

ФГБОУ ВО Новосибирский ГАУ
Факультет Экономики и управления
Кафедра Информационных технологий и моделирования

Машинное обучение и нейронные сети
Методические указания для проведения лабораторно-практических занятий и
самостоятельной работы студентов

Новосибирск 2021

УДК 004.85+004.032.26 (07)
ББК 32.81, я 7
М 382

Составитель: О.В.Агафонова, доцент кафедры Информационных технологий и моделирования, к.э.н.

Рецензент: О.С.Шинделова, доцент кафедры Учета и финансовых технологий, к.э.н.

Машинное обучение и нейронные сети: методические указания для проведения лабораторно-практических занятий и самостоятельной работы студентов / Новосиб. гос. аграр. ун-т. Фак. ЭиУ; сост. О.В. Агафонова. - Новосибирск, 2021. – 9 с.

Методические указания для проведения лабораторно-практических занятий и самостоятельной работы студентов по дисциплине «Машинное обучение и нейронные сети» предназначены для студентов направления подготовки 09.03.03 Прикладная информатика всех форм обучения.

Методические указания утверждены и рекомендованы к изданию учебно-методическим советом факультета экономики и управления (протокол №4 от «28» декабря 2021 г.)

© Новосибирский государственный аграрный университет, 2021

1. Введение

Лабораторно-практическое занятие – форма организации обучения, которая направлена на формирование практических умений и навыков и является связующим звеном между самостоятельным теоретическим освоением обучающимися учебной дисциплины и применением ее положений на практике.

Лабораторно-практические занятия проводятся в целях: выработки практических умений и приобретения навыков в решении задач, выполнении заданий, производстве расчетов, разработке и оформлении документов, практического овладения компьютерными технологиями. Главным их содержанием является практическая работа каждого обучающегося.

Самостоятельная работа обучающихся – вид деятельности, при котором в условиях систематического уменьшения прямого контакта с преподавателем, обучающимися выполняются учебные задания.

При этом специфика самостоятельной работы обучающихся заключается в том, чтобы они самостоятельно получали новые знания.

Цель и задачи дисциплины

Целью изучения курса «Машинное обучение и нейронные сети» является: формирование у обучающегося целостной системы знаний о теоретических, методологических и практических подходах используемых в работе с данными.

Исходя из цели, в процессе изучения дисциплины решаются следующие задачи:

– описание автоматизируемых с помощью информационных систем бизнес процессы (овладение навыками анализа бизнес процессов с помощью современных средств автоматизации.);

– применение информационных технологий (программных средств и платформ) инфраструктуры информационных технологий организаций, используя современные подходы и стандарты автоматизации;

– анализ функциональных разрывов и корректировка на его основе существующей модели бизнес-процессов (описание действий для внесения изменений в бизнес модель, овладение навыками корректировки существующей модели бизнес-процессов).

Дисциплина «Машинное обучение и нейронные сети» в соответствии с требованиями ФГОС ВО направлена на формирование следующих компетенций бакалавра:

ПК-2. Способен разрабатывать модели бизнес-процессов заказчика, с учетом требований к информационным системам.

ПК-5. Способен проводить адаптацию бизнес-процессов заказчика к возможностям информационной системы.

2. Цель проведения лабораторно-практических занятий и самостоятельной работы обучающихся

Лабораторно-практические занятия направлены на закрепление и расширение знаний, полученных на лекциях.

Лабораторно-практические занятия по курсу «Машинное обучение и нейронные сети» направлены на практическое углубленное изучение учебной дисциплины, привитие обучающимся навыков самостоятельного поиска и анализа учебной информации, формирование и развитие у них научного мышления, умения активно участвовать в творческой дискуссии, делать правильные выводы, аргументировано излагать и отстаивать свое мнение.

Цель самостоятельной работы обучающихся – овладение методами получения новых знаний, приобретение навыков самостоятельного анализа, усиление научных основ практической деятельности.

3. Содержание занятий

Тема 1. Вводная лекция

Машинное обучение в жизни. Спектр применений. Использование методов машинного обучения в науке, технике, медицине, ритейле, рекламе, генерации мультимедиа и других областях. Машинное обучение и самоуправляемые автомобили, системы распознавания речи и видео. Спектр задач машинного обучения: предсказывать зарплату по описанию вакансии, предлагать пользователю музыку на основании его анкеты в интернете, предсказывать отток клиентов компании и т.д.

Тема 2. Математическая база

Функции. Производные и оптимизация. Векторы. Матрицы. Библиотека NumPy. Теория вероятностей.

Тема 3. Введение в машинное обучение

Линейные модели, логистическая регрессия, логические методы классификации.

Тема 4. Глубокое обучение

Применение. Типы нейронных сетей. Модель нейрона. Многослойные нейронные сети. Многоклассовая классификация. Эффективное обучение нейронных сетей.

Тема 5. Свёрточные нейронные сети

Операция свёртки. Свёрточный и пулинг слои. Техника TransferLearning. Архитектуры.

Тема 6. Рекуррентные нейронные сети

Forwardpass. Backwardpass. Проблемы. Архитектура рекуррентного нейрона-LSTM. Схема LSTM-нейрона-GPU. Двухнаправленные рекуррентные нейронные сети.

4. Контролирующие материалы для аттестации по дисциплине

Задача для контрольной работы

В контрольной работе обучающимся предстоит посоревноваться между собой в решении настоящей ML-задачи. Нужно будет предсказать стоимость торгового помещения.

Что входит в успешное решение:

1. Подготовка и тщательный анализ данных. Разбить данные на training sample и validation sample. Проанализировать признаки. Отобрать те, которые лично вам кажутся наиболее важными. Посмотреть на их распределение (в том числе на попарные). Посмотреть на совместное распределение целевой переменной и выбранных признаков.

Сформулировать гипотезы о том, какие сочетания признаков могут быть полезны. Добавить их (если придумаете). Подсказка (одна из идей): отрисуйте выборку на карте, посчитайте расстояние до важных объектов в этом городе, посчитайте статистики по районам и по ближайшим объектам.

2. Выбор алгоритма и его гиперпараметров. Обосновать выбор алгоритма. Если гиперпараметры подбирались, нужно обосновать необходимость в этом. Учесть логику работы алгоритма.

3. Финальный анализ. Проанализировать важности признаков. Соотнести полученную из модели важность признаков с гипотезами из пункта 1, сделать выводы.

Вопросы к зачету

1. Машинное обучение в жизни.
2. Использование методов машинного обучения в науке, технике, медицине, ритейле, рекламе, генерации мультимедиа и других областях.
3. Машинное обучение и самоуправляемые автомобили, системы распознавания речи и видео.
4. Спектр задач машинного обучения.
5. Примеры машинного обучения.
6. Какие ошибки могут возникать в процессе построения модели?
7. Как строится решающее дерево.
8. Деревья и переобучение.
9. Деревья и композиции.
10. Как работает случайный лес.
11. Градиентный бустинг.
12. Градиентный бустинг над решающими деревьями.

13. Задача ранжирования поисковой выдачи.
14. Компьютерное зрение.
15. Обработка естественного языка и речи.
16. Управление: робототехника.
17. Задачи на структурированных данных.
18. Нейросеть.
19. Полносвязная нейросеть.
20. Искусственный нейрон.
21. Нейрон: обучение.
22. Многослойные нейросети.
23. Многоклассовая классификация.
24. Эффективное обучение нейронных сетей.
25. Обучение на больших выборках.
26. Визуализация на линиях уровня.
27. Итерация и эпоха.
28. Алгоритмы оптимизации.

6. Рекомендуемая литература

Основная литература

1. Григорьев, А. А. Методы и алгоритмы обработки данных: учебное пособие / А. А. Григорьев, Е. А. Исаев. — 2-е изд., перераб. и доп. - Москва: ИНФРА-М, 2021. - 383 с. - (Высшее образование:Бакалавриат). - ISBN 978-5-16-015581-4. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1032305>

2. Ниматулаев, М. М. Информационные технологии в профессиональной деятельности: учебник / М. М. Ниматулаев. - Москва: ИНФРА-М, 2020. - 250 с. - (Высшее образование:Бакалавриат). - ISBN 978-5-16-015399-5. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1031122>

Дополнительная литература

1. Гобарева, Я. Л. Бизнес-аналитика средствами Excel: учебное пособие / Я.Л. Гобарева, О.Ю. Городецкая, А.В. Золотарюк. - 3-е изд., перераб. и доп. - Москва: Вузовский учебник: ИНФРА-М, 2021. - 350 с. + Доп. материалы [Электронный ресурс]. - ISBN 978-5-9558-0560-3. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1668637>

2. Пятаева, А. В. Интеллектуальные системы и технологии: учеб.пособие / А. В. Пятаева, К. В. Раевич. - Красноярск: Сиб. федер. ун-т, 2018. - 144 с. - ISBN 978-5-7638-3873-2. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1032131>

3. Современные информационно-коммуникационные технологии для успешного ведения бизнеса: учебное пособие / Ю.Д. Романова, Л.П. Дьяконова, Н.А. Женова [и др.]. - 2-е изд., испр. и доп. - Москва: ИНФРА-М, 2021. - 257 с. + Доп. материалы [Электронный ресурс]. - (Учебники для программы MBA). -

DOI 10.12737/1073931. - ISBN 978-5-16-017053-4. - Текст: электронный. - URL:
<https://znanium.com/catalog/product/1073931>

Информационное обеспечение

1. Электронно-библиотечная система издательства «Лань» // www.e.lanbook.com.
2. Электронно-библиотечная система издательства «ИНФРА-М» // www.znanium.com.
3. Научная электронная библиотека eLibrary.ru
4. Электронно-библиотечная система НГАУ// <http://nsau.edu.ru/library/ebooks/e-lib-sys-nsau/>
5. Библиотечно-информационный ресурс «Федеральное собрание образовательных материалов для студентов».
6. Среда разработки приложений Python 3.7.1 (Anaconda3 2018.12 64 bit)
7. Colaboratory <https://colab.research.google.com/notebooks/welcome.ipynb?hl=ru>
8. Интерактивный туториал <https://pythontutor.ru/> <https://pythonworld.ru/>
9. Большая коллекция материалов по машинному обучению на русском языке <http://www.machinelearning.ru/>
10. Дистрибутив Python с большинством необходимых библиотек <http://anaconda.org>
11. Библиотека для научных вычислений для языка программирования Python <http://scipy.org/>
12. Библиотека для анализа данных pandas <http://pandas.pydata.org/>
13. Документация библиотеки sklearn http://scikitlearn.org/stable/user_guide.html
14. Примеры решения некоторых задач <http://scikitlearn.org/stable/tutorial/index.html>
15. Платформа для проведения конкурсов по решению задач машинного обучения. Содержит обучающие ресурсы с примерами решений задач и их обсуждением <http://kaggle.com>
16. Коллекция данных и задач <http://archive.ics.uci.edu/ml/>
17. Курс «Программирование на Python» по основам программирования на языке Python <https://stepik.org/course/>
18. Курс по основам машинного обучения от Эндрю Ы на (Andrew Ng). Преподается на английском языке <https://www.coursera.org/learn/machinelearning>
19. Курс по основам машинного обучения с использованием Python + pandas + sklearn. Преподается на русском языке <https://ru.coursera.org/learn/vvedeniemashinnoe-obuchenie>
20. Справочно-правовая система «Консультант Плюс» <http://consultant.ru/>

21. Официальный сайт Новосибирского государственного аграрного университета <http://nsau.edu.ru/>

22. Федеральный образовательный портал – Экономика. Социология. Менеджмент <http://ecsocman.hse.ru>

Составитель
Агафонова Ольга Витальевна

Методические указания для проведения лабораторно-практических занятий и
самостоятельной работы студентов

Объем 0,56 уч. – изд. л.

Новосибирский государственный аграрный университет

630039, Новосибирск, ул. Добролюбова, 160