Машинное обучение в физических задачах

Лекторы:
Станислав Колодный

Язык:
Русский

Трудоемкость:
4 з.е.

Форма контроля:
Экзамен

Образовательная программа:
Теоретическая и экспериментальная физика
8 семестр

Пререквизиты:
Линейная алгебра
Теория вероятностей и математическая статистика
Численные методы
Методы оптимизации
Программирование на Python

Лекции (ак.час)*	Практические занятия (ак.час)	Лабораторные занятия (ак.час)
12	24	
*1 академический час = 45 минутам		

Курс направлен на знакомство с классическим машинным обучением в его современном виде, включая глубокое обучение, рекуррентные сети, автоэнкодеры, генеративно-состязательные сети и обучение с подкреплением. В рамках курса студенты приобретают практические навыки использования Python для машинного обучения и пакета PyTorch в частности. Обучение происходит на примерах из областей, в которых изучаемые подходы наиболее эффективно используются, включая распознавание образов, генерацию текстов, генерацию изображений. Формируется понимание того, какие задачи в работе физика могут быть решены с помощью машинного обучения, приводятся примеры успешного решения физических задач методами машинного обучения.

После успешного освоения курса студенты знают основные методы машинного обучения, современные подходы к решению задач искусственного интеллекта, умеют самостоятельно решать типовые задачи с помощью искусственных нейронных сетей с помощью фреймворка PyTorch, понимают принципы функционирования нейронных сетей, методы верификации, обладают навыками самостоятельной идентификации задач, пригодных для решения методами машинного обучения, выбора подходящего метода.

Машинное обучения в последние несколько лет начало успешно применяться в физике и других науках, позволяя значительно поднять продуктивность исследователей. Ежегодно публикуются десятки статей про применение машинного обучения в физики и их число только растет. Данный курс призван подготовить специалистов, готовых самостоятельно использовать методы машинного обучения, делая особенный акцент на корректности использования методов, верификации результатов. В рамках курса раскрываются современные подходы к решению задач машинного обучения, опубликованные в научной литературе в последние годы и используемые в современном коммерческом и свободной программном обеспечении. Лабораторные работы выполняются с использованием фреймворка PyTorch, который достаточно прост в освоении студентами, при этом достаточно производителен для решения практических задач физики. Отличительной чертой курса является то, что предпринимается попытка обоснования почему каждый конкретный метод и конкретная архитектура нейронной сети позволяет успешно решить конкретную задачу, для чего проводятся параллели с хорошо обоснованными подходами классического программирования.

Содержание курса

8 семестр

Машинное обучение в физических задачах

Структура курса

Разделы	Лабораторные	Лекции (ак.ч.)	Лаб.(ак.ч.)
1. Введение в машинное обучение.			
Статистика, анализ данных, машинное обучение и искусственный интеллект. Основные понятия и обозначения. Постановки и прикладные примеры задач машинного обучения (обучение с учителем, обучение без учителя, обучение с подкреплением). Классификация моделей и методов машинного обучения. Параметрические и непараметрические модели. Методология решения задач машинного обучения. Методы построения, преобразования и отбора признаков. Обнаружение и отсев выбросов. Обобщающая способность модели. Кросс-валидация. Матрица несоответствий. Метрики оценки качества обучения: точность, полнота, чувствительность, специфичность, F-мера, байесовский информационный критерий, критерий Акаике. Дилемма смещения-разброса. Линейная регрессия. Метод наименьших квадратов. Линейный дискриминантный анализ. Мультиколлинеарность и переобучение. Регуляризация. Гребневая регрессия, метод лассо. Логистическая регрессия. Минимизация эмпирического риска и метод стохастического градиента. Вероятностная постановка задачи классификации. Принцип максимума правдоподобия. Наивный байесовский классификатор. Подходы к оценке плотности распределения. Метрический классификатор. Метод ближайших соседей. Метод парзеновских окон. Линейный дискриминантный анализ. Метод опорных векторов. Ансамблевые методы. Бэггинг. Бустинг. Градиентный бустинг на решающих деревьях.	Нейронные сети и задача регрессии	6	6
2. Искусственные нейронные сети.			
Области применения искусственных нейронных сетей. Модель Маккаллока-Питтса. Виды нейронов. Перцептрон Розенблатта. Нейронная реализация логических функций. Функции активации. Сети прямого распространения. Персептронные сети. Метод обратного распространения ошибки. Метод оптимального усечения структуры сети. Самоорганизующиеся сети Кохоннена. Ограниченные машины Больцмана. Алгоритм Contrastive Divergence. Сети доверия. Алгоритм wake-sleep. Глубокое обучение. Использование машин Больцмана для глубоких сетей доверия.		4	
3. Основные приложения нейронных сетей.			
Автоэнкодеры. Сверточные сети. Рекуррентные нейронные сети, сети Хопфилда, долгая краткосрочная память (LSTM). Генеративно-состязательные сети, методы вычисления равновесий по Нэшу. Обучение с подкреплением, отдельные разделы теория игр, методы Монте-Карло. Алгоритм AlphaZero. Методы распознавания образов. Методы генерации изображений. Методы повышения качества изображения. Методы обработки естественных текстов. Планирование движения. Обзор физических задач, решенных методами машинного обучения.	Сверточные сети, классификация изображений. Реккурентные сети, генерация текстов. Генеративно-состязательные сети, улучшение изображения Обучение с подкреплением, задачи планирования	2	6 4 4 4

Рекомендуемые ресурсы

1. Курвилль Аарон, Гудфеллоу Ян. Глубокое обучение. — ДМК Пресс, 2018 — http://www.deeplearningbook.org
2. Андреас Мюллер, Сара Гвидо. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными — O'Reilly Media, 2017 — https://www.oreilly.com/library/view/introduction-to-machine/9781449369880/
3. Жерон Орельен. Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем, 2-е издание. — O'Reilly Media, 2020 — https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/

Дополнительная литература:

4. Machine Learning / Tom M. Mitchell — McGraw-Hill Education, 1997 — http://www.cs.cmu.edu/~tom/mlbook.html
5. Understanding Machine Learning: From Theory to Algorithms / Shai Shalev-Shwartz and Shai Ben-David — Cambridge University Press, 2014 — https://www.cambridge.org/core/books/understanding-machine-learning/3059695661405D25673058E43C8BE2A6
6. Bayesian Reasoning and Machine Learning / David Barber — Cambridge University Press, 2012 —- http://web4.cs.ucl.ac.uk/staff/D.Barber/textbook/200620.pdf

Политика оценивания

Оценочные средства дисциплины: лабораторная работа, презентация, коллоквиум, устный экзамен.

В процессе изучения дисциплины студенты самостоятельно выполняют и далее защищают на занятиях лабораторные работы. Выполнение лабораторных работ в первой части семестра в полном объеме является допуском к коллоквиуму, который в таком случае засчитывается автоматически. В случае, если студент не выполнил данное условие, то дополнительно к сдаче коллоквиума необходимо подготовить презентацию на заранее согласованную с преподавателем тему. Презентация оценивается в 10 баллов.
В конце семестра проводится устный экзамен, в билете два теоретических вопроса, возможны дополнительны вопросы по задачам и материалу.

Максимальное количество баллов за курс – 100.
Оценка формируется исходя из количества баллов:
от 90 до 100 – «отлично», от 74 до 90 – «хорошо», от 60 до 74 – «удовлетворительно».
Максимальное количество баллов за выполнение и защиту лабораторных работ – 16 (минимальное 8) за одну лабораторную работу, 80 (минимальное 40) суммарно.
Максимальное количество баллов за презентацию - 10.
Максимальное количество баллов за устный экзамен – 20.
При невыполнении минимальных требований по любому из пунктов студент получает оценку «неудовлетворительно».