не разделимых классов.	4
4	2	Использование логистической регрессии для решения задачи множественной классификации. Распознавание рукописных цифр от 0 до 9	4
5	3	Классификация ирисов Фишера с помощью метода kNN. Использование ансамблевых моделей на основе деревьев решений.	4
6	4	Использование SVM для решения задачи бинарной классификации. Построение классификатора спама на основе SVM.	4
7	5	Использование PCA для уменьшения размерности данных. Метод t-SNE.	4
8	6	Валидация модели машинного обучения. Изучение метрик качества в задаче с несбалансированными классами.	4

5.3. Лабораторные работы

Не предусмотрены

5.4. Самостоятельная работа студента

Выполнение СРС			
Подвид СРС	Список литературы (с указанием разделов, глав, страниц) / ссылка на ресурс	Семестр	Кол-во часов
Изучение тем и проблем, не выносимых на лекции и практические занятия	Методические указания к дисциплине Основы машинного обучения Тема конструирование признаков (Ибряева ОЛ)	7	9
Подготовка к защите практических работ	Презентации в курсе Электронного ЮУрГУ	7	20
Подготовка к итоговому тесту	Презентации в курсе Электронного ЮУрГУ	7	8,5
Подготовка к промежуточным тестам 1-7	Презентации, выложенные в курсе Электронного ЮУрГУ	7	14

6. Текущий контроль успеваемости, промежуточная аттестация

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	7	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	7	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	7	Текущий контроль	Test 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	7	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	7	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	7	Текущий	Test 6. PCA	1	2	Компьютерный тест содержит 2	экзамен

		контроль	tSNE			равнозначных вопроса. Число баллов равно числу правильных ответов. Время на прохождение теста - 3 минуты.	
7	7	Текущий контроль	Test 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	7	Текущий контроль	Практическая работа 1	1	12	Максимальное количество баллов - 12. 1. Правильно реализована функция 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. Найденны 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.	экзамен
9	7	Текущий контроль	Практическая работа 2	1	12	Максимальное количество баллов - 12. 1. Правильно реализована функция сигмоида – 1 балл. 2. Правильно реализована функция costFunction – 1 балл. 3. Правильно реализована функция	экзамен

					<p>gradientFunc – 1 балл.</p> <p>4. Построена граница решений в первой задаче, видно, что прямая отделяет два класса между собой – 1 балл.</p> <p>5. Правильно предсказана вероятность поступления абитуриента с баллами 45, 85 – 1 балл.</p> <p>6. Верно добавлены признаки в задаче классификации с линейно неразделимыми классами -1 балл.</p> <p>7. Верно добавлено слагаемое, отвечающее за регуляризацию в функции costFunction и gradientFunc – 1 балл.</p> <p>8. Верно построена граница решений во второй задаче при трех разных значениях параметра регуляризации – 1 балл.</p> <p>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.</p> <p>10. Студент загрузил и верно подготовил данные для обучения в третьей задаче, добавил столбец из 1 - 1 балл</p> <p>11. Верно построены 10 бинарных классификаторов в третьей задаче – 1 балл</p> <p>12. Найдена доля правильных ответов модели – 1 балл.</p>		
10	7	Текущий контроль	Практическая работа 3	1	9	<p>Максимальное количество баллов - 9.</p> <p>1. Правильно реализована функция, находящая евклидово расстояние между двумя векторами, выдает указанные ответы при заданном входе – 1 балл.</p> <p>2. Верно реализована функция predict_classification(train_set, labels, test_row, num_neighbors) – 1 балл.</p> <p>3. Построен график зависимости точности классификатора от числа соседей 1, ..., 60 – 1 балл</p> <p>4. Студент демонстрирует умение использовать алгоритм ближайших соседей из пакета sklearn – 1 балл.</p> <p>5. Проведено сравнение с методом логистической регрессии – 1 балл.</p> <p>6. Верно применен алгоритм дерева решений для решения задачи классификации ирисов Фишера -1 балл.</p> <p>7. Верно применен алгоритм случайного леса для решения задачи классификации ирисов Фишера -1 балл.</p> <p>8. Верно применен алгоритм</p>	экзамен

						<p>градиентного бустинга для решения задачи классификации ирисов Фишера - 1 балл.</p> <p>9. Для хотя бы одного из методов (деревья решений, случайный лес, градиентный бустинг) получена переобученная модель и путем настройки параметров модели устранено переобучение – 1 балл.</p>	
11	7	Текущий контроль	Практическая работа 4	1	12	<p>Максимальное количество баллов - 12.</p> <p>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.</p> <p>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.</p> <p>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.</p> <p>4. В задаче Breast Cancer студент производит разбиение на тренировочную и тестовую выборку, строит модель нелинейного SVM, оценивает ее точность на обоих выборках – 1 балл.</p> <p>5. Студент верно определяет максимальное и минимальное значения признаков в задаче Breast Cancer, приводит их к одному масштабу – 1 балл.</p> <p>6. Студент проводит обучение модели в задаче Breast Cancer на нормализованных данных, верно делает вывод о необходимости нормализации при использовании SVM -1 балл.</p> <p>7. Студент верно объясняет, как происходит перевод письма в набор чисел в задаче классификации спама – 1 балл.</p> <p>8. Студент верно объясняет, как сформирован словарь в задаче классификации спама– 1 балл.</p> <p>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</p>	экзамен

						<p>- 1 балл.</p> <p>10. Студент демонстрирует способность решать задачу классификации спама с помощью метода линейного SVM, определяет точность модели на тренировочной и тестовой выборках - 1 балл</p> <p>11. Студент демонстрирует способность решать задачу классификации спама с помощью метода нелинейного SVM, определяет точность модели на тренировочной и тестовой выборках – 1 балл</p> <p>12. Студент демонстрирует способность проверить работу модели на конкретных письмах и определить класс письма – 1 балл.</p>	
12	7	Текущий контроль	Практическая работа 5	1	12	<p>Максимальное количество баллов - 12.</p> <p>1. Студент верно объясняет, для чего используются методы PCA и tSNE, в чем их сходство и различие – 1 балл.</p> <p>2. Студент демонстрирует способность применять PCA для визуализации данных, на примере MNIST – 1 балл.</p> <p>3. Студент демонстрирует способность применять tSNE для визуализации данных, на примере MNIST – 1 балл.</p> <p>4. Студент понимает задачу, которая досталась ему по номеру варианта, знает смысл признаков, тип задачи – 1 балл.</p> <p>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.</p> <p>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.</p> <p>11. Студент понимает необходимость разделения на тестовую и тренировочную выборку, демонстрирует это в своей работе – 1 балл</p> <p>12. Студент знает, для чего и как используется кросс-валидация, демонстрирует это в своей работе – 1 балл.</p>	экзамен
13	7	Промежуточная аттестация	Итоговый тест	-	20	Компьютерный тест содержит 20 равнозначных вопросов. Число баллов равно числу правильных ответов. Время	экзамен

6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
экзамен	На экзамене происходит оценивание учебной деятельности обучающихся по дисциплине на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля и промежуточной аттестации. При оценивании результатов учебной деятельности обучающегося по дисциплине используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179) 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 %. Допускается выставление оценки на основе текущего рейтинга (автоматом).	В соответствии с пп. 2.5, 2.6 Положения

6.3. Оценочные материалы

Компетенции	Результаты обучения	№ КМ												
		1	2	3	4	5	6	7	8	9	10	11	12	13
ПК-7	Знает: основные алгоритмы машинного обучения и технологию их создания, валидации, оптимальной настройки	+	+	+	+	+	+	+	+	+	+	+	+	+
ПК-7	Умеет: реализовывать алгоритмы машинного обучения, производить их оптимальную настройку, с использованием библиотек языка Python			+	+		+		+	+	+	+		+
ПК-7	Имеет практический опыт: решения задач машинного обучения, их анализа и оптимизации	+							+	+	+	+		+

Фонды оценочных средств по каждому контрольному мероприятию находятся в приложениях.

7. Учебно-методическое и информационное обеспечение дисциплины

Печатная учебно-методическая документация

а) основная литература:

1. Демидов, А. К. Искусственный интеллект [Текст] учеб. пособие А. К. Демидов, Б. М. Кувшинов ; Юж.-Урал. гос. ун-т, Каф. Приклад. математика ; ЮУрГУ. - Челябинск: Издательство ЮУрГУ, 2008. - 65, [1] с. ил.
2. Ясницкий, Л. Н. Введение в искусственный интеллект [Текст] учеб. пособие Л. Н. Ясницкий. - 2-е изд., испр. - М.: Академия, 2008. - 174, [1] с.

б) дополнительная литература:

1. Ясницкий, Л. Н. Введение в искусственный интеллект Учеб. пособие для вузов по специальности 010100 "Математика" Л. Н. Ясницкий. - М.: Academia, 2005. - 174, [1] с.

в) отечественные и зарубежные журналы по дисциплине, имеющиеся в библиотеке:

Не предусмотрены

г) методические указания для студентов по освоению дисциплины:

1. Методические указания к дисциплине Основы машинного обучения Тема конструирование признаков (Ибряева ОЛ)

из них: учебно-методическое обеспечение самостоятельной работы студента:

1. Методические указания к дисциплине Основы машинного обучения Тема конструирование признаков (Ибряева ОЛ)

Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в электронной форме	Библиографическое описание
1	Основная литература	Электронно-библиотечная система издательства Лань	Коэльо, Л. П. Построение систем машинного обучения на языке Python / Л. П. Коэльо, В. Ричарт ; перевод с английского А. А. Слинкин. — 2-е изд. — Москва : ДМК Пресс, 2016. — 302 с. — ISBN 978-5-97060-330-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/82818 (дата обращения: 08.10.2021). — Режим доступа: для авториз. пользователей.
2	Основная литература	Электронно-библиотечная система издательства Лань	Воронина, В. В. Теория и практика машинного обучения : учебное пособие / В. В. Воронина. — Ульяновск : УлГТУ, 2017. — 290 с. — ISBN 978-5-9795-1712-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/165053 (дата обращения: 08.10.2021). — Режим доступа: для авториз. пользователей.

Перечень используемого программного обеспечения:

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
Практические занятия и семинары	110 (3г)	Компьютерный класс
Лекции	110 (3г)	Проектор
Экзамен	110	Компьютерный класс

	(3r)	
--	------	--