

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Российский государственный гуманитарный университет»
(ФГБОУ ВО «РГГУ»)

ОТДЕЛЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ В ГУМАНИТАРНОЙ СФЕРЕ
Кафедра математики, логики и интеллектуальных систем в гуманитарной сфере

БАЙЕСОВСКИЕ МЕТОДЫ В СТАТИСТИКЕ И МАШИННОМ ОБУЧЕНИИ

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

45.04.04 – Интеллектуальные системы в гуманитарной среде

Когнитивное и программное обеспечение интеллектуальных роботов и программирование
интеллектуальных систем

Уровень высшего образования: магистратура

Форма обучения: очная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2024

Байесовские методы в статистике и машинном обучении
Рабочая программа дисциплины

Составитель(и):

Кандидат технических наук, доцент Л.О. Шашкин

.....

УТВЕРЖДЕНО

Протокол заседания кафедры МЛИИС

№6 от 08.02.2024

ОГЛАВЛЕНИЕ

1. Пояснительная записка	5
1.1 Цель и задачи дисциплины	5
1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций	5
1.3. Место дисциплины в структуре образовательной программы	6
2. Структура дисциплины	6
3. Содержание дисциплины	7
4. Образовательные технологии	8
5. Оценка планируемых результатов обучения	9
5.1 Система оценивания	9
5.2 Критерии выставления оценки по дисциплине	10
5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине	11
6. Учебно-методическое и информационное обеспечение дисциплины	12
6.1 Список источников и литературы	12
6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».	12
7. Материально-техническое обеспечение дисциплины	12
8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов	12
9. Методические материалы	13
9.1 Планы семинарских/ практических/ лабораторных занятий	13
9.2 Методические рекомендации по подготовке письменных работ	18
Приложение 1. Аннотация рабочей программы дисциплины	21

1. Пояснительная записка

1.1 Цель и задачи дисциплины

Цель дисциплины. В процессе обучения байесовскому машинному обучению преследуются несколько целей. Одна из них – подготовить выпускника, способного внедрить алгоритмы машинного обучения в интеллектуальных системах. Другой целью курса можно считать научение слушателей применять технику, основанную на теореме Байеса, для разработки более робастных алгоритмов интеллектуального анализа данных.

Задачи дисциплины: освоение байесовского подхода к статистике и машинному обучению и навыков, необходимых для получения требуемых компетенций в области создания интеллектуальных систем с машинным обучением.

1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Компетенция (код и наименование)	Индикаторы компетенций (код и наименование)	Результаты обучения
<i>УК-1</i> Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	<i>УК-1.1</i> Знает принципы сбора, отбора и обобщения информации <i>УК-1.2</i> Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности <i>УК-1.3</i> Имеет практический опыт работы с информационными объектами и сетью Интернет, опыт библиографического разыскания, создания научных текстов	Знать: <ul style="list-style-type: none">● математические основы байесовского подхода к статистике и машинному обучению;● байесовские модели точечного оценивания и регрессии; Уметь: <ul style="list-style-type: none">● применять методы байесовской статистики для интеллектуального анализа данных;● разрабатывать робастные версии алгоритмов на основе байесовского подхода. Владеть: <ul style="list-style-type: none">● простейшими навыками применения алгоритмов байесовского машинного обучения для переработки больших объемов информации.
<i>ОПК-1</i> Способен применять в профессиональной деятельности методы математического анализа, логики и моделирования, теоретического и экспериментального исследования в информатике, лингвистике	<i>ОПК-1.1</i> Знает основы математического анализа, логики и математического моделирования <i>ОПК-1.2</i> Умеет использовать математические методы для построения моделей в информатике, лингвистике и некоторых	

гуманитарных наук	гуманитарных дисциплинах <i>ОПК-1.3</i> Владеет методами теоретического и экспериментального исследования в информатике, лингвистике и гуманитарных науках	
<i>ОПК-2</i> Способен выявлять сущность проблем, возникающих в ходе профессиональной деятельности, привлекать соответствующий математический аппарат и информационные технологии для их решения	<i>ОПК-2.1</i> Знает примеры решения разных классов задач, возникающих в профессиональной деятельности <i>ОПК-2.2</i> Умеет использовать инструментальные средства для построения математических моделей, используемых для решения прикладных задач из разных классов <i>ОПК-2.3</i> Имеет практический опыт участия в решении практических задач, предполагающий использование математического аппарата и информационных технологий	

1.3. Место дисциплины в структуре образовательной программы

Дисциплина «Статистическая теория машинного обучения» относится к обязательной части блока дисциплин учебного плана.

Для освоения дисциплины необходимы компетенции, сформированные (в рамках бакалавриата) в ходе изучения следующих дисциплин: «Дискретная математика», «Теория алгоритмов», «Теория вероятностей и математическая статистика».

2. Структура дисциплины¹

Общая трудоёмкость дисциплины составляет 3 з.е., 108 академических часа (ов).

Структура дисциплины для очной формы обучения

¹ При реализации образовательной программы на очно-заочной и заочной формах обучения, таблица составляется для каждой формы.

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Семестр	Тип учебных занятий	Количество часов
3	Лекции	12
	Практические занятия	18
	Всего:	30

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 78 академических часа(ов).

3. Содержание дисциплины²

В курсе изучаются понятия байесовского подхода к статистике и машинному обучению и его применения в интеллектуальных системах. На практических занятиях студенты приобретают практические навыки обработки небольших массивов данных, навыки анализа алгоритмов машинного обучения с использованием байесовского подхода.

В результате изучения курса студенты должны овладеть основными идеями и понятиями байесовского машинного обучения, моделями точечного оценивания и регрессии в байесовской статистике и машинном обучении, уметь использовать их при решении практических задач интеллектуального анализа данных.

Курс должен сочетать современность и строгость изложения материала с его доступностью для слушателей. В основе курса лежит обсуждение ключевых теорем и методов применения байесовского машинного обучения для решения прикладных задач.

№ п/п	Наименование раздела дисциплины	Содержание
1.	Обзор теории вероятностей	Случайные величины. Среднее с.в. Примеры с.в. Сходимость по вероятности. Сходимость почти наверное. Полнота пространства с.в. Сходимость по распределению. Монотонная сходимость средних. Лемма Фату. Доминируемая сходимость средних. Пространства интегрируемых с.в. Неравенство Йенсена. Полнота пространства L_1 . Борелевские функции от с.в. Теорема факторизации. Квадратично-интегрируемые с.в. Сходимость в среднем-квадратичном.
2.	Условные средние и мартингалы	Ортогональное подпространство и проекции. Свойства оператора проецирования. Теорема Пифагора. Условное среднее как проекция. Расширение на пространство L_1 . Свойства условных средних. Неравенство Дуба. Мартингалы. Примеры мартингалов. Суб- и супермартингалы. Марковские моменты остановки. Неравенства для суб- и супермартингалов. Сходимость супермартингалов.
3.	Байесовское обучение для экспоненциального семейства	Теорема Байеса. Дизынтегрирование. Плотность распределения. Совместная и условная плотности. Теорема Байеса для плотностей. Виды априорных плотностей. Экспоненциальный класс распределений с.в. Моменты достаточной статистики. Семейство бета-

² Раздел может быть представлен как в текстовой форме, так и в таблице

		распределений. Семейство гамма-распределений. Семейство нормальных распределений. Сопряженное семейство. Параметры апостериорного распределения. Лемма Скорохода. Байесовское обучение. Пример байесовского обучения для пуассоновского потока. Состоятельность байесовского обучения в этом случае.
4.	Классификация и точечное оценивание	Задача классификации. Байесовский классификатор. Оптимальность байесовского классификатора. Логистическая функция. Линейный дискриминатор Фишера. Байесовское оценивание и предсказание. Пример байесовского оценивания. Пример байесовского предсказания. Оценка максимума апостериорной вероятности. Устранение переобучения. Логистическая регрессия. Метод Ньютона-Рафсона для логистической регрессии.
5.	Байесовский взгляд на регрессию	Модель логистической релевантной регрессии. Логарифм обоснованности модели. Формулы логистической релевантной регрессии. Оптимальность байесовской регрессии. Модель линейной регрессии (напоминание). Обобщенная линейная регрессия. Ридж-регрессия. Линейная регрессия для релевантных векторов. Тождество Вудбери. Лемма об определителе матрицы. Явный вид обоснованности модели. Дифференцирование квадратичной формы. Вариационные нижние оценки. Опорное семейство для обоснованности. Формулы пересчета весов.
6.	Состоятельность байесовского обучения	Допущение отделимости и измеримости. Метрики и меры на произведениях пространств. Измеримое восстановление параметров. Равномерная интегрируемость. Предел равномерно-интегрируемой последовательности. Равномерно интегрируемые мартингалы. Предельная независимость. Состоятельность байесовской оценки. Состоятельность апостериорного распределения.

4. Образовательные технологии³

Применительно к данной дисциплине, следует отметить, что она, по существу, является введением в прикладную математическую теорию, поэтому применение в ней информационных технологий носит вспомогательный характер. В ней используются следующие информационные технологии.

- Преподаватель представляет изучаемый материал посредством демонстрации презентаций.

- Электронная почта и другие средства Интернет используется студентами для общения с преподавателем.

³ В разделе указываются образовательные технологии, используемые при реализации различных видов учебных занятий для наиболее эффективного освоения дисциплины. При проведении учебных занятий обеспечивается развитие у обучающихся навыков командной работы, межличностной коммуникации, принятия решений, лидерских качеств (включая при необходимости проведение интерактивных лекций, групповых дискуссий, ролевых игр, тренингов, анализ ситуаций и имитационных моделей, преподавание дисциплин (*модулей*) в форме курсов, составленных на основе результатов научных исследований, в том числе с учётом региональных особенностей профессиональной деятельности выпускников и потребностей работодателей (п.34. Приказ №301).

● Сервер учебных материалов отделения используется студентами для создания словаря по курсу, самостоятельной работы и тестирования.

5. Оценка планируемых результатов обучения

5.1 Система оценивания⁴

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
- контрольная работа (темы 1-3)	30 баллов	30 баллов
- контрольная работа (темы 4-6)	30 баллов	30 баллов
Промежуточная аттестация зачет		40 баллов
Итого за семестр (дисциплину)		100 баллов

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

5.2 Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ A,B	отлично/ зачтено	Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации. Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения. Свободно ориентируется в учебной и профессиональной литературе. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».
82-68/	хорошо/	Выставляется обучающемуся, если он знает теоретический и практический

⁴ Система оценивания выстраивается в соответствии с учебным планом, где определены формы промежуточной аттестации (зачёт/зачёт с оценкой/экзамен), и структурой дисциплины, где определены формы текущего контроля. Указывается распределение баллов по формам текущего контроля и промежуточной аттестации, сроки отчётности.

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
С	зачтено	материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей. Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами. Достаточно хорошо ориентируется в учебной и профессиональной литературе. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».
67-50/ D,E	удовлетворительно/ зачтено	Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации. Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами. Демонстрирует достаточный уровень знания учебной литературы по дисциплине. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».
49-0/ F,FX	неудовлетворительно/ не зачтено	Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации. Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами. Демонстрирует фрагментарные знания учебной литературы по дисциплине. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.

5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине⁵

Контрольная работа 1

1. Вычислить апостериорную плотность распределения параметров, если априорная плотность имеет бета-распределение с параметрами $\alpha, \beta > 0$, а наблюдается отрицательно-биномиальная с.в. с параметрами n, θ .
2. Найти каноническую параметризацию распределения Парето $p(x|\alpha, \mu) = (\alpha \cdot \mu^\alpha) / x^{(\alpha+1)}$ при $x > \mu$ в виде экспоненциального семейства.

Контрольная работа 2

1. Рассмотрим априорное распределение $p(\theta) = 1/\theta \cdot (\log U - \log L)$, где $L < \theta < U$. Доказать, что апостериорным будет $p(\theta | y_1, \dots, y_n) = n / (\theta^{(n+1)} \cdot ((xV L)^{-n})$

⁵ Приводятся примеры оценочных средств в соответствии со структурой дисциплины и системой контроля: варианты тестов, тематика письменных работ, примеры экзаменационных билетов, типовые задачи, кейсы и т.п. Оценочными средствами должны быть обеспечены все формы текущего контроля и промежуточной аттестации. Они должны быть ориентированы не только на проверку сформированности знаний, но также умений и владений.

$-U^{(-n)}$), где u_i - независимые равномерно распределенные величины на полуинтервале $(0, \theta]$, $(x \vee L) = \max\{L, u_1, \dots, u_n\}$ и $(x \vee L) < \theta < U$.

2. Пусть из урны, содержащей первоначально α белых и β черных шаров, извлекается на каждом шаге по одному шару, причем после этого в урну добавляется 2 шара того же самого цвета, что и извлеченный на текущем шаге шар. Доказать, что вероятность того, что после n шагов будет извлечено k белых шаров, будет иметь бета-биномиальное распределение.

1 Контрольные вопросы к зачету

1. Теорема Байеса.
2. Плотность распределения. Совместная и условная плотности.
3. Теорема Байеса для плотностей.
4. Виды априорных плотностей.
5. Экспоненциальный класс распределений с.в.
6. Моменты достаточной статистики.
7. Семейство бета-распределений.
8. Семейство гамма-распределений.
9. Семейство нормальных распределений.
10. Сопряженное семейство. Параметры апостериорного распределения.
11. Байесовское обучение. Пример байесовского обучения для пуассоновского потока.
12. Состоятельность байесовского обучения для пуассоновского потока (без доказательства).
13. Оптимальность байесовского классификатора.
14. Логистическая функция. Линейный дискриминатор Фишера.
15. Байесовское оценивание и предсказание.
16. Пример байесовского оценивания.
17. Пример байесовского предсказания.
18. Оценка максимума апостериорной вероятности. Устранение переобучения.
19. Логистическая регрессия. Метод Ньютона-Рафсона для логистической регрессии.
20. Модель логистической релевантной регрессии.
21. Формулы логистической релевантной регрессии.
22. Оптимальность байесовской регрессии.
23. Линейная регрессия для релевантных векторов.
24. Вариационные нижние оценки. Опорное семейство для обоснованности.
25. Формулы пересчета весов линейной регрессии для релевантных векторов.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1 Список источников и литературы⁶

а) Основная литература

1. Ветров Д.П., Кропотов Д.А. Байесовские методы машинного обучения. М.: МГУ, 2007

б) Дополнительная литература

1. Виноградов, Д.В. Байесовские методы в машинном обучении, 2020
<https://vk.com/club174045824>

6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».

Национальная электронная библиотека (НЭБ) www.rusneb.ru

⁶ Рекомендуется включать в списки издания из ЭБС и не более 15 печатных изданий.

ELibrary.ru Научная электронная библиотека www.elibrary.ru
Электронная библиотека Grebennikon.ru www.grebennikon.ru
Cambridge University Press
ProQuest Dissertation & Theses Global
SAGE Journals
Taylor and Francis
JSTOR

7. Материально-техническое обеспечение дисциплины

Для обеспечения дисциплины используется материально-техническая база образовательного учреждения: компьютерный класс, компьютер преподавателя, компьютеры студентов, проектор, экран, доступ в интернет.

Состав программного обеспечения:

1. Windows
2. Microsoft Office

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

● для слепых и слабовидящих: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением или могут быть заменены устным ответом; обеспечивается индивидуальное равномерное освещение не менее 300 люкс; для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств; письменные задания оформляются увеличенным шрифтом; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

● для глухих и слабослышащих: лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования; письменные задания выполняются на компьютере в письменной форме; экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

● для лиц с нарушениями опорно-двигательного аппарата: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями

обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих: в печатной форме увеличенным шрифтом, в форме электронного документа, в форме аудиофайла.

- для глухих и слабослышащих: в печатной форме, в форме электронного документа.

- для обучающихся с нарушениями опорно-двигательного аппарата: в печатной форме, в форме электронного документа, в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих: устройством для сканирования и чтения с камерой SARA CE; дисплеем Брайля PAC Mate 20; принтером Брайля EmBraille ViewPlus;

- для глухих и слабослышащих: автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих; акустический усилитель и колонки;

- для обучающихся с нарушениями опорно-двигательного аппарата: передвижными, регулируемые эргономическими партами СИ-1; компьютерной техникой со специальным программным обеспечением.

9. Методические материалы⁷

9.1 Планы семинарских/ практических/ лабораторных занятий⁸

Тема 1. (2 ч.) Распределения вероятностей и их плотности

Цель занятий: напомнить основные понятия и факты теории вероятностей

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

Что такое случайная величина?

Что такое плотность случайной величины?

Как устроены пространства случайных величин?

Контрольные вопросы:

1. Распределения вероятностей
2. Дискретные случайные величины. Целочисленные с.в.
3. Серии Бернулли. Биномиальная с.в.
4. Отрицательно-биномиальная случайная величина.

⁷ Методические материалы по дисциплине могут входить в состав рабочей программы, либо разрабатываться отдельным документом.

⁸ План занятий строится в соответствии со структурой дисциплины (п.2). Разделы плана включают: название темы, количество часов, форму проведения занятия, его содержание (вопросы для обсуждения, задания, контрольные вопросы, кейсы и т.п.), список литературы. При необходимости, планы практических и лабораторных занятий могут содержать указания по выполнению заданий и требования к материально-техническому обеспечению занятия.

5. Пуассоновская случайная величина.
6. Интеграл Римана-Стилтьеса
7. Интеграл Лебега-Стилтьеса
8. Считаящая мера
9. Плотность распределения вероятностей для целочисленной с.в.
10. Мера Лебега на прямой
11. Нормальное распределение

Список источников и литературы:

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

2. Виноградов, Д.В. Байесовские методы в машинном обучении, 2020
<https://vk.com/club174045824>
3. www.wikipedia.org

Материально-техническое обеспечение занятия: доска, видеопроектор, ноутбук.

Тема 2. (4 ч.) Сопряженные семейства распределений

Цель занятий: изучить понятие сопряженных семейств распределений.

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

- Как применяется теорема Байеса для плотностей?
- Когда два семейства распределений сопряжены?
- Что такое плоское априорное распределение?
- Что такое априорное распределение Джеффриса?

Контрольные вопросы:

12. Гамма-функция Эйлера
13. Семейство гамма-распределений
14. Экспоненциальное распределение
15. Распределение χ^2 .
16. Бета-функция Эйлера
17. Связь бета-функции с гамма-функцией.
18. Бета-распределения
19. Равномерное распределение
20. Плоское априорное распределение.
21. Совместная и условная плотности.
22. Теорема Байеса для плотностей.
23. Априорная и апостериорные плотности и теорема Байеса.
24. Сопряженные семейства.
25. Сопряженность гамма-распределений и распределений Пуассона.
26. Сопряженность бета-распределений и биномиальных распределений.
27. Байесовский вывод для сопряженных семейств.
28. Априорное распределение Джеффриса

Список источников и литературы:

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

4. Виноградов, Д.В. Байесовские методы в машинном обучении, 2020
<https://vk.com/club174045824>
5. www.wikipedia.org

Материально-техническое обеспечение занятия: доска, видеопроектор, ноутбук.

Тема 3. (2 ч.) Экспоненциальные семейства распределений

Цель занятий: усвоить алгоритм машины опорных векторов.

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

Когда семейство распределений является экспоненциальным?

Что такое каноническая параметризация?

Что такое производящая функция моментов?

Как вычислить моменты от достаточной статистики?

Контрольные вопросы:

29. Экспоненциальный класс распределений.

30. Достаточная статистика, репараметризация, начало, статсумма (обоснованность).

31. Каноническая параметризация.

32. Каноническая параметризация гамма-распределений.

33. Каноническая параметризация бета-распределений.

34. Каноническая параметризация нормальных распределений.

35. Производящая функция моментов.

36. Вычисление моментов гамма-распределений.

37. Вычисление моментов бета-распределений.

38. Вычисление моментов нормальных распределений.

39. Класс распределений Парето.

Список источников и литературы:

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

6. Виноградов, Д.В. Байесовские методы в машинном обучении, 2020
<https://vk.com/club174045824>

7. www.wikipedia.org

Материально-техническое обеспечение занятия: доска, видеопроектор, ноутбук.

Тема 4. (4 ч.) Байесовский вывод для семейств экспоненциального класса

Цель занятий: усвоить методологию байесовского вывода.

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

Как устроено байесовское обучение?

Что такое оценка максимума апостериорной вероятности?

Чем оценка максимума апостериорной вероятности лучше, чем оценка максимального правдоподобия?

Контрольные вопросы:

40. Байесовское обучение.

41. Пример байесовского обучения для пуассоновского потока.

42. Состоятельность байесовского обучения для пуассоновского потока (на

картинках).

43. Байесовское оценивание и предсказание.
44. Пример байесовского оценивания для пуассоновского потока.
45. Пример байесовского предсказания для пуассоновского потока.
46. Оценка максимума апостериорной вероятности.
47. Оценка максимума апостериорной вероятности для биномиальной с.в.
48. Оценка максимума апостериорной вероятности для пуассоновской с.в.
49. Феномен переобучения для оценок максимального правдоподобия
50. Устранение переобучения с помощью максимума апостериорной вероятности.

Список источников и литературы:

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

8. Виноградов, Д.В. Байесовские методы в машинном обучении, 2020
<https://vk.com/club174045824>
9. www.wikipedia.org

Материально-техническое обеспечение занятия: доска, видеопроектор, ноутбук.

Тема 5. (2 ч.) Составные распределения

Цель занятий: изучить распределения, возникающие при байесовском предсказании.

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

Что такое составные распределения?

Что будет после байесовского обучения по биномиальной выборке?

Что будет после байесовского обучения по пуассоновской выборке?

Контрольные вопросы:

51. Составные распределения.

52. Бета-биномиальное распределение как результат байесовского обучения по биномиальной выборке.

53. Моменты бета-биномиального распределения.

54. Гамма-пуассоновское распределение как результат байесовского обучения по пуассоновской выборке.

55. Гамма-пуассоновское распределение = отрицательно-биномиальное распределение.

56. Моменты гамма-пуассоновского распределения.

Список источников и литературы:

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

10. Виноградов, Д.В. Байесовские методы в машинном обучении, 2020
<https://vk.com/club174045824>
11. www.wikipedia.org

Материально-техническое обеспечение занятия: доска, видеопроектор, ноутбук.

Тема 6. (4 ч.) Байесовский взгляд на регрессию

Цель занятий: разобраться с вычислениями в регрессии на релевантных векторах.

Форма проведения – обсуждение, решение задач.

Вопросы для обсуждения:

Что такое регрессия на релевантных векторах?

Как дифференцировать квадратичные формы и определители?

Что такое вариационные нижние оценки?

Чем выделяются релевантные вектора?

Контрольные вопросы:

57. Линейная регрессия для релевантных векторов.

58. Тождество Вудбери.

59. Лемма об определителе матрицы.

60. Явный вид обоснованности модели.

61. Дифференцирование квадратичной формы.

62. Вариационные нижние оценки.

63. Опорное семейство для обоснованности.

64. Формулы пересчета весов и релевантные вектора.

Список источников и литературы:

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

12. Виноградов, Д.В. Байесовские методы в машинном обучении, 2020
<https://vk.com/club174045824>

13. www.wikipedia.org

Материально-техническое обеспечение занятия: доска, видеопроектор, ноутбук.

9.2 Методические рекомендации по подготовке письменных работ⁹

Наименование раздела дисциплины	Кол-во часов	Вопросы для изучения	Литература
Обзор теории вероятности	2	Случайные величины. Среднее с.в. Примеры с.в. Сходимость по вероятности. Сходимость почти наверное. Полнота пространства с.в. Сходимость по распределению. Монотонная сходимость средних. Лемма Фату. Доминируемая сходимость средних. Пространства интегрируемых с.в. Неравенство Йенсена. Полнота пространства L^1 . Борелевские функции от с.в. Теорема факторизации. Квадратично-интегрируемые с.в. Сходимость в среднем-	Lecture1.pdf и Lecture2.pdf (начало) из https://vk.com/club174045824

⁹ В раздел включаются требования к подготовке, содержанию, и оформлению письменных работ предусмотренных учебным планом или рабочей программой (курсовая работа, эссе, реферат, доклад и т.п.). При наличии кафедральных или факультетских рекомендаций по подготовке письменных работ раздел включается в РПД по усмотрению преподавателя - составителя. Если письменная работа не предусмотрена, раздел не включается.

		квадратичном.	
Условные средние и мартингалы	2	Ортогональное подпространство и проекции. Свойства оператора проецирования. Теорема Пифагора. Условное среднее как проекция. Расширение на пространство L_1 . Свойства условных средних. Неравенство Дуба. Мартингалы. Примеры мартингалов. Суб- и супермартингалы. Марковские моменты остановки. Неравенства для суб- и супермартингалов. Сходимость супермартингалов.	Lecture2.pdf (конец) и Lecture3.pdf из https://vk.com/club174045824
Байесовское обучение для экспоненциального семейства	2	Теорема Байеса. Дизинтегрирование. Плотность распределения. Совместная и условная плотности. Теорема Байеса для плотностей. Виды априорных плотностей. Экспоненциальный класс распределений с.в. Моменты достаточной статистики. Семейство бета-распределений. Семейство гамма-распределений. Семейство нормальных распределений. Сопряженное семейство. Параметры апостериорного распределения. Лемма Скорохода. Байесовское обучение. Пример байесовского обучения для пуассоновского потока. Состоятельность байесовского обучения в этом случае.	Lecture4.pdf и Lecture5.pdf (начало) из https://vk.com/club174045824
Классификация и точечное оценивание	2	Задача классификации. Байесовский классификатор. Оптимальность байесовского классификатора. Логистическая функция. Линейный дискриминатор Фишера. Байесовское оценивание и предсказание. Пример байесовского оценивания. Пример байесовского предсказания. Оценка максимума апостериорной вероятности. Устранение переобучения. Логистическая регрессия. Метод Ньютона-	Lecture5.pdf (конец) и Lecture6.pdf (начало) из https://vk.com/club174045824

		Рафсона для логистической регрессии.	
Байесовский взгляд на регрессию	2	<p>Модель логистической релевантной регрессии. Логарифм обоснованности модели. Формулы логистической релевантной регрессии. Оптимальность байесовской регрессии. Модель линейной регрессии (напоминание). Обобщенная линейная регрессия. Ридж-регрессия. Линейная регрессия для релевантных векторов. Тождество Вудбери. Лемма об определителе матрицы. Явный вид обоснованности модели. Дифференцирование квадратичной формы. Вариационные нижние оценки. Опорное семейство для обоснованности. Формулы пересчета весов.</p>	Lecture6.pdf (конец) и Lecture7.pdf из https://vk.com/club174045824
Состоятельность байесовского оценивания	2	<p>Допущение отделимости и измеримости. Метрики и меры на произведениях пространств. Измеримое восстановление параметров. Равномерная интегрируемость. Предел равномерно-интегрируемой последовательности. Равномерно интегрируемые мартингалы. Предельная независимость. Состоятельность байесовской оценки. Состоятельность апостериорного распределения.</p>	Lecture8.pdf из https://vk.com/club174045824

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Дисциплина «Байесовские методы в статистике и машинном обучении» реализуется на Отделении интеллектуальных систем в гуманитарной сфере кафедрой математики, логики и интеллектуальных систем.

В процессе обучения байесовскому машинному обучению преследуются несколько целей. Одна из них – подготовить выпускника, способного внедрить алгоритмы машинного обучения в интеллектуальных системах. Другой целью курса можно считать научение слушателей применять технику, основанную на теореме Байеса, для разработки более робастных алгоритмов интеллектуального анализа данных.

Задачи дисциплины: освоение байесовского подхода к статистике и машинному обучению и навыков, необходимых для получения требуемых компетенций в области создания интеллектуальных систем с машинным обучением.

Дисциплина направлена на формирование следующих компетенций:

- *УК-1* Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий;
- *ОПК-1* Способен применять в профессиональной деятельности методы математического анализа, логики и моделирования, теоретического и экспериментального исследования в информатике, лингвистике и гуманитарных науках;
- *ОПК-2* Способен выявлять сущность проблем, возникающих в ходе профессиональной деятельности, привлекать соответствующий математический аппарат и информационные технологии для их решения.

В результате освоения дисциплины обучающийся должен:

Знать:

- байесовский подход к статистике и машинному обучению;
- байесовские модели точечного оценивания и регрессии;
- математические основы байесовского подхода к статистике и машинному обучению

Уметь:

- разрабатывать робастные версии алгоритмов на основе байесовского подхода;
- применять методы байесовской статистики для интеллектуального анализа данных

Владеть:

- простейшими навыками применения алгоритмов байесовского машинного обучения для переработки больших объемов информации;
- простейшими навыками встраивания алгоритмов байесовского машинного обучения в новые информационные технологии

По дисциплине предусмотрена промежуточная аттестация в форме **зачета**.

Общая трудоемкость освоения дисциплины составляет 3 зачетные единицы.