

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

УТВЕРЖДЕНО

**Директор физтех-школы
электроники, фотоники и
молекулярной физики
А.С. Батурин**

	Рабочая программа дисциплины (модуля)
по дисциплине:	Статистические методы анализа научных данных
по направлению:	Прикладные математика и физика
профиль подготовки:	Физика перспективных технологий: альтернативная энергетика, научное программирование и функциональные материалы Физтех-школа Электроники, Фотоники и Молекулярной Физики кафедра химической физики функциональных материалов
курс:	4
квалификация:	бакалавр

Семестры, формы промежуточной аттестации:

7 (осенний) - Дифференцированный зачет

8 (весенний) - Экзамен

Аудиторных часов: 60 всего, в том числе:

лекции: 60 час.

семинары: 0 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 90 час.

Подготовка к экзамену: 30 час.

Всего часов: 180, всего зач. ед.: 4

Программу составил: В.В. Новиков, д-р хим. наук

Программа обсуждена на заседании кафедры химической физики функциональных материалов 03.03.2023

Аннотация

Курс предназначен для ознакомления обучающихся с продвинутыми методами обработки экспериментальных данных в применении к конкретным выходным данным физических исследований:

1. Комбинаторика, основные комбинаторные объекты
2. Вероятностное пространство как математическая модель, понятия сигма-алгебры и вероятностной меры
3. Условная вероятность, независимость и формула Байеса. Схема Бернулли. Предельные теоремы
4. Случайная величина и распределение. Математическое ожидание и дисперсия. Независимость
5. Закон больших чисел
6. Случайные векторы. Корреляция и ковариация
7. Основная задача статистики. Точечные оценки. Критерии качества оценок
8. Функция штрафа и функция риска
9. Интервальные оценки. Доверительный интервал и квантиль. Эмпирическое распределение
10. Методы построения оценок и статистические гипотезы
11. Преобразование Фурье в физических методах исследования
12. Методы фильтрации сигнала. Распространенные модели и экспериментальные примеры
13. Методы машинного обучения для обработки данных физического эксперимента
14. Обзор продвинутого использования методов машинного обучения в передовых областях науки

1. Цели и задачи

Цель дисциплины

Ознакомление обучающихся с продвинутыми методами обработки экспериментальных данных в применении к конкретным выходным данным физических исследований.

Задачи дисциплины

Формирование у студентов знаний по теоретическим основам статистических методов обработки и анализа данных, а также навыков применения методов машинного обучения и работы с научными библиотеками языка Python: matplotlib, numpy, scipy, scikit-learn.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1 Формулирует совокупность взаимосвязанных задач в рамках поставленной цели работы, обеспечивающих ее достижение. Определяет ожидаемые результаты решения поставленных задач
	УК-2.2 Проектирует решение конкретной задачи проекта, выбирая оптимальный способ ее решения, исходя из действующих правовых норм и имеющихся ресурсов и ограничений
ПК-1 Способен планировать и проводить научные эксперименты (в избранной	ПК-1.3 Владеет культурой постановки научной задачи и моделирования естественнонаучных объектов и систем

предметной области) и (или) теоретические (аналитические и имитационные) исследования	ПК-1.4 Умеет строить математические модели для описания и исследования процессов и явлений в соответствующих научных областях
ПК-2 Способен анализировать полученные в ходе научно-исследовательской работы данные и делать научные выводы (заключения)	ПК-2.1 Владеет методами статистической обработки и анализа научных данных
	ПК-2.2 Умеет находить ключевые параметры, определяющие изучаемое явление, и производить численные оценки по порядку величины
ПК-4 Способен критически оценивать применимость используемых методик и методов	ПК-4.1 Знает численные порядки величин, характерных для соответствующей профессиональной области
	ПК-4.2 Знает источники происхождения и умеет производить оценку погрешности измерений и достоверности экспериментальных результатов
	ПК-4.3 Способен обосновать причинно-следственные отношения используемых понятий и моделей

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- теоретические основы статистических методов обработки и анализа данных;
- продвинутые методы обработки экспериментальных данных в применении к конкретным выходным данным физических исследований;
- специфику совмещения методы обработки экспериментальных данных для различных физико-химических методов.

уметь:

- выдвигать статистические гипотезы;
- строить модели обработки данных в применении к выбранному методу исследования;
- сочетать различные методы анализа данных;
- использовать современное программное обеспечение и методики.

владеть:

- инструментами статистической проверки поставленных гипотез;
- методологией использования современных методов обработки и анализа экспериментальных данных в применении к конкретным выходным данным физических исследований;
- практическими навыками использования современных методов обработки и анализа экспериментальных данных в применении к конкретным выходным данным физических исследований;
- методологией совместного использования различных методов обработки и анализа экспериментальных данных и статистических проверок поставленных гипотез.

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	Комбинаторика, основные комбинаторные объекты	2			6
2	Вероятностное пространство как математическая модель, понятия сигма-алгебры и вероятностной меры	2			6

3	Условная вероятность, независимость и формула Байеса. Схема Бернулли. Предельные теоремы	6			6
4	Случайная величина и распределение. Математическое ожидание и дисперсия. Независимость	2			6
5	Закон больших чисел	2			6
6	Случайные векторы. Корреляция и ковариация	2			6
7	Основная задача статистики. Точечные оценки. Критерии качества оценок	4			6
8	Функция штрафа и функция риска	2			6
9	Интервальные оценки. Доверительный интервал и квантиль. Эмпирическое распределение	4			6
10	Методы построения оценок и статистические гипотезы	4			6
11	Преобразование Фурье в физических методах исследования	4			6
12	Методы фильтрации сигнала. Распространенные модели и экспериментальные примеры	10			10
13	Методы машинного обучения для обработки данных физического эксперимента	14			4
14	Обзор продвинутого использования методов машинного обучения в передовых областях науки	2			10
Итого часов		60			90
Подготовка к экзамену		30 час.			
Общая трудоёмкость		180 час., 4 зач.ед.			

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

Семестр: 7 (Осенний)

1. Комбинаторика, основные комбинаторные объекты

История. Введение в комбинаторику. Комбинаторные правила суммы и произведения. Выборки. Размещения с повторениями. Размещения без повторений. Сочетания без повторений. Сочетания с повторениями. Примеры физических систем с различными видами выборок. Распределения Бозе-Эйнштейна и Ферми-Дирака.

2. Вероятностное пространство как математическая модель, понятия сигма-алгебры и вероятностной меры

Случайные события. σ -алгебра. Вероятностная норма, аксиомы нормированности и σ -аддитивности. Непрерывность вероятности. Классическая вероятность, геометрическая вероятность.

3. Условная вероятность, независимость и формула Байеса. Схема Бернулли. Предельные теоремы

Условная вероятность и формула полной вероятности. Независимость событий. Модель нечестной монетки, схема Бернулли. Предельная теорема Пуассона. Предельные теоремы Муавра-Лапласа. Машина Гамильтона

4. Случайная величина и распределение. Математическое ожидание и дисперсия. Независимость

Случайная величина и функция распределения. Плотность распределения. Дискретные и абсолютно непрерывные функции распределения. Неравенство Коши-Шварца. Свойства математического ожидания. Независимость случайных величин. Свойства дисперсии. Стандартное отклонение

5. Закон больших чисел

Центральная предельная теорема. Неравенство Маркова. Неравенство Чебышева. Закон больших чисел. Условия применимости законов больших чисел для данных физического эксперимента

6. Случайные векторы. Корреляция и ковариация

Коэффициент корреляции. Некоррелированность и независимость случайных величин. Случайный вектор, совместное распределение. Матрица ковариаций и матрица корреляций. Задача линейной регрессии с точки зрения математической статистики. Оценка ошибки линейной регрессии. Реализация построения линейной регрессии для экспериментальной зависимости в Python 3

7. Основная задача статистики. Точечные оценки. Критерии качества оценок

Статистическая выборка. Понятие параметрического множества. Точечная оценка параметров распределения. Несмещенность и состоятельность оценки. Оценка дисперсии при известном математическом ожидании. Асимптотически несмещенная оценка

8. Функция штрафа и функция риска

Функция потерь. Байесовский и минимаксный подходы к поиску оптимальной оценки с помощью функции риска. Асимптотическая нормальность. Асимптотическая дисперсия. Критерии выбора наилучшей оценки

9. Интервальные оценки. Доверительный интервал и квантиль. Эмпирическое распределение

Уровень доверия. t-критерий Стьюдента. Квантиль, квартиль, медиана. Распределение хи-квадрат. Характеристики нормального распределения. Нормальное распределение в физическом эксперименте, правила записи и интерпретации полученных результатов. Эмпирическая функция распределения. Вариационный ряд. Порядковая статистика. Теорема Гливенко-Кантелли. Теорема Колмогорова

10. Методы построения оценок и статистические гипотезы

Метод моментов. Метод наибольшего правдоподобия. Проверка статистических гипотез. Критерий Колмогорова. Гипотезы о согласии. Уровень статистической значимости гипотезы. Критерий хи-квадрат. Теорема Пирсона

Семестр: 8 (Весенний)

11. Преобразование Фурье в физических методах исследования

Фурье-спектрометры. Примеры использующих преобразование Фурье физико-химических приборов, их отличия от иных типов приборов. Разрешающая способность Фурье-спектрометров. Аппаратная функция. Аподизация. Треугольная аподизация, функции Хатта-Гензеля и Блекмана-Харриса. Применение различных типов аподизации для данных ЯМР и ИК-спектрометров в Python

12. Методы фильтрации сигнала. Распространенные модели и экспериментальные примеры

Источники шумов в экспериментальных данных. Понятие белого шума, Броуновского и фликкер-шума. Методы сглаживания. Медианное сглаживание, скользящее среднее. Метод Савицкого-Голея. Линейные методы. Фильтр Калмана. Фильтрация с использованием преобразования Фурье. Фильтры, основанные на вейвлет-преобразовании. Адаптивный метод фильтрации шумов. Методы сравнения методов сглаживания и рекомендуемые подходы при работе с реальными данными. Методы аппроксимации спектров для повышения разрешающей способности

13. Методы машинного обучения для обработки данных физического эксперимента

Библиотека `scikit learn` в Python, её базовый синтаксис и область применения. Предварительная подготовка экспериментальных данных для задач машинного обучения. Стандартизация. Нормализация. Уменьшение регрессии. Выбор и работа с моделью. Обучение с учителем. Обучение без учителя. Регрессия. Линейная регрессия, стохастическая регрессия градиентного спуска, эластичная чистая регрессия, регрессия с усилением градиента. Классификация. Логистическая регрессия, метод k -ближайших соседей, метод опорных векторов, наивный байесовский классификатор. Коэффициенты детерминации. Кластерный анализ. Метод k -средних, спектральная кластеризация, агломеративная кластеризация. Метод главных компонент (`principal component analysis`). Обучение с подкреплением. Проблема переобучения

14. Обзор продвинутого использования методов машинного обучения в передовых областях науки

Основные задачи машинного обучения в экспериментальной науке. Машинное зрение. Медицинские применения классификации и кластеризации. Хемоинформатика. Техническая диагностика. Предсказание структуры и свойств веществ и материалов

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

Учебная аудитория, оснащенная мультимедийным проектором и экраном.

Самостоятельная работа студента обеспечивается доступностью библиотечного фонда

6. Перечень рекомендуемой литературы

Основная литература

1. Проверка статистических гипотез. / Леман Э. – М.: Наука, 1979. – 408 с.
2. Анализ экспериментальных данных: учебное пособие / В. А. Ананьев. – Кемерово: Кемеровский государственный университет, 2009. – Часть 1.
3. Физические методы в химии: в 2 т. Т. 1 / Р. Драго; пер. с англ. А. А. Соловьянова; под ред. О. А. Реутова. — М. : Мир, 1981. — 422 с.
4. Физические методы в химии: в 2 т. Т. 2 : / Р. Драго ; пер. с англ. А. А. Соловьянова ; под ред. О. А. Реутова. — М. : Мир, 1981. — 456 с.
5. Теория вероятностей и математическая статистика: учеб. пособие для вузов / В. Е. Гмурман. — 12-е изд., перераб. — М. : Юрайт, 2010, 2011. — 479 с.
6. Масс-спектрометрия в органической химии: учеб. пособие для вузов / А. Т. Лебедев — М.: БИНОМ. Лаб. знаний, 2009. — 493 с.
7. Методы исследований в экспериментальной физике: учеб. пособие для вузов / М. И. Пергамент. — М. : Интеллект, 2010. — 304 с.

Дополнительная литература

1. Основы молекулярной спектроскопии : учеб. пособие для вузов / Н. Н. Кудрявцев ; М-во образования РФ, МФТИ .— М. : ВЭПИ, 1990.— 158 с.
2. Введение в фурье-спектроскопию / Р.Дж.Белл. — М., "Мир", 1975.
3. Методы и техника обработки сигналов при физических измерениях: в 2 т. / Ж.Макс. — М., "Мир", 1983.
4. Введение в анализ данных с применением непрерывного вейвлет-преобразования. / Левалле, Дж. Syracuse University / Пер. Грибунина ВГ, 1995.

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)

База данных Национального института стандартизации и технологии США по свойствам соединений - <http://webbook.nist.gov/chemistry/>

Научная электронная библиотека РФФИ www.elibrary.ru

Единое окно доступа к образовательным ресурсам Федерального портала Российское образование <http://www.window.edu.ru>

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 дисциплины, уметь применять полученные знания для решения различных задач.

Успешное освоение курса требует:

- посещения всех занятий, предусмотренных учебным планом по дисциплине;
- ведения конспекта занятий;
- напряжённой самостоятельной работы студента.

Самостоятельная работа включает в себя:

- чтение рекомендованной литературы;
- проработку учебного материала, подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- решение задач, предлагаемых студентам на занятиях;
- подготовку к выполнению заданий текущей и промежуточной аттестации.

Показателем владения материалом служит умение без конспекта отвечать на вопросы по темам дисциплины.

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

Возможен промежуточный контроль знаний студентов в виде решения задач в соответствии с тематикой занятий.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Прикладные математика и физика
профиль подготовки:	Физика перспективных технологий: альтернативная энергетика, научное программирование и функциональные материалы Физтех-школа Электроники, Фотоники и Молекулярной Физики кафедра химической физики функциональных материалов
курс:	<u>4</u>
квалификация:	бакалавр
Семестры, формы промежуточной аттестации:	
7 (осенний) - Дифференцированный зачет	
8 (весенний) - Экзамен	
Разработчик:	В.В. Новиков, д-р хим. наук

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1 Формулирует совокупность взаимосвязанных задач в рамках поставленной цели работы, обеспечивающих ее достижение. Определяет ожидаемые результаты решения поставленных задач
	УК-2.2 Проектирует решение конкретной задачи проекта, выбирая оптимальный способ ее решения, исходя из действующих правовых норм и имеющихся ресурсов и ограничений
ПК-1 Способен планировать и проводить научные эксперименты (в избранной предметной области) и (или) теоретические (аналитические и имитационные) исследования	ПК-1.3 Владеет культурой постановки научной задачи и моделирования естественнонаучных объектов и систем
	ПК-1.4 Умеет строить математические модели для описания и исследования процессов и явлений в соответствующих научных областях
ПК-2 Способен анализировать полученные в ходе научно-исследовательской работы данные и делать научные выводы (заключения)	ПК-2.1 Владеет методами статистической обработки и анализа научных данных
	ПК-2.2 Умеет находить ключевые параметры, определяющие изучаемое явление, и производить численные оценки по порядку величины
ПК-4 Способен критически оценивать применимость используемых методик и методов	ПК-4.1 Знает численные порядки величин, характерных для соответствующей профессиональной области
	ПК-4.2 Знает источники происхождения и умеет производить оценку погрешности измерений и достоверности экспериментальных результатов
	ПК-4.3 Способен обосновать причинно-следственные отношения используемых понятий и моделей

2. Показатели оценивания компетенций

В результате изучения дисциплины «Статистические методы анализа научных данных» обучающийся должен:

знать:

- теоретические основы статистических методов обработки и анализа данных;
- продвинутые методы обработки экспериментальных данных в применении к конкретным выходным данным физических исследований;
- специфику совмещения методы обработки экспериментальных данных для различных физико-химических методов.

уметь:

- выдвигать статистические гипотезы;
- строить модели обработки данных в применении к выбранному методу исследования;
- сочетать различные методы анализа данных;
- использовать современное программное обеспечение и методики.

владеть:

- инструментами статистической проверки поставленных гипотез;
- методологией использования современных методов обработки и анализа экспериментальных данных в применении к конкретным выходным данным физических исследований;
- практическими навыками использования современных методов обработки и анализа экспериментальных данных в применении к конкретным выходным данным физических исследований;
- методологией совместного использования различных методов обработки и анализа экспериментальных данных и статистических проверок поставленных гипотез.

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

В начале каждого занятия проводится краткий опрос по теме предыдущей лекции

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

Вопросы к дифференцированному зачету:

Основные определения в машинном обучении. Применение при анализе научных данных.

Основные типы задач машинного обучения на экспериментальных данных.

Статистическое определение вероятности, его особенности и связь с классическим определением.

Формула полной вероятности. Формула Байеса в экспериментальных данных.

Особенности Байесовских методов. Применение при анализе научных данных.

Комбинаторика: размещения, сочетания, перестановки. Размещения, сочетания и перестановки с повторениями. Примеры в экспериментальных данных.

Случайные величины и случайные события в экспериментальных данных. Дискретные и непрерывные случайные величины. Закон распределения случайной величины и способы его задания.

Математическое ожидание случайной величины в условиях физического эксперимента. Его смысл и примеры. Свойства математического ожидания.

Биномиальное распределение. Наивероятнейшее число появлений события. Гипергеометрическое распределение.

Функция Лапласа в условиях физического эксперимента: график, свойства, таблицы. Локальная и интегральная теорема Лапласа. Применение в анализе научных данных.

Основные положения теории выборочного метода. Генеральная совокупность и выборка.

Статистические оценки параметров распределения: несмещенность, состоятельность, эффективность оценок.

Точечные оценки: выборочная средняя, дисперсия, среднее квадратическое отклонение.

Точечная оценка генеральной средней по выборочной средней в экспериментальных данных.

Точечная оценка генеральной дисперсии. "Исправленные" выборочная дисперсия и среднее квадратическое отклонение.

Статистическая проверка гипотез в условиях физического эксперимента. Статистическая гипотеза: параметрическая и непараметрическая; нулевая и альтернативная.

Ошибки I и II рода. Уровень значимости и мощность критерия. Применение в анализе научных данных.

Статистический критерий проверки нулевой гипотезы. Наблюдаемое значение критерия. Критическая область. Область принятия гипотезы.

Критические точки в экспериментальных данных. Отыскание односторонней и двусторонней критических областей.

Основные этапы проверки статистических гипотез в экспериментальных данных.

Проверка гипотезы о равенстве выборочной средней и гипотетической генеральной средней нормальной совокупности при известной и неизвестной генеральной дисперсии.

Проверка гипотезы о равенстве наблюдаемой относительной частоты и гипотетической вероятности появления события в условиях физического эксперимента.

Проверка гипотезы о равенстве долей признака в двух совокупностях.

Проверка гипотезы о законе распределения случайной величины. Критерий согласия Пирсона в условиях физического эксперимента.

Метрические классификаторы – основные методы и алгоритмы задач.

Основные принципы кластеризации.

Принципы подхода «дерево решений».

Представимость функций в виде нейросети в применимости к анализу экспериментальных данных.

Метод опорных векторов. Основные идеи. Применение в анализе научных данных.

Анализ смещения разброса экспериментальных данных. Примеры.

Вопросы к экзамену:

Статистические оценки параметров распределения. Точечные оценки.

Статистическая проверка гипотез в условиях физического эксперимента. Статистическая гипотеза: параметрическая и непараметрическая; нулевая и альтернативная. Ошибки I и II рода. Уровень значимости и мощность критерия.

Основные определения в машинном обучении: объект, целевая функция, признак, модель, обучающая выборка, функционал качества, обучение, переобучение.

Задачи машинного обучения обучение с учителем, без учителя в условиях физического эксперимента.

Задачи регрессии и классификации. Задачи снижения размерности и кластеризации.

Типы признаков в машинном обучении. Приведите примеры различных признаков применимых в условиях физического эксперимента.

Определение ROC-кривой. Применение в анализе научных данных.

Метод ближайших соседей в задаче классификации. Метод ближайших соседей в задаче регрессии.

Методы отбора признаков. Жадный метод. Применимость для анализа научных данных.

Определение отступа в метрических алгоритмах классификации. Алгоритм Condensed Nearest Neighbor.

Обобщение метода ближайших соседей через взвешенный учет объектов. Ядерная оценка плотности.

Проклятие размерности. Зависимость метода ближайших соседей от масштабирования признаков. Способы стандартизации признаков.

Постановка задачи кластеризации. Цели кластеризации. Типы кластерных структур. Чувствительность к нормировке и масштабированию признаков.

Степени свободы метода к средним. Метод k-means++. Метод Xmeans в условиях физического эксперимента.

Метод распространения близости.

Графовые алгоритмы кластеризации.

Алгоритм Ланса-Уильямса. Применимость для анализа научных данных.

Логическая закономерность. Интерпретируемость и информативность.

Решающий список. Достоинства и недостатки.

Структура решающего дерева, метод спуска по дереву в общем случае.

Достоинства и недостатки решающих деревьев в условиях физического эксперимента. Подрезание решающих деревьев. Небрежные решающие деревья.

Деревья принятия решений в задаче регрессии.

Вероятностная постановка задачи классификации. Функция правдоподобия и априорная вероятность.

Функционал среднего риска. Общая формула байесовского классификатора.

Наивный байесовский классификатор в условиях физического эксперимента.

Восстановление плотности распределения по выборке.

Аддитивное сглаживание для байесова классификатора.

Модель МакКаллока-Питтса. Применимость для анализа научных данных.

Обобщённая модель линейного классификатора. Определение отступа. Минимизация эмпирического риска.

Метод градиентного спуска. Выбор величины шага в условиях физического эксперимента.

L2 регуляризация.

Метод стохастического градиента. Недостатки метода SG и как с ними бороться.

Внутренний и внешний функционал качества. Кросс-валидация.

Критерий непротиворечивости моделей.

Аналитическая оценка вероятности переобучения. Схема использования при работе с физическими методами исследования.

Неравенство Бернштейна-Хефлинга в применении к задаче выбора модели.

Дихотомии. Функция роста. Точка разрыва. Применение в анализе научных данных.

Оценка на максимальное число дихотомий в условиях физического эксперимента.

Представимость функций в виде нейросети для анализа экспериментальных данных.

Метод обратного распространения ошибок. Основные недостатки и способы их устранения.

Выбор начального приближения в градиентных методах настройки нейронных сетей в условиях физического эксперимента. Функции активации.

Устройство свёрточной нейросети и целесообразность её применения для анализа различных физико-химических данных.

Нейронные сети для задачи регрессии в условиях физического эксперимента.

Постановка задачи SVM для анализа экспериментальных данных. Регуляризация в задаче SVM.

Двойственная задача SVM. Применение в анализе научных данных.

Ядерный алгоритм SVM.

Представление метода опорных векторов в виде нейронной сети.

Метод SVR для задачи регрессии в условиях физического эксперимента.

Постановка задачи многомерной линейной регрессии для анализа экспериментальных данных. Матричная запись.

Использование сингулярного разложения для решения задачи наименьших квадратов.

Проблема «мультиколлинеарности» в задачах многомерной линейной регрессии.

Гребневая регрессия. Регуляризация Лассо в условиях физического эксперимента.

Нелинейная регрессия. Метод Ньютона-Гаусса для анализа экспериментальных данных.

Задача уменьшения размерности. Метод главных компонент.

Постановка задачи анализа смещения и разброса. Качество обучения в зависимости от пространства моделей.

Внутренний и внешний функционал качества. Средний метод в условиях физического эксперимента.

Определение смещения и разброса. Кривые обучения для анализа экспериментальных данных.

Определение композиции алгоритмов. Типы композиций в условиях физического эксперимента.

Взвешенное голосование. Бустинг. Алгоритм AdaBoost для анализа экспериментальных данных.

Простое голосование. Бэггинг и метод случайных подпространств. Применение в анализе научных данных.

Случайный лес. Стэкинг в условиях физического эксперимента.

Примеры экзаменационных билетов:

Пример 1:

1. Нейронные сети для задачи регрессии в условиях физического эксперимента.
2. Метод ближайших соседей в задаче классификации. Метод ближайших соседей в задаче регрессии

Пример 2:

1. Нелинейная регрессия. Метод Ньютона-Гаусса для анализа экспериментальных данных
2. Типы признаков в машинном обучении. Приведите примеры различных признаков применимых в условиях физического эксперимента

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

Оценка в осеннем семестре выставляется по результатам устного опроса на дифференцированном зачете. Оценка в весеннем семестре выставляется по результатам устного экзамена. Опрос не должен превышать 60 минут.