

Минобрнауки России

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«Оренбургский государственный университет»

Кафедра компьютерной безопасности и математического обеспечения информационных систем

РАБОЧАЯ ПРОГРАММА

ДИСЦИПЛИНЫ

«Б1.Д.Б.13.7 Байесова статистика»

Уровень высшего образования

БАКАЛАВРИАТ

Направление подготовки

01.03.05 Статистика

(код и наименование направления подготовки)

Статистика и управление данными

(наименование направленности (профиля) образовательной программы)

Квалификация

Бакалавр

Форма обучения

Очная

Год набора 2024

Рабочая программа дисциплины «Б1.Д.Б.13.7 Байесова статистика» рассмотрена и утверждена на заседании кафедры

Кафедра компьютерной безопасности и математического обеспечения информационных систем

наименование кафедры

протокол № 8 от "25" марта 2024 г.

Заведующий кафедрой

Кафедра компьютерной безопасности и математического обеспечения информационных систем

наименование кафедры

подпись

И.В. Влацкая

расшифровка подписи

Исполнители:

доцент

должность

подпись

И.В. Влацкая

расшифровка подписи

должность

подпись

расшифровка подписи

СОГЛАСОВАНО:

Председатель методической комиссии по направлению подготовки

01.03.05 Статистика

код наименование

личная подпись

В.Н. Афанасьев

расшифровка подписи

/ Заведующий отделом формирования фонда и научной обработки документов

личная подпись

Н.Н. Бигалиева

расшифровка подписи

Уполномоченный по качеству факультета

личная подпись

И.В. Крючкова

расшифровка подписи

№ регистрации _____

© Влацкая И.В., 2024

© ОГУ, 2024

1 Цели и задачи освоения дисциплины

Цель освоения дисциплины:

освоение байесовского подхода к теории вероятностей как одного из последовательных способов математических рассуждений в условиях неопределенности.

Задачи:

изучение байесовского подхода и его теоретического обоснования и практическое применение байесовского подхода в задачах анализа данных.

2 Место дисциплины в структуре образовательной программы

Дисциплина относится к базовой части блока Д «Дисциплины (модули)»

Пререквизиты дисциплины: *Б1.Д.Б.12.3 Численные методы, Б1.Д.Б.13.2 Математическая статистика*

Постреквизиты дисциплины: *Б2.П.В.П.2 Научно-аналитическая работа*

3 Требования к результатам обучения по дисциплине

Процесс изучения дисциплины направлен на формирование следующих результатов обучения

Код и наименование формируемых компетенций	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
ОПК-3 Способен осознанно применять методы математической и дескриптивной статистики для анализа количественных данных, в том числе с применением необходимой вычислительной техники и стандартных компьютерных программ, содержательно интерпретировать полученные результаты, готовить статистические материалы для докладов, публикаций и других аналитических материалов	ОПК-3-В-1 Применяет современный статистический и математический инструментарий для решения профессиональных задач ОПК-3-В-2 Владеет навыками использования информационно-коммуникационных технологий и программных средств для анализа количественных данных	<u>Знать:</u> – основные задачи и области применения методов локализации пространственных данных; – особенности объектов моделирования и методики экспериментальной оценки их свойств; <u>Уметь:</u> работать с прикладным программным обеспечением общего, специального и профессионального назначения. <u>Владеть:</u> методами и средствами поиска, сбора, обработки, хранения, представления, распространения информации.

4 Структура и содержание дисциплины

4.1 Структура дисциплины

Общая трудоемкость дисциплины составляет 3 зачетные единицы (108 академических часов).

Вид работы	Трудоемкость, академических часов	
	7 семестр	всего
Общая трудоёмкость	108	108
Контактная работа:	34,25	34,25
Лекции (Л)	18	18
Лабораторные работы (ЛР)	16	16
Промежуточная аттестация (зачет, экзамен)	0,25	0,25
Самостоятельная работа: - самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий); - изучение разделов курса в системе электронного обучения; - подготовка к лабораторным занятиям; - подготовка к рубежному контролю и т.п.)	73,75	73,75
Вид итогового контроля (зачет, экзамен, дифференцированный зачет)	диф. зач.	

Разделы дисциплины, изучаемые в 7 семестре

№ раздела	Наименование разделов	Количество часов				
		всего	аудиторная работа			внеауд. работа
			Л	ПЗ	ЛР	
1	Параметрическая Байесовская статистика	26	4		4	18
2	Использование Байесовской статистики в машинном обучении	30	6		4	20
3	Непараметрическая Байесовская статистика	26	4		4	18
4	Регрессия на основе Гауссовских процессов	26	4		4	18
	Итого:	108	18		16	74
	Всего:	108	18		16	74

4.2 Содержание разделов дисциплины

1. **Байесовский подход к теории вероятностей.** Введение. Частотный и байесовский подходы к теории вероятностей. Примеры байесовских рассуждений.

2. **Полный байесовский вывод.** Сопряжённые распределения. Примеры. Экспоненциальный класс распределений, его свойства.

3. **Байесовский выбор модели.** Принцип наибольшей обоснованности. Интерпретация понятия обоснованности, ее геометрический смысл, иллюстративные примеры, связь с принципом Оккама.

4. **Модель релевантных векторов для задачи регрессии.** Обобщенные линейные модели, вероятностная модель линейной регрессии. Метод релевантных векторов, вывод формул для регрессии. Свойства решающего правила. Матричные вычисления и нормальное распределение.

5. **Модель релевантных векторов для задачи классификации.** Логистическая и мультиномиальная регрессия. Метод релевантных векторов для задачи классификации. Приближение Лапласа для оценки обоснованности в случае задачи классификации, его достоинства и недостатки.

6. **ЕМ-алгоритм.** ЕМ-алгоритм в общем виде. ЕМ-алгоритм как покоординатный подъем. ЕМ-

алгоритм для задачи разделения смеси нормальных распределений. Байесовский метод главных компонент.

7. Вариационный подход. Приближенные методы байесовского вывода. Минимизация дивергенции Кульбака-Лейблера и факторизованное приближение. Идея вариационного подхода, вывод формул для вариационной смеси нормальных распределений.

8. Методы Монте Карло по схеме марковских цепей (МСМС). Методы генерации выборки из одномерных распределений. Методы МСМС для оценки статистик вероятностных распределений. Теоретические свойства марковских цепей. Схема Метрополиса-Хастингса и схема Гиббса. Примеры использования.

9. Стохастические методы МСМС. Продвинутое методы сэмплирования, использующие градиент лог-правдоподобия. Динамика Гамильтона и Ланжевена. Масштабируемые обобщения этих методов.

10. Гауссовские процессы для регрессии и классификации. Гауссовские случайные процессы. Выбор наиболее адекватной ковариационной функции.

11. Непараметрические байесовские методы. Процессы Дирихле. Распределение Дирихле. Свойства накопления и нейтральности. Генерация выборки из Дирихле через гамма-распределения и через stick-breaking. Представление процесса Дирихле с помощью процесса китайского ресторана. Схема Гиббса для разделения смеси распределений с процессом Дирихле. Представление процесса Дирихле с помощью stick-breaking. Вариационный вывод для разделения смеси распределений с процессом Дирихле.

4.3 Лабораторные работы

№ ЛР	№ раздела	Наименование лабораторных работ	Кол-во часов
1, 2	1	Построением атематических моделей на основе байесовского подхода.	4
3, 4	2	Построение линейной регрессии	4
5, 6	3	Метод главных компонент	4
7, 8	4	Марковские цепи	4
		Итого:	16

5 Учебно-методическое обеспечение дисциплины

5.1 Основная литература

1. Математическая статистика [Текст]: учебник для вузов / Г.А. Соколов, И.М. Гладких.- М. : Экзамен, 2004. - 432 с. - Библиогр.: с .429-431.- ISBN 5-94692-982-8.

2. Лисьев В. П. Теория вероятностей и математическая статистика. Учебное пособие [Электронный ресурс] / Лисьев В. П. - Евразийский открытый институт, 2010. Режим доступа: <http://biblioclub.ru/index.php?page=book&id=90420>

5.2 Дополнительная литература

1. Айвазян, С. А. Прикладная статистика. Основа эконометрики : в 2 т.: учеб. для вузов / С. А. Айвазян, В. С. Мхитарян . - М. : ЮНИТИ-ДАНА, 2001.. - ISBN 5-238-00304-8. Т. 1 : Теория вероятностей и прикладная статистика. - , 2001. - 656 с.

2. Теория статистики [Текст] : учебник / под ред. Г. Л. Громыко.- 2-е изд., перераб. и доп. - М. : ИНФРА-М, 2006. - 476 с.

5.3 Периодические издания

1. Журнал «Информационные технологии».

2. Журнал «Открытые системы. СУБД».
3. Журнал «Программная инженерия».

5.4 Интернет-ресурсы

1. Open Machine Learning Course mlcourse.ai. [Электронный ресурс]. – Режим доступа: <https://mlcourse.ai/>
2. Kaggle [Электронный ресурс]. – Режим доступа: <https://www.kaggle.com/>
3. Введение в машинное обучение [Электронный ресурс]. – Режим доступа: <https://www.coursera.org/learn/vvedenie-mashinnoe-obuchenie>

5.5 Программное обеспечение, профессиональные базы данных и информационные справочные системы

1. *Операционная система Microsoft Windows текущей версии. Доступна в рамках подписки Microsoft DreamSpark Premium. Разработчик: компания Microsoft. Режим доступа: https://e5.onthefhub.com/WebStore/ProductsByMajorVersionList.aspx?cmi_mnuMain=bdba23cf-e05e-e011-971f-0030487d8897&ws=58727022-4bac-e211-88b7-f04da23e67f4&vsro=8*
2. *Офисный пакет Microsoft Office (Word, Excel, Power Point) текущей версии. Доступен в рамках лицензионного соглашения OVS-ES. Разработчик: компания Microsoft. Режим доступа: <https://products.office.com/en/home>*

6 Материально-техническое обеспечение дисциплины

Учебные аудитории для проведения занятий лекционного типа, семинарского типа, курсового проектирования, для проведения групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации.

Аудитории оснащены комплектами ученической мебели, техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Для проведения лабораторных занятий используется компьютерный класс, оснащенный персональными компьютерами с установленным программным обеспечением.

Помещение для самостоятельной работы обучающихся оснащены компьютерной техникой, подключенной к сети "Интернет", и обеспечением доступа в электронную информационно-образовательную среду ОГУ.

К рабочей программе прилагаются:

- Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине;
- Методические указания для обучающихся по освоению дисциплины.