

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 квалификации выпускника: магистр

Форма обучения очная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2019

«Байесовские методы в статистике и машинном обучении»

Рабочая программа дисциплины

Составитель:

доктор физико-математических наук, профессор

Д.В. Виноградов

УТВЕРЖДЕНО

Протокол заседания кафедры МЛиИС

№__7__ от__10.06.2019

ОГЛАВЛЕНИЕ

1. Пояснительная записка

1.1 Цель и задачи дисциплины (*модуля*)

1.2. Формируемые компетенции, соотнесённые с планируемыми результатами обучения по дисциплине

1.3. Место дисциплины в структуре образовательной программы

2. Структура дисциплины

3. Содержание дисциплины

4. Образовательные технологии

5. Оценка планируемых результатов обучения

5.1. Система оценивания

5.2. Критерии выставления оценок

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

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Список источников и литературы

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

7. Материально-техническое обеспечение дисциплины

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

9. Методические материалы

9.1. Планы семинарских занятий

9.2. Методические указания для обучающихся по освоению дисциплины

9.3. Иные материалы

Приложения

Приложение 1. Аннотация дисциплины

Приложение 2. Лист изменений

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. Формируемые компетенции, соотнесённые с планируемыми результатами обучения по дисциплине:

Коды компетенций	Содержание компетенций	Перечень планируемых результатов обучения по дисциплине
ОК-6	способностью находить творческие решения социальных и профессиональных задач	Знать: <ul style="list-style-type: none">байесовский подход к статистике и машинному обучению. Уметь: <ul style="list-style-type: none">разрабатывать робастные версии алгоритмов на основе байесовского подхода.
ОПК-7	готовностью перерабатывать большие объемы информации и вычленять главное (анализ информации)	Знать: <ul style="list-style-type: none">байесовские модели точечного оценивания и регрессии; Владеть: <ul style="list-style-type: none">простейшими навыками применения алгоритмов байесовского машинного обучения для переработки больших объемов информации.
ПК-25	готовностью использовать современные достижения науки и передовой технологии в научно-исследовательских работах	Знать: <ul style="list-style-type: none">математические основы байесовского подхода к статистике и машинному обучению; Уметь: <ul style="list-style-type: none">применять методы байесовской статистики для интеллектуального анализа данных Владеть: <ul style="list-style-type: none">простейшими навыками встраивания алгоритмов байесовского машинного обучения в новые информационные технологии.

1.3. Место дисциплины в структуре образовательной программы

Дисциплина «Статистическая теория машинного обучения» относится к вариативной части дисциплин по выбору 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 дисциплины

Структура дисциплины для очной формы обучения

Общая трудоёмкость дисциплины составляет 3 з.е., 114 ч., в том числе контактная работа обучающихся с преподавателем 30 ч., промежуточная аттестация 18 ч., самостоятельная работа обучающихся 84 ч.

№ п/п	Раздел дисциплины/темы	С е м е с т р	Виды учебной работы (в часах)							Формы текущего контроля успеваемости, форма промежуточной аттестации (<i>по семестрам</i>)
			Контактная				Пром ежуро чная аттест ация		Са мо сто яте ль- на я ра бо та	
			Л е к ц и и	С е м и н а р	Пр акт иче ски е зан яти я	Лаб ора тор ны е зан яти я				
1	Обзор теории вероятностей	1	2			2			14	Оценка выполнения практических заданий
2	Условные средние и мартингалы	1	2			4			14	Оценка выполнения практических заданий
3	Байесовское обучение для экспоненциального семейства	1	2			2			14	Контрольная работа
4	Классификация и точечное оценивание	1	2			4			14	Оценка выполнения практических заданий
5	Байесовский взгляд на регрессию	1	2			2			14	Оценка выполнения практических заданий

6	Состоятельность Байесовского обучения	1	2			4		14	Контрольная работа
	Зачет	1							Итоговая контрольная работа
	итого:		12			18		84	

3. Содержание дисциплины

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-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 технологии

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

5. Оценка планируемых результатов обучения

5.1. Система оценивания

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) зачет		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 (отлично)»/ «зачтено»	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>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.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ C	«хорошо»/ «зачтено (хорошо)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач</p>

		<p>профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,E	«удовлетворительно»/ «зачтено (удовлетворительно)»/ «зачтено»	<p>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.</p> <p>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.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	«неудовлетворительно»/ не зачтено	<p>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.</p> <p>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.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

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

ФОС для проведения промежуточной аттестации по дисциплине состоит из 4 разделов:

1 раздел. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы;

2 раздел. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания;

3 раздел. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы;

4 раздел. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

Для каждого результата обучения по дисциплине (модулю) или практике определяются показатели и критерии оценивания сформированности компетенций на различных этапах их формирования, шкалы и процедуры оценивания.

Оценочные средства, соответствующие конкретным этапам формирования компетенций

Код компетенции	Описание этапов формирования компетенции	Наименование оценочных средств
ОК-6, ОПК-7, ПК-25	Знать: - байесовский подход к статистике и машинному обучению; - байесовские модели точечного оценивания и регрессии; - математические основы байесовского подхода к статистике и машинному обучению	Выполнение практических заданий Зачет
	Уметь: - разрабатывать робастные версии алгоритмов на основе байесовского подхода; - применять методы байесовской статистики для интеллектуального анализа данных	Выполнение практических заданий Контрольная работа 1 Контрольная работа 2 Зачет
	Владеть: - простейшими навыками применения алгоритмов байесовского машинного обучения для переработки больших объемов информации; - простейшими навыками встраивания алгоритмов байесовского машинного обучения в новые информационные технологии	Выполнение практических заданий Контрольная работа 1 Контрольная работа 2 Зачет

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 $p(\theta | y_1, \dots, y_n) = n/(\theta^{(n+1)}) \cdot ((x \vee L)^{-n} - U^{-n})$, где y_i - независимые равномерно распределенные величины на полуинтервале $(0, \theta]$, $(x \vee L) = \max\{L, y_1, \dots, y_n\}$ 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.

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 литературы

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 БД и ИСС

№п /п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2019 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2019 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
3	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам Электронная библиотека Grebennikon.ru
4	Компьютерные справочные правовые системы Консультант Плюс, Гарант

7. Материально-техническое обеспечение дисциплины

Занятия проводятся в компьютерном классе ауд. 706, расположенном по адресу 125993, Москва, Миусская пл., д. 6, стр.2,.

Этот компьютерный класс оснащен

- достаточным количеством объединенных в локальную сеть рабочих станций,
- медиапроектором и экраном,
- меловой доской.

В классе имеются возможности

- подключения ноутбука к медиапроектору,
- одновременного доступа в Интернет для преподавателя и студентов.

1. Перечень ПО

№п /п	Наименование ПО	Производитель	Способ распространения (лицензионное или свободно распространяемое)
1	Adobe Master Collection CS4	Adobe	лицензионное
2	ОС «Альт Образование» 8	ООО «Базальт СПО	лицензионное
3	Windows 10 Pro	Microsoft	лицензионное
4	Kaspersky Endpoint Security	Kaspersky	лицензионное
5	Microsoft Office	Microsoft	лицензионное

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;
 - зачёт проводится в устной форме или выполняется 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;
- зачёт проводится 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;
 - зачёт проводится в устной форме или выполняется 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 материалы

9.1. Планы лабораторных занятий

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.
5. Пуассоновская случайная величина.
6. Интеграл Римана-Стилтьеса
7. Интеграл Лебега-Стилтьеса
8. Считающая 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 указания для обучающихся по освоению дисциплины

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-квадратичном.	Lecture1.pdf и Lecture2.pdf (начало) из https://vk.com/club174045824
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. Дизынтегрирование. 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.	Lecture4.pdf и Lecture5.pdf (начало) из https://vk.com/club174045824

		Лемма Скорохода. Байесовское обучение. Пример байесовского обучения для пуассоновского потока. Состоятельность байесовского обучения в этом случае.	
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-Рафсона для логистической регрессии.	Lecture5.pdf (конец) и Lecture6.pdf (начало) из https://vk.com/club174045824
Байесовский взгляд на регрессию	2	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.	Lecture6.pdf (конец) и Lecture7.pdf из https://vk.com/club174045824
Состоятельность байесовского оценивания	2	Допущение отделимости и измеримости. Метрики и меры на произведениях пространств. Измеримое восстановление параметров. Равномерная интегрируемость. Предел равномерно-интегрируемой	Lecture8.pdf из https://vk.com/club174045824

		<p>последовательности. Равномерно интегрируемые мартингалы. Предельная независимость. Состоятельность байесовской оценки. Состоятельность апостериорного распределения.</p>	
--	--	---	--

9.3 Иные материалы

АННОТАЦИЯ 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:

- ОК-6 способностью находить творческие решения социальных и профессиональных задач;
- ОПК-7 готовностью перерабатывать большие объемы информации и вычленять главное (анализ информации);
- ПК-25 готовностью использовать современные достижения науки и передовой технологии в научно-исследовательских работах.

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.

ЛИСТ ИЗМЕНЕНИЙ

№	Текст актуализации или прилагаемый к РПД документ, содержащий изменения	Дата	№ протокола
1	Приложение №1	08.06.2020	№6

Приложение к листу изменений №1

1. Образовательные технологии (к п.4 на 2020 г.)

В период временного приостановления посещения обучающимися помещений и территории РГГУ. для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:

- аудио-файлы с лекциями из группы <https://vk.com/club174045824> (Виноградов, Д.В. Байесовские методы в машинном обучении, 2020)
- PDF-презентации лекций из группы <https://vk.com/club174045824> (Виноградов, Д.В. Байесовские методы в машинном обучении, 2020)
- PDF-листочки с задачами для лабораторных работ из группы <https://vk.com/club174045824> (Виноградов, Д.В. Байесовские методы в машинном обучении, 2020)
- видео-лекции;
- онлайн-лекции в режиме реального времени;
- электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
- системы для электронного тестирования;
- консультации с использованием телекоммуникационных средств.

2. Перечень БД и ИСС (к п. 6.2 на 2020 г.)

№п/п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2020 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2020 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
3	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам Электронная библиотека Grebennikon.ru
4	Компьютерные справочные правовые системы Консультант Плюс, Гарант

3. Состав программного обеспечения (ПО) (к п. 7 на 2020 г.)

№п /п	Наименование ПО	Производитель	Способ распространения (лицензионное или свободно распространяемое)
1	Adobe Master Collection CS4	Adobe	лицензионное
2	ОС «Альт Образование» 8	ООО «Базальт СПО	лицензионное
3	Windows 10 Pro	Microsoft	лицензионное
4	Kaspersky Endpoint Security	Kaspersky	лицензионное
5	Microsoft Office 2016	Microsoft	лицензионное
6	Zoom	Zoom	лицензионное