

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО

**Проректор по учебной работе и
довузовской подготовке**

А.А. Воронов

	Рабочая программа дисциплины (модуля)
по дисциплине:	Практикум по научному программированию
по направлению:	Информатика и вычислительная техника
профиль подготовки:	Физтех-школа Прикладной Математики и Информатики кафедра компьютерной лингвистики
курс:	3
квалификация:	бакалавр

Семестр, формы промежуточной аттестации: 6 (весенний) - Дифференцированный зачет

Аудиторных часов: 30 всего, в том числе:

лекции: 15 час.

семинары: 15 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 15 час.

Всего часов: 45, всего зач. ед.: 1

Количество контрольных работ, заданий: 1

Программу составил: А.Е. Жуковский

Программа обсуждена на заседании кафедры компьютерной лингвистики 04.06.2020

Аннотация

Курс нацелен на получение студентами базовых знаний о прикладном программировании в научной сфере, а также о грамотной постановке и выполнении научного эксперимента.

Он охватывает как повторение основных свойств языка Python (базовый синтаксис и объектно-ориентированное программирование), так и изучение основных библиотек визуализации и математических расчетов: pandas, matplotlib и пр. Немаловажной частью курса является грамотное применение методов машинного обучения – для этого студентам будет предложено работать с библиотекой scikit-learn.

1. Цели и задачи

Цель дисциплины

данный курс должен сформировать представление о навыках решения наиболее важных с прикладной точки зрения задач научного программирования, в особенности на языке Python.

Задачи дисциплины

- понимание специфики научной задачи;
- изучения принципов построения адекватных моделей и эволюции,
- изучение выбора архитектуры сети в зависимости от решаемых задач;
- познакомиться с современными технологиями и методиками решения научных задач.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1 Формулирует совокупность взаимосвязанных задач в рамках поставленной цели работы, обеспечивающих ее достижение. Определяет ожидаемые результаты решения поставленных задач
	УК-2.2 Проектирует решение конкретной задачи проекта, выбирая оптимальный способ ее решения, исходя из действующих правовых норм и имеющихся ресурсов и ограничений
ОПК-3 Способен составлять и оформлять научные и (или) технические (технологические, инновационные) отчеты (публикации, проекты)	ОПК-3.1 Знает основные правила оформления научных публикаций и научно-технической документации, в том числе с использованием прикладного программного обеспечения
	ОПК-3.2 Владеет на практике методологией составления научно-технических отчетов (проектов)
	ОПК-3.3 Владеет методами визуального и графического представления результатов научной (научно-технической, инновационной технологической) деятельности в виде отчетов, научных публикаций
ОПК-5 Способен участвовать в проведении фундаментальных и прикладных	ОПК-5.2 Обладает способностью к освоению новых знаний на основе изучения литературы, научных статей и других источников

исследований и разработок, самостоятельно осваивать новые теоретические, в том числе, математические методы исследований и работать на современной экспериментальной научно-исследовательской, измерительно-аналитической и технологической аппаратуре)	ОПК-5.1 Способен решать поставленные задачи в области теоретических и экспериментальных исследований и разработок
	ОПК-5.3 Способен к профессиональной эксплуатации современной экспериментальной научно-исследовательской (измерительно-аналитической и технологической) аппаратуры
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценить качество разработанной модели
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
ПК-2 Способен самостоятельно или в качестве члена (руководителя) малого коллектива организовывать и проводить научные исследования и их апробацию	ПК-2.1 Знает принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации
	ПК-2.2 Способен планировать и проводить научные исследования самостоятельно или в качестве члена (руководителя) малого научного коллектива
	ПК-2.3 Способен проводить апробацию результатов научно-исследовательской работы посредством публикации научных статей и участия в конференциях

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- контролировать точность вычислений в Python;
- уметь проводить вычисления в Python;
- способы валидации гипотезы;
- основные программные расширения для научного программирования (библиотеки scikit-learn, pandas), среды разработки (Jupyter).

уметь:

- строить и проверять научные гипотезы;
- применять инструменты для эффективной работы с языком Python.

владеть:

- навыками освоения большого объема информации;
- навыками самостоятельной работы;
- культурой грамотной разработки архитектур нейронных сетей.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа

1	Введение в язык Python. Python как инструмент для анализа данных.	2	2		2
2	Базовые понятия языка. ООП, основные структуры данных, функции, интерпретатор, основные библиотеки.	2	2		2
3	Базовые принципы распределения данных и вычислений.	1	1		1
4	Модули для математических вычислений: math, cmath и другие.	1	1		1
5	Работа с массивами данных: вектора признаков, словари, строки.	1	1		1
6	Визуализация данных. Работа с визуализированными математическими данными и их анализ (библиотеки matplotlib и другие).	2	2		2
7	Работа с данными из внешних источников. Чтение из файла.	1	1		1
8	Введение в классы. Классы и их специальные методы, применимость в научном программировании.	2	2		2
9	Выборки случайных чисел. Генерация и анализ выборок. Статистический анализ данных.	1	1		1
10	Основы машинного обучения: обучение без учителя/с учителем, обучение с подкреплением, бустинг, байесовская сеть.	1	1		1
11	Обзор основных профильных статей.	1	1		1
Итого часов		15	15		15
Подготовка к экзамену		0 час.			
Общая трудоёмкость		45 час., 1 зач.ед.			

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 6 (Весенний)

1. Введение в язык Python. Python как инструмент для анализа данных.

Основы языка Python. Почему и какой ЯП пригоден для научных исследований. Особенности синтаксиса и парсинга данных в общем виде.

2. Базовые понятия языка. ООП, основные структуры данных, функции, интерпретатор, основные библиотеки.

Общие принципы передачи данных. Классы, функции, интерпретатор, основные библиотеки. Учимся использовать готовые функции библиотеки Pandas, объединять данные из разных источников, работать с большими файлами.

3. Базовые принципы распределения данных и вычислений.

Вертикальное и горизонтальное распределение. Распределённые базы данных. Многозвенные архитектуры. Репликация. Основные проблемы, возникающие при распределении ИС и пути их решения.

4. Модули для математических вычислений: math, cmath и другие.

Встроенные функции языка, например, размещенные в math. Старшинство арифметических операций, использование встроенных математических функций, использование интерактивной оболочки Python в качестве калькулятора. Ошибки: ошибки целочисленного деления и погрешности вычислений.

5. Работа с массивами данных: вектора признаков, словари, строки.

Проведение анализа и рекурсивного выбора признаков на базе моделей. Методы оценки значимости и отбора признаков, их использование. «Проклятие размерности», основные алгоритмы и принципы их работы. Продвинутое типы данных: массивы, множества, словари.

6. Визуализация данных. Работа с визуализированными математическими данными и их анализ (библиотеки matplotlib и другие).

Библиотеки:

- numpy и scipy;
- pandas;
- Визуализация данных: seaborn, plotly, matplotlib.

7. Работа с данными из внешних источников. Чтение из файла.

Получение данных с внешних сайтов и API. Data mining и парсинг.

8. Введение в классы. Классы и их специальные методы, применимость в научном программировании.

Основы Python и Git как инструмента совместной работы. Базовые типы данных и циклы. Функции и классы. Продвинутое типы данных: массивы, множества, словари.

9. Выборки случайных чисел. Генерация и анализ выборок. Статистический анализ данных.

Понятия: среднее, медиана и квартили, одномерный и многомерный анализ, коллинеарность. Центральная предельная теорема и статистический анализ данных в Python. Статистические тесты и проверка гипотез. Статистические показатели в Python. Выборка и доверительный интервал для стат. значимости теста и проектирование дизайна A/B-тестов. Описательная статистика и виды распределений в Python.

10. Основы машинного обучения: обучение без учителя/с учителем, обучение с подкреплением, бустинг, байесовская сеть.

Библиотека Sklearn. Алгоритмы классификации: линейные методы, логистическая регрессия и SVM, деревья решений. Алгоритмы регрессии: линейная и полиномиальная. Алгоритмы кластеризации. Ансамблирование. Оценка точности модели, переобучение, регуляризация. Улучшение качества модели. Анализ и обработка временных рядов.

11. Обзор основных профильных статей.

Анализ профильной литературы по NLP/CV тематике. Основные архитектуры моделей обеих областей: векторные представления, тематические смещения, машинный перевод; оценка датасета; поиск по картинкам, сегментация, GAN/RNN/CNN-архитектуры.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

6.Перечень рекомендуемой литературы

Основная литература

1. Программирование на Python 3 : Подробное руководство [Текст] = Programming in Python 3 : [учеб. пособие для вузов] / М. Саммерфилд; пер. с англ. А. Киселева .— СПб : Символ-Плюс, 2015 .— 608 с.

Дополнительная литература

1. Introduction to Machine Learning with Python. A Guide for Data Scientists /Andreas C. Muller, Sarah Guido. Beiling ; Boston, O'REILLY, 2017
2. Hands-on data structures and algorithms with Python, Write complex and powerful code using the latest features of Python 3.7 /B. Agarwal, B. Baka. Birmingham ; Mumbai, Packt, 2018
3. Python machine learning, Machine learning and deep learning with Python, scikit-learn, and TensorFlow/S. Raschka, V. Mirjalili, -Birmingham ; Mumbai, Packt, 2017

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

Не используются

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

на практических занятиях используются мультимедийные технологии, включая демонстрацию презентаций.

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Студент, изучающий дисциплину, должен с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике. В результате изучения дисциплины студент должен знать основные определения дисциплины, уметь применять полученные знания для решения различных задач.

Успешное освоение курса требует:

- посещения всех занятий, предусмотренных учебным планом по дисциплине;
- ведения конспекта занятий;
- напряжённой самостоятельной работы студента.

Самостоятельная работа включает в себя:

- чтение рекомендованной литературы;
- проработку учебного материала, подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- решение задач, предлагаемых студентам на занятиях;
- подготовку к выполнению заданий текущей и промежуточной аттестации.

Показателем владения материалом служит умение без конспекта отвечать на вопросы по темам дисциплины.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов, следует обращаться за консультациями преподавателю.

Возможен промежуточный контроль знаний студентов в виде решения задач в соответствии с тематикой занятий.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению: Информатика и вычислительная техника

профиль подготовки: Физтех-школа Прикладной Математики и Информатики
кафедра компьютерной лингвистики

курс: 3

квалификация: бакалавр

Семестр, формы промежуточной аттестации: 6 (весенний) - Дифференцированный зачет

Разработчик: А.Е. Жуковский

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1 Формулирует совокупность взаимосвязанных задач в рамках поставленной цели работы, обеспечивающих ее достижение. Определяет ожидаемые результаты решения поставленных задач
	УК-2.2 Проектирует решение конкретной задачи проекта, выбирая оптимальный способ ее решения, исходя из действующих правовых норм и имеющихся ресурсов и ограничений
ОПК-3 Способен составлять и оформлять научные и (или) технические (технологические, инновационные) отчеты (публикации, проекты)	ОПК-3.1 Знает основные правила оформления научных публикаций и научно-технической документации, в том числе с использованием прикладного программного обеспечения
	ОПК-3.2 Владеет на практике методологией составления научно-технических отчетов (проектов)
	ОПК-3.3 Владеет методами визуального и графического представления результатов научной (научно-технической, инновационной технологической) деятельности в виде отчетов, научных публикаций
ОПК-5 Способен участвовать в проведении фундаментальных и прикладных исследований и разработок, самостоятельно осваивать новые теоретические, в том числе, математические методы исследований и работать на современной экспериментальной научно-исследовательской, измерительно-аналитической и технологической аппаратуре)	ОПК-5.2 Обладает способностью к освоению новых знаний на основе изучения литературы, научных статей и других источников
	ОПК-5.1 Способен решать поставленные задачи в области теоретических и экспериментальных исследований и разработок
	ОПК-5.3 Способен к профессиональной эксплуатации современной экспериментальной научно-исследовательской (измерительно-аналитической и технологической) аппаратуры
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценить качество разработанной модели
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
ПК-2 Способен составлять математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-2.1 Знает принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации

ПК-2 Способен самостоятельно или в качестве члена (руководителя) малого коллектива организовывать и проводить научные исследования и их апробацию

ПК-2.2 Способен планировать и проводить научные исследования самостоятельно или в качестве члена (руководителя) малого научного коллектива

ПК-2.3 Способен проводить апробацию результатов научно-исследовательской работы посредством публикации научных статей и участия в конференциях

2. Показатели оценивания компетенций

В результате изучения дисциплины «Практикум по научному программированию» обучающийся должен:

знать:

- контролировать точность вычислений в Python;
- уметь проводить вычисления в Python;
- способы валидации гипотезы;
- основные программные расширения для научного программирования (библиотеки scikit-learn, pandas), среды разработки (Jupyter).

уметь:

- строить и проверять научные гипотезы;
- применять инструменты для эффективной работы с языком Python.

владеть:

- навыками освоения большого объема информации;
- навыками самостоятельной работы;
- культурой грамотной разработки архитектур нейронных сетей.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

1. Библиотеки numpy и scipy;
2. Библиотеки pandas;
3. Библиотеки визуализации данных: seaborn, plotly, matplotlib;
4. Библиотека Sklearn.
5. Основные архитектуры моделей обеих областей;
6. Векторные представления;
7. Тематические смещения;
8. Машинный перевод;
9. Оценка датасета;
10. Поиск по картинкам;
11. Сегментация, GAN/RNN/CNN-архитектуры.

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

1. Общие принципы передачи данных.
2. Классы, функции, интерпретатор, основные библиотеки.
3. Вертикальное и горизонтальное распределение.
4. Распределённые базы данных.
5. Многозвенные архитектуры.
6. Репликация.
7. Основные проблемы, возникающие при распределении ИС и пути их решения.
8. Встроенные функции языка, например, размещенные в math.
9. Старшинство арифметических операций, использование встроенных математических функций, использование интерактивной оболочки Python в качестве калькулятора.
10. Ошибки: ошибки целочисленного деления и погрешности вычислений.
11. Проведение анализа и рекурсивного выбора признаков на базе моделей.
12. Методы оценки значимости и отбора признаков, их использование.

13. «Проклятие размерности», основные алгоритмы и принципы их работы.
14. Продвинутое типы данных: массивы, множества, словари.
15. Получение данных с внешних сайтов и API.
16. Data mining и парсинг.
17. Основы Python и Git как инструмента совместной работы.
18. Базовые типы данных и циклы.
19. Функции и классы.
20. Продвинутое типы данных: массивы, множества, словари.
21. Понятия: среднее, медиана и квартили, одномерный и многомерный анализ, коллинеарность.
22. Центральная предельная теорема и статистический анализ данных в Python.
23. Статистические тесты и проверка гипотез. Статистические показатели в Python.
24. Выборка и доверительный интервал для стат. значимости теста и проектирование дизайна A/B-тестов.
25. Описательная статистика и виды распределений в Python.
26. Алгоритмы классификации: линейные методы, логистическая регрессия и SVM, деревья решений
27. Алгоритмы регрессии: линейная и полиномиальная
28. Алгоритмы кластеризации
29. Ансамблирование
30. Оценка точности модели, переобучение, регуляризация
31. Улучшение качества модели
32. Анализ и обработка временных рядов
33. Анализ профильной литературы по NLP/CV тематике.

Критерии оценивания

отлично (10) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

отлично (9) - выставляется студенту, показавшему свободное оперирование знаниями учебной программы дисциплины, выполнение заданий творческого характера.

отлично (8) - выставляется студенту, показавшему владение программным учебным материалом с наличием несущественных ошибок в действиях, самостоятельно исправляемых учащимся.

хорошо (7) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускается в ответе или в решении задач некоторые неточности.

хорошо (6) - выставляется студенту если он осознает воспроизведение программного учебного материала, в том числе и различной степени сложности, с несущественными ошибками, затруднения в применении отдельных навыков.

хорошо (5) - выставляется студенту если теоретическое содержание освоено не полностью, некоторые практические навыки сформированы недостаточно, в некоторых случаях были допущены ошибки.

удовлетворительно (4) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации.

удовлетворительно (3) - выставляется студенту в случае большого количества недочетов и неправильных ответов, а также пассивной работе в ходе занятий, многие учебные задания не выполнены.

неудовлетворительно (2) - выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач.

неудовлетворительно (1) - выставляется студенту, который не освоил теоретическое и практическое содержание курса, все выполненные учебные задания содержат грубые ошибки.

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Во время проведения дифференцированного зачета обучающиеся могут пользоваться программой дисциплины, а также справочной литературой, конспектами лекций или другими материалами.