

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

УТВЕРЖДЕНО

**Директор физтех-школы
прикладной математики и
информатики**

А.М. Райгородский

	Рабочая программа дисциплины (модуля)
по дисциплине:	Байесовские модели в машинном обучении
по направлению:	Информатика и вычислительная техника
профиль подготовки:	Прикладная математика и информатика Физтех-школа Прикладной Математики и Информатики кафедра проблем передачи информации и анализа данных
курс:	1
квалификация:	магистр

Семестр, формы промежуточной аттестации: 1 (осенний) - Дифференцированный зачет

Аудиторных часов: 30 всего, в том числе:

лекции: 30 час.

семинары: 0 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 60 час.

Всего часов: 90, всего зач. ед.: 2

Программу составили:

Е.В. Бурнаев, канд. физ.-мат. наук, доцент

А.А. Зайцев

Ю.А. Янович

Программа обсуждена на заседании кафедры проблем передачи информации и анализа данных 01.02.2024

Аннотация

Описывается современное состояние прикладной теории байесовского статистического оценивания. Особое внимание уделяется непараметрическим методам и способам выбора априорного распределения. Исследуется ряд новых типов байесовских оценок: квазипараметрические оценки, байесовские условно минимаксные оценки, оценки вероятности безотказной работы на основе параметрических моделей работоспособности с аддитивной погрешностью. Показываются возможности байесовского подхода при оценивании показателей надежности в расчетных ситуациях, которые практически исчерпывают многообразие задач анализа безотказности технических объектов.

1. Цели и задачи

Цель дисциплины

- дать представление о современном состоянии байесовской статистики и ее использовании в анализе данных.

Задачи дисциплины

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

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-5 Способен и готов к профессиональному росту и руководству коллективом в области информатики и вычислительной техники, толерантно воспринимая социальные, этнические, конфессиональные и культурные различия	ОПК-5.1 Способен работать в коллективе, толерантно воспринимая социальные, этнические, конфессиональные и культурные различия
	ОПК-5.2 Владеет навыком руководства малым коллективом в сфере своей профессиональной деятельности
	ОПК-5.3 Стремится к получению новых знаний, профессиональному и личностному росту
ПК-3 Владеет навыками участия в научных дискуссиях, выступления с сообщениями и докладами устного, письменного и виртуального (размещение в информационных сетях) характера, представления материалов собственных исследований	ПК-3.1 Знает основы ведения научной дискуссии и формы устного научного высказывания
	ПК-3.2 Умеет вести корректную дискуссию в области информационных технологий задавать вопросы и отвечать на поставленные вопросы по теме научной работы
	ПК-3.3 Имеет практический опыт участия в научных студенческих конференциях, очных, виртуальных, заочных обсуждениях научных проблем в области информационных технологий

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- основы байесовского подхода;
- подходы к приближенному байесовскому выводу;
- асимптотические и неасимптотические результаты в байесовской статистике;
- основы непараметрической байесовской статистики;
- примеры использования байесовской статистики в прикладных задачах.

уметь:

- производить байесовский вывод;
- использовать аппарат байесовской статистики в прикладных задачах;
- выбирать априорное распределение;
- использовать аппарат непараметрической байесовской статистики.

владеть:

- навыком освоения большого объема информации;
- навыками постановки научно-исследовательских задач и навыками самостоятельной работы.

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

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

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Основные понятия и задачи статистического оценивания.	6			12
2	Асимптотическая нормальность и неасимптотические результаты о свойствах апостериорного распределения.	6			12
3	Подходы к байесовскому выводу.	6			12
4	Непараметрическая байесовская статистика.	6			12
5	Гауссовские случайные процессы.	6			12
Итого часов		30			60
Подготовка к экзамену		0 час.			
Общая трудоёмкость		90 час., 2 зач.ед.			

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

Семестр: 1 (Осенний)

1. Основные понятия и задачи статистического оценивания.

Основные понятия и задачи статистического оценивания. Экспоненциальное и регулярное семейства распределений. Правдоподобие. Статистическая теория принятия решений Байесовский вывод. Сравнение байесовского и частотного подхода к статистическому оцениванию. Байесовская теория принятия решений Исключение мешающих параметров. Перестановочность. Теорема де Финетти. Выбор модели.

Выбор априорного распределения. Информативное, неинформативное, сопряженное априорные распределения. Априорное распределение Джеффри. Выбор априорного распределения с геометрической точки зрения.

2. Асимптотическая нормальность и неасимптотические результаты о свойствах апостериорного распределения.

Асимптотическая нормальность апостериорного распределения. Теорема Дуба. Условия Ибрагимова и Хасьминского. Состоятельность байесовских оценок. Теорема Бернштейна фон Мизеса.

Неасимптотические результаты о свойствах апостериорного распределения. Подход квазикасимального правдоподобия. Квадратичное приближение правдоподобия в окрестности точки максимума математического ожидания правдоподобия. Неасимптотическая нормальность апостериорного распределения для нормального априорного распределения.

3. Подходы к байесовскому выводу.

Аналитические подходы к приближенному байесовскому выводу. Аппроксимация Лапласа, вариационный вывод. Минимизация расстояния Кульбака-Лейблера и факторизация распределения.

Подходы на основе методов Монте-Карло к приближенному байесовскому выводу. Базовые методы. Схема Метрополиса-Хастингса, схема Гиббса. Оценка нормировочной константы распределения с помощью схемы Гиббса.

Примеры использования байесовского подхода. Машина релевантных векторов, вероятностный метод главных компонент, выбор числа компонент в гауссовской смеси.

4. Непараметрическая байесовская статистика.

Непараметрическая байесовская статистика. Априорные распределения в непараметрическом случае. Случайный процесс Дирихле. Свойства случайного процесса Дирихле.

Непараметрическая байесовская статистика. Сильная и слабая состоятельность непараметрических байесовских оценок. Теорема Шварца.

5. Гауссовские случайные процессы.

Гауссовские случайные процессы. Регрессия на основе гауссовских процессов. Верхняя граница для риска оценки процесса.

Регрессия и классификация на основе гауссовских процессов. Приближенный байесовский вывод. Адаптивное планирование эксперимента и суррогатная оптимизация.

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

доска, ноутбук и мультимедийное оборудование (проектор или плазменная панель).

6.Перечень рекомендуемой литературы

Основная литература

Фонд базовой кафедры

2. C.M. Bishop. Pattern recognition and machine learning, volume 4. Springer New York, 2006.
3. J.K. Ghosh, D. Mohan, and S. Tapas. An introduction to Bayesian analysis. Springer New York, 2006.
4. J.K. Ghosh and R.V. Ramamoorthi. Bayesian nonparametrics. Springer, 2003.
5. B. Kleijn, A. van der Vaart, and H. van Zanten. Lectures on Nonparametric Bayesian Statistics. Springer, 2013.
6. M. La'zaro-Gredilla and M. Titsias. Variational heteroscedastic gaussian process re-gression. ICML, 2011.
7. D.J.C. MacKay. Information theory, inference and learning algorithms. Cambridge university press, 2003.
8. C.E. Rasmussen and C.K.I. Williams. Gaussian processes for machine learning, volume 1. MIT press Cambridge, MA, 2006.
9. V. Spokoiny. Basics of Modern Parametric Statistics. Springer, 2013.
10. L. Wasserman. All of statistics: a concise course in statistical inference. [Springer, 2003.
11. Ибрагимов И.А., Хасьминский Р.З. Асимптотическая теория оценивания. М.: Наука, 1979.

Дополнительная литература

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

Не используются

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

На занятиях используются мультимедийные технологии, включая демонстрацию презентаций.

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

Студент, изучающий дисциплину, должен с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике.

В результате изучения дисциплины студент должен знать основные определения, понятия, аксиомы, алгоритмы.

Успешное освоение курса требует напряжённой самостоятельной работы студента. В программе курса приведено минимально необходимое время для работы студента над темой. Самостоятельная работа включает в себя:

- чтение и конспектирование рекомендованной литературы,
- проработку учебного материала (по учебной и научной литературе), подготовку ответов на вопросы, предназначенных для самостоятельного изучения, доказательство отдельных утверждений, свойств;
- подготовку к дифференцированному зачету.

Руководство и контроль за самостоятельной работой студента осуществляется в форме индивидуальных консультаций.

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

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Информатика и вычислительная техника
профиль подготовки:	Прикладная математика и информатика Физтех-школа Прикладной Математики и Информатики кафедра проблем передачи информации и анализа данных
курс:	<u>1</u>
квалификация:	магистр

Семестр, формы промежуточной аттестации: 1 (осенний) - Дифференцированный зачет

Разработчики:

Е.В. Бурнаев, канд. физ.-мат. наук, доцент

А.А. Зайцев

Ю.А. Янович

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-5 Способен и готов к профессиональному росту и руководству коллективом в области информатики и вычислительной техники, толерантно воспринимая социальные, этнические, конфессиональные и культурные различия	ОПК-5.1 Способен работать в коллективе, толерантно воспринимая социальные, этнические, конфессиональные и культурные различия
	ОПК-5.2 Владеет навыком руководства малым коллективом в сфере своей профессиональной деятельности
	ОПК-5.3 Стремится к получению новых знаний, профессиональному и личностному росту
ПК-3 Владеет навыками участия в научных дискуссиях, выступления с сообщениями и докладами устного, письменного и виртуального (размещение в информационных сетях) характера, представления материалов собственных исследований	ПК-3.1 Знает основы ведения научной дискуссии и формы устного научного высказывания
	ПК-3.2 Умеет вести корректную дискуссию в области информационных технологий задавать вопросы и отвечать на поставленные вопросы по теме научной работы
	ПК-3.3 Имеет практический опыт участия в научных студенческих конференциях, очных, виртуальных, заочных обсуждениях научных проблем в области информационных технологий

2. Показатели оценивания компетенций

В результате изучения дисциплины «Байесовские модели в машинном обучении» обучающийся должен:

знать:

- основы байесовского подхода;
- подходы к приближенному байесовскому выводу;
- асимптотические и неасимптотические результаты в байесовской статистике;
- основы непараметрической байесовской статистики;
- примеры использования байесовской статистики в прикладных задачах.

уметь:

- производить байесовский вывод;
- использовать аппарат байесовской статистики в прикладных задачах;
- выбирать априорное распределение;
- использовать аппарат непараметрической байесовской статистики.

владеть:

- навыком освоения большого объема информации;
- навыками постановки научно-исследовательских задач и навыками самостоятельной работы.

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

С целью контроля освоения обучающимися учебного материала проводится устный опрос в начале занятия или в конце занятия по пройденной теме.

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

1. Байесовский вывод. Сравнение байесовского и вероятностного подходов.
2. Выбор априорного распределения. Виды априорных распределений.
3. Асимптотические результаты в байесовской статистике.
4. Неасимптотические результаты в байесовской статистике. Теорема Бернштейна-фон Мизеса.
5. Аналитические подходы к приближенному байесовскому выводу.
6. Подходы на основе методов Монте-Карло к приближенному байесовскому выводу.

7. Использование байесовского вывода в анализе данных.
8. Априорные распределения в непараметрической байесовской статистике. Случайный процесс Дирихле.
9. Свойства непараметрических байесовских оценок.
10. Регрессия и классификация на основе гауссовских случайных процессов.
11. Оценки риска для регрессии на основе гауссовских случайных процессов.

Критерии оценивания

Оценка отлично 10 баллов - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины, проявляющему интерес к данной предметной области, продемонстрировавшему умение уверенно и творчески применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично 9 баллов - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично 8 баллов - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений, с некоторыми недочетами.

Оценка хорошо 7 баллов - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

Оценка хорошо 6 баллов - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности.

Оценка хорошо 5 баллов - выставляется студенту, если он в основном знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач достаточно большое количество неточностей.

Оценка удовлетворительно 4 балла - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он освоил основные разделы учебной программы, необходимые для дальнейшего обучения, и может применять полученные знания по образцу в стандартной ситуации.

Оценка удовлетворительно 3 балла - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, допускающему ошибки в формулировках базовых понятий, нарушения логической последовательности в изложении программного материала, слабо владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и с трудом применяет полученные знания даже в стандартной ситуации.

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

Оценка неудовлетворительно 1 балл - выставляется студенту, который не знает основного содержания учебной программы дисциплины, допускает грубейшие ошибки в формулировках базовых понятий дисциплины и вообще не имеет навыков решения типовых практических задач.

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

Во время проведения дифференцированного зачёта обучающиеся могут пользоваться программой дисциплины, а также справочной литературой, вычислительной техникой.

Дифференцированный зачёт проводится путем организации специального опроса, проводимого в устной форме.