Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Максимов Алексей Борисович
Должность: директор департамента по образовательной политике
Дата подписания: 31.05.2024 17:12:15
Уникальный программный ключ:
8db180d1a3f02ac9e60521a5672742735c18b1d6

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное автономное образовательное учреждение
высшего образования
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
Факультет информационных технологий

УТВЕРЖДАЮ
Декан факультета
«Информационные технологии»
_____ / Д.Г.Демидов /
«15» февраля 2024г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Технологии анализа больших данных»

Направление подготовки/специальность
01.04.02 Прикладная математика и информатика
Профиль/специализация
«Системная аналитика больших данных»

Квалификация
Магистр

Формы обучения
Очная

Москва, 2024 г.

Разработчик(и):

к.ф-м.н, доцент / А.В. Филимонов/

Согласовано:
Заведующий кафедрой «Прикладная информатика»,
к.э.н, доцент /С.В. Суворов/

Содержание

Оглавление

1 Цели, задачи и планируемые результаты обучения по дисциплине 4
2 Место дисциплины в структуре образовательной программы 6
3 Структура и содержание дисциплины 6
3.1 Виды учебной работы и трудоемкость 6
3.2 Тематический план изучения дисциплины 7
3.3 Содержание дисциплины 7
3.4 Тематика семинарских/практических и лабораторных занятий 9
3.5 Тематика курсовых проектов (курсовых работ) 10
4 Учебно-методическое и информационное обеспечение 10
4.1 Нормативные документы и ГОСТы 10
4.2 Основная литература 10
4.3 Дополнительная литература 10
4.4 Электронные образовательные ресурсы 11
4.5 Лицензионное и свободно распространяемое программное обеспечение 11
4.6 Современные профессиональные базы данных и информационные справочные системы 11
5 Материально-техническое обеспечение 11
6 Методические рекомендации 12
6.1 Методические рекомендации для преподавателя по организации обучения 12
6.2 Методические указания для обучающихся по освоению дисциплины 12
7 Фонд оценочных средств 12
7.1 Методы контроля и оценивания результатов обучения 12
7.2 Шкала и критерии оценивания результатов обучения 13
7.3 Оценочные средства 13

1 Цели, задачи и планируемые результаты обучения по дисциплине

Целями дисциплины являются получение представления о теории (технологии) работы с большими данными.

Уделить основное внимание способам уменьшения размерности данных и методам кластеризации.

Задачами дисциплины являются:

- Овладение технологиями, используемыми для уменьшения размерности данных с минимальными потерями информативности.
- Ознакомление с методами кластеризации.

Обучение по дисциплине «Технологии анализа больших данных» направлено на формирование у обучающихся следующих компетенций:

Код и наименование компетенций	Индикаторы достижения компетенции
УК-4. Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия	УК-4.1. Знать: правила и закономерности личной и деловой устной и письменной коммуникации; современные УК-4.2. Уметь: применять на практике коммуникативные технологии, методы и способы делового общения для академического и профессионального взаимодействия УК-4.3. Владеть: методикой межличностного делового общения на русском и иностранном языках, с применением профессиональных языковых форм, средств и современных коммуникативных технологий
ПК-4. Способен к разработке систем управления базами данных	ПК-4.1. Знать: теория баз данных, основные структуры данных, основные модели данных и их организация, принципы построения языков запросов и манипулирования данными, методы обработки данных, основы современных систем управления базами данных, методы поддержки, контроля и оптимизации баз данных, системы хранения и анализа баз данных, методы повышения надежности работы системы управления базами данных, методы построения баз знаний и принципы построения экспертных систем, синтаксис, особенности программирования и стандартные библиотеки выбранного языка программирования, конструкции распределенного и параллельного программирования, способы и механизмы управления данными, принципы организации, состав и схемы работы операционных систем, принципы управления ресурсами, методы организации файловых систем, принципы

	построения сетевого взаимодействия, основы архитектуры, устройства и функционирования вычислительных систем, архитектура и принципы функционирования коммуникационного оборудования, устройство и принципы функционирования информационных систем, стандарты информационного взаимодействия систем, принципы организации инфокоммуникационных систем, основы информационной безопасности, подходы к автоматизации и стандарты автоматизации организации, локальные правовые акты, действующие в организации, английский язык на уровне чтения технической документации в области информационных и компьютерных технологий ПК-4.2. Уметь: Идентифицировать класс разрабатываемой системы управления базами данных в зависимости от выполняемых ею задач, определенных в техническом задании на разработку системы управления базами данных, идентифицировать класс разрабатываемой системы управления базами данных в зависимости от аппаратных средств, определенных в техническом задании на разработку системы управления базами данных, создавать блок-схемы алгоритмов функционирования разрабатываемых компонентов системы управления базами данных, оценивать вычислительную сложность алгоритмов функционирования разрабатываемых компонентов системы управления базами данных, применять языки программирования, определенные в техническом задании на разработку системы управления базами данных, для написания программного кода, осуществлять подготовку и сохранение резервных копий данных, применять нормативно-техническую документацию при использовании систем управления базами данных ПК-4.3. Владеть: навыками получения технической документации на разработку системы управления базами данных, изучения технической документации на разработку системы управления базами данных, разработки структуры системы управления базами данных в целом и ее

	отдельных компонентов, создания блок-схемы системы управления базами данных, разработки системы многозадачного и многопользовательского режимов, разработки системы администрирования данных, разработки системы поддержки транзакционных механизмов, разработки системы масштабируемости системы управления базами данных, разработки системы контроля целостности данных, разработки системы безопасности системы управления базами данных, разработки системы резервного копирования, написанияисходного кода системы управления базами данных на языке программирования системы управления базами данных, передачи исходного кода системы управления базами данных на тестирование

2 Место дисциплины в структуре образовательной программы

Дисциплина относится к части, формируемой участниками образовательных отношений) блока Б1 «Дисциплины (модули)».

Дисциплина «Технологии анализа больших данных» предполагает успешное освоение студентами дисциплин магистратуры «Технологии анализа больших данных» (входит в Базовую часть), «Большие данные и теория вычислительного эксперимента».

Материалы дисциплины востребованы при изучении курса «Анализ данных» и при выполнении выпускной квалификационной работы.

3 Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 3 зачетных(е) единиц(ы) (108часов).

3.1 Виды учебной работы и трудоемкость (по формам обучения)

3.1.1 Очная форма обучения

№ п/п	Вид учебной работы	Количество часов	Семестры	
			2	
1	**Аудиторные занятия**	**18**		
	В том числе:			
1.1	Лекции	12	12	
1.2	Семинарские/практические занятия			
1.3	Лабораторные занятия	24	24	
2	**Самостоятельная работа**	**72**	72	
3	**Промежуточная аттестация**		экзамен	
	Итого:	**108**		

3.1.2 Очно-заочная форма обучения

3.1.3 Заочная форма обучения

3.2 Тематический план изучения дисциплины (по формам обучения)

3.2.1 Очная форма обучения

№ п/п	Разделы/темы дисциплины	Трудоемкость, час					
		Всего	Аудиторная работа				Самостоятельная работа
			Лекции	Семинарские/ практические занятия	Лабораторные занятия	Практическая подготовка	
1	Лекция Л-1. Задача понижения размерности данных	27	3		6		18
2	Лекция Л-2. Метод главных компонент (PCA) и родственные с ним методы	27	3		6		18
3	Лекция Л-3. Реализация метода главных компонент с использованием нейронных сетей	27	3		6		18
4	Лекция Л-4. Методы кластеризации	27	3		6		18
Итого		**108**	**12**		**24**		**72**

3.3 Содержание дисциплины

Л-1	Задача понижения размерности данных	3 ак. часа

Краткое содержание (перечень рассматриваемых вопросов) лекции:

- Проблемы, возникающие при анализе данных большого размера.
- Два режима сокращения данных.
- Требования к алгоритмам снижения размерности данных.

Контрольные вопросы:

1. Что такое «большие данные?»
2. Чем большие данные отличаются от традиционных?
3. В чем сходство между большими данными и традиционными данными?
4. Риски, связанные с большими данными?
5. Где большие данные можно использовать?
6. Какую структуру имеют ваши большие данные?
7. Для чего нужна эффективная фильтрация больших данных?
8. Что позволяют обнаружить Веб-данные?
9. Что такое сегментация клиентов?
10. Что такое телематические данные и как их можно использовать?
11. Что такое интеллектуальные сети и как их данные можно использовать?
12. Промышленные двигатели и оборудование. Как использовать данные, полученные от датчиков?

13. Какие данные можно получить из социальных сетей?
14. Чем отличается традиционная и современная архитектура аналитической системы?
15. Что такое массово-параллельные системы обработки?
16. Что такое облачные вычисления?
17. Модель MapReduce?
18. Что такое аналитическая песочница, ее определение и сфера применения?
19. Преимущества аналитической песочницы.
20. Что такое аналитический набор данных?
21. Что входит в аналитический набор данных предприятия?
22. Сводные таблицы или представления?
23. Встроенный скоринг.
24. Какая эволюция произошла в аналитических методах?
25. Принцип работы групповых методов.
26. Что такое экспресс-модель?
27. Какие имеются способы применения экспресс-моделей?
28. Как производится анализ неструктурированных данных?
29. Как происходило распространение графических пользовательских интерфейсов?
30. Почему набирают популярность точечные решения?
31. Современные средства визуализации
32. В чем важность визуализации для передовой аналитики?
33. Чем отличается анализ и отчетность?
34. Какие критерии хорошего анализа?
35. Чем отличается базовая аналитика от углубленной аналитики?
36. Что такое статистическая значимость?
37. Кто такой профессионал в области аналитики?
38. Что такое достаточно чистые данные?
39. Почему необходимо выбрать нужный уровень детализации?
40. Нужна ли аналитику сертификация?
41. Какие способы сокращения данных вы знаете?
42. Перечислите требования к алгоритмам снижения размерности данных.

Л-2	Метод главных компонент (PCA) и родственные с ним методы	3 ак. часов

Краткое содержание (перечень рассматриваемых вопросов) лекции:

- Основные этапы метода.
- Нахождение лучшего пространства меньшей размерности.
- Метод главных компонент через критерий минимизации ошибки проектирования на гиперплоскость.
- Метод главных компонент как метод максимизации разброса данных.
- Оценка погрешности метода главных компонент.
- Понятие о вероятностном методе главных компонент.
- Ядровой (нелинейный) метод главных компонент.
- Понятие о методе независимых компонент (ICA).
- Программная реализация метода главных компонент.

Контрольные вопросы:

43. Для чего нужно исследовать данные для применения метода PCA?
44. Как подготовить данные к использованию метода?
45. Что такое ковариационная матрица?
46. Собственные вектора и значения (айгенпары)?
47. Как используется снижение размерности (проекция) в PCA
48. Как производится восстановление данных?
49. Что такое геометрическая интерпретация метода главных компонент?
50. Как рассчитываются погрешности метода главных компонент?

51. Как сформулировать вероятностную модель для метода главных компонент
52. Как осуществляется выбор числа главных компонент?
53. Что такое ядровая функция?
54. На каких допущениях и эффектах базируется метод независимых компонентах?

Л-3	Реализация метода главных компонент с использованием нейронных сетей	3 ак. часа

Краткое содержание (перечень рассматриваемых вопросов) лекции:

- Алгоритм обучения нейронной сети без учителя.
- Правило Хебба и обобщенное правило Хебба для адаптации весов нейронов.
- Обобщенный алгоритм Хебба для анализа главных компонент.
- Оценка погрешности обобщенного алгоритма Хебба для нейронной сети.

Контрольные вопросы:

55. Что значит понятие «обучения без учителя» в теории распознавания образов?
56. Какие типы входных данных используются в алгоритмах обучения нейросети без учителя?
57. Какие задачи можно решать с помощью алгоритмов обучения нейросети без учителя?
58. Какие функции активации нейронов используются в нейронных сетях Хебба?
59. Приведите принципиальную схему нейронной сети Хебба
60. Как корректируются веса связей нейронов в сети Хебба?
61. Для каких целей осуществляется сокращение размерности в обобщенном алгоритме Хебба?
62. Как определяется погрешность алгоритма Хебба?

Л-4	Методы кластеризации	3 ак. часа

Краткое содержание (перечень рассматриваемых вопросов) лекции:

- Постановка задачи кластеризации.
- Некорректность задачи кластеризации.
- Цели кластеризации.
- Типы кластерных структур.
- Методы кластеризации, использующие графы.
- Иерархическая кластеризация.
- Статистические методы кластеризации.
- Визуализация кластерной структуры.
- Классификация при помощи самоорганизующихся карт Кохонена.

Контрольные вопросы:

63. Для каких целей производится кластеризация данных?
64. Какие методы кластеризации вы знаете?
65. Какие типы входных данных используются при класстеризации?
66. Что такое функция расстояния?
67. Что такое алгоритм кластеризации?
68. Структура сети Кохонена?
69. Алгоритм, реализующий сеть Кохонена?
70. Особенности модели сети Кохонена?

Раскрывается содержание темы

…

3.4 Тематика семинарских/практических и лабораторных занятий

Лабораторная работа 1. Задача понижения размерности данных

Лабораторная работа 2. Метод главных компонент (PCA) и родственные с ним методы
Лабораторная работа 3. Реализация метода главных компонент с использованием нейронных сетей
Лабораторная работа 4. Методы кластеризации

3.5 Тематика курсовых проектов (курсовых работ)

Не предусмотрено

4 Учебно-методическое и информационное обеспечение

4.1 Нормативные документы и ГОСТы

1.Федеральный государственный образовательный стандарт высшего образования (уровень магистратуры) по направлению подготовки 01.04.02 Прикладная математика и информатика, утвержденный приказом Министерства науки и высшего образования Российской Федерации от 10.01.2018 №13.

2. Приказ Минобрнауки России от 09.02.2016 N 86 "О внесении изменений в Порядок проведения государственной итоговой аттестации по образовательным программам высшего образования - программам бакалавриата, программам специалитета и программам магистратуры, утвержденный приказом Министерства бразования и науки Российской Федерации от 29 июня 2015 г. N636"(Зарегистрировано в Минюсте России 02.03.2016 N 41296).

3. Приказ ректора Московского политехнического университета от 01.09.2016 No 128-ОД о введение в действие положения о порядке проведения государственной итоговой аттестации федерального государственного бюджетного образовательного учреждения высшего образования «Московский политехнический университет».

…

4.2 Основная литература

1. Фрэнкс, Б. Укрощение больших данных: как извлекать знания из массивов информации с помощью глубокой аналитики. [Электронный ресурс] — Электрон. дан. — М. : Манн, Иванов и Фербер, 2014. — 352 с. — Режим доступа: http://e.lanbook.com/book/62154

2. Петросов А. А. «Стратегическое планирование и прогнозирование». учебное пособие [Электронный ресурс] Московский государственный горный университет 2001 г — Режим доступа: http://www.knigafund.ru/books/177036

…

4.3 Дополнительная литература

1. Чубукова И. А. « Data Mining». [Электронный ресурс] Издательство: Интернет-Университет Информационных Технологий, 2008 г. –Режим доступа: http://www.knigafund.ru/books/177992

2.

…

4.4 Электронные образовательные ресурсы

1. https://online.mospolytech.ru/course/view.php?id=12081

4.5 Лицензионное и свободно распространяемое программное обеспечение

1. Операционная система, Windows 11 (или ниже) - Microsoft Open License
2. Офисные приложения, Microsoft Office 2013(или ниже) - Microsoft Open License

…

4.6 Современные профессиональные базы данных и информационные справочные системы

1.
2.

…

Ссылки на ресурсы должны содержать актуальный электронный адрес и быть доступными для перехода с любого компьютера.

5 Материально-техническое обеспечение

5.1 Требования к оборудованию и помещению для занятий

Лабораторные работы и самостоятельная работа студентов должны проводиться в специализированной аудитории, оснащенной современной оргтехникой и персональными компьютерами с программным обеспечением в соответствии с тематикой изучаемого материала. Число рабочих мест в аудитории должно быть достаточным для обеспечения индивидуальной работы студентов. Рабочее место преподавателя должно быть оснащены современным компьютером с подключенным к нему проектором на настенный экран, или иным аналогичным по функциональному назначению оборудованием.

Лекционные занятия должны проводиться в специализированных аудиториях с комплектом мультимедийного оборудования и/или доской для записей материалов. Число рабочих мест в аудитории должно быть достаточным для обеспечения индивидуальной работы студентов.

5.2 Требования к программному обеспечению

Для выполнения лабораторных работ и самостоятельной работы необходимо следующее программное обеспечение:

1. Microsoft windows.
2. Офисные приложения, Microsoft Office.
3. Веб-браузер, Chrome.
4. Microsoft Visio

Для проведения лекционных занятий специального программного обеспечения для освоения дисциплины не требуется.

6 Методические рекомендации

6.1 Методические рекомендации для преподавателя по организации обучения

1. При подготовке к занятиям следует предварительно проработать материал занятия, предусмотрев его подачу точно в отведенное для этого время занятия. Следует подготовить необходимые материалы – теоретические сведения, задачи и др. При проведении занятия следует контролировать подачу материала и решение заданий с учетом учебного времени, отведенного для занятия.

2. При проверке работ и отчетов следует учитывать не только правильность выполнения заданий, но и оптимальность выбранных методов решения, правильность выполнения всех его шагов.

3. При организации и проведения экзаменов в практико-ориентированной форме следует использовать утвержденные кафедрой Методические рекомендации.

6.2 Методические указания для обучающихся по освоению дисциплины

Изучение дисциплины осуществляется в строгом соответствии с целевой установкой в тесной взаимосвязи учебным планом. Основой теоретической подготовки студентов являются аудиторные занятия, лекции.

В процессе самостоятельной работы студенты закрепляют и углубляют знания, полученные во время аудиторных занятий, дорабатывают конспекты и записи, готовятся к проведению и обрабатывают результаты лабораторных работ, готовятся к промежуточной аттестации, а также самостоятельно изучают отдельные темы учебной программы.

На занятиях студентов, в том числе предполагающих практическую деятельность, осуществляется закрепление полученных, в том числе и в процессе самостоятельной работы, знаний. Особое внимание обращается на развитие умений и навыков установления связи положений теории с профессиональной деятельностью будущего специалиста.

Самостоятельная работа осуществляется индивидуально. Контроль самостоятельной работы организуется в двух формах:

- самоконтроль и самооценка студента;
- контроль со стороны преподавателей (текущий и промежуточный).

Критериями оценки результатов самостоятельной работы студента являются:

- уровень освоения студентом учебного материала;
- умения студента использовать теоретические знания при выполнении практических задач;
- сформированность компетенций;
- оформление материала в соответствии с требованиями.

7 Фонд оценочных средств

7.1 Методы контроля и оценивания результатов обучения

В процессе обучения используются следующие оценочные формы самостоятельной работы студентов, оценочные средства текущего контроля успеваемости и промежуточных аттестаций:

- Во втором семестре изучения дисциплины: проверка домашних заданий, экзамен.

7.2 Шкала и критерии оценивания результатов обучения

Промежуточная аттестация обучающихся в форме экзамена проводится по результатам выполнения всех видов учебной работы, предусмотренных учебным планом по данной дисциплине (модулю), при этом учитываются результаты текущего контроля успеваемости в течение семестра. Оценка степени достижения обучающимися планируемых результатов обучения по дисциплине (модулю) проводится преподавателем, ведущим занятия по дисциплине (модулю) методом экспертной оценки. По итогам промежуточной аттестации по дисциплине выставляется оценка «отлично», «хорошо», «удовлетворительно» или «неудовлетворительно».

Шкала оценивания результатов промежуточной аттестации определена в п 5.6 «Положении о проведении текущего контроля успеваемости и промежуточной аттестации обучающихся федерального государственного бюджетного образовательного учреждения высшего образования «Московский политехнический университет», утвержденным приказом ректора Московского политехнического университета от 31.08.2017 № 843-ОД. В случае внесения изменений в документ или утверждения нового Положения, следует учитывать принятые правки.

К промежуточной аттестации допускаются только студенты, выполнившие все виды учебной работы, предусмотренные рабочей программой дисциплины. При этом используется балльно-рейтинговая система, включающая следующие критерии оценки.

Критерий	**Значение критерия**
Выполнение экзаменационного задания	Максимальное значение критерия – 100 баллов.

Максимальная сумма набираемых по дисциплине баллов – 100. С началом каждого нового семестра изучения дисциплины набранные баллы обнуляются и рейтинг студента ведется заново. Перевод набранных баллов в оценку промежуточной аттестации производится согласно следующей таблице.

Оценка по балльно-рейтинговой системе	**Оценка по итоговой аттестации**
0 … 49	Неудовлетворительно
50 … 59	Удовлетворительно
60 … 75	Хорошо
76 … 100	Отлично

7.3 Оценочные средства

7.3.1 Текущий контроль

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

(МОСКОВСКИЙ ПОЛИТЕХ)

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №1
по дисциплине
«**Технологии анализа Big data**»
направление подготовки 01.04.02 Прикладная математика и информатика

ВОПРОСЫ:

1. Методы сегментации клиентов.
2. Ядровой (нелинейный) метод главных компонент.
3. Методы кластеризации, использующие графы
4. Реализовать обучение нейронной сети с двумя нейронами по правилу Хебба для задачи классификации. Привести исходные данные и результаты обучения сети в табл. 1 (таблица прилагается)

Утверждено: _____ / _____ / «__» _____ 20__г.

№	Входные данные	Исходные данные		Нормированные значения	
		Потребляемая мощность, Вт x_1	Цена, руб. x_2	Потребляемая мощность, Вт x_1	Цена, руб. x_2
1	LG MF-6543AFS	2350	10490	0,49	0,28
2	Daewoo KOR-6L65	1800	4990	0,37	0,14
3	Bosch HMT 75G421	1270	6690	0,26	0,18
4	BBK 20MWS-709M/B	1050	4290	0,22	0,12
5	Panasonic NN-DS592M	1820	25590	0,38	0,69
6	Midea AG823A3V	1250	6490	0,26	0,18
7	Rolsen MS1770MU	1150	4290	0,24	0,12
8	Samsung ME81KRW-1/BWT	1550	5790	0,32	0,16

Типовые практические задания

1. Реализовать задачу понижения размерности;

2. Использовать метод главных компонент для выявления предпочтений.
3. Реализовать обучение нейросети с использованием правила Хебба
4. Организовать самоорганизующуюся карту Кохонена .