

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Тверской государственный университет»



Утверждаю:

Проректор по ОДиМП

Сердитова Н.Е.

«2» сентября 2024 г.

Рабочая программа факультативной дисциплины (с аннотацией)

Машинное обучение: как это работает и практика

Для всех направлений подготовки физико-технического, математического факультетов и факультета прикладной математики и кибернетики

Для обучающихся
магистратуры, 3,4 курсов бакалавриата, 3-5 курсов специалитета,
очной формы обучения

Составитель: к.ф.-м.н., Романов А.А.

Тверь, 2024

I. Аннотация

1. Цель и задачи дисциплины

Целью освоения дисциплины является изучение принципов работы машинного обучения.

Задачами освоения дисциплины являются:

- подробный разбор принципов работы, лежащих в основе машинного обучения;
- изучение основных моделей и задач, решаемых машинным обучением;
- особенности нейронных сетей;

2. Место дисциплины в структуре ООП

Факультативная дисциплина «Машинное обучение: как это работает и практика» тесно связана с дисциплиной «Обработка физического эксперимента». Дисциплина «Машинное обучение: как это работает и практика» содержит, с одной стороны, материал, который ввиду ограниченности часов, не может быть в полной мере раскрыт в рамках основного курса «Обработка физического эксперимента». С другой стороны, дисциплина «Машинное обучение: как это работает и практика» содержит дополнительный материал, выходящий за рамки основного курса «Обработка физического эксперимента».

Теоретические дисциплины (или модули) и практики, для которых освоение данной дисциплины (или модуля) необходимо как предшествующее: «Математическая статистика», «Теория вероятности», «Информатика».

3. Объем дисциплины: 3 зачетных единицы, 108 академических часов, **в том числе:**

контактная аудиторная работа: лекции 36 часов;

самостоятельная работа: 72 часа.

4. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы

Планируемые результаты освоения образовательной программы (формируемые компетенции)	Планируемые результаты обучения по дисциплине
УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1. Анализирует задачу, выделяя ее базовые составляющие; УК-1.2. Определяет, интерпретирует и ранжирует информацию, требуемую для решения поставленной задачи;
ОПК-1. Способен применять базовые знания в области физико-математических и (или) естественных наук в сфере своей профессиональной деятельности.	ОПК-1.2. Применяет знания в области физико-математических наук при решении практических задач в сфере профессиональной деятельности.

5. Форма промежуточной аттестации зачет.

6. Язык преподавания: русский.

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

1. Для студентов очной формы обучения

Учебная программа – наименование разделов и тем	Всего (час.)	Контактная работа (час.)		Самостоятельная работа, в том числе Контроль (час.)
		Лекции		
		всего	в т.ч. ПП	
1. Регрессионный анализ. Линейная, полиномиальная и логарифмическая регрессия	9	3		6
2. Классификация: Логистическая регрессия и SVM	6	2		4
3. Функции потерь и оптимизация	9	3		6
4. 2. Оценка точности модели, переобучение, регуляризация	6	2		4
5. Проблема качества данных	9	3		6
6. Работа с пропусками	9	3		6
7. Работа с переменными	9	3		6
8. Деревья решений	6	2		4
9. Поиск выбросов и генерация новых признаков	6	2		4

10. Feature Selection	9	3		6
11. Машинное обучение. Введение в НС и библиотеку Keras	9	3		6
12. Углубление в НС и библиотеку Keras	9	3		6
13. Введение в сверточные НС	6	2		4
14. Введение в рекуррентные НС	6	2		4
ИТОГО	108	36		72

III. Образовательные технологии

Учебная программа- наименование разделов и тем	Вид занятия	Образовательные технологии
1. Регрессионный анализ. Линейная, полиномиальная и логарифмическая регрессия	<i>Лекция</i>	<i>Традиционная лекция, активное слушание.</i>
2. Классификация: Логистическая регрессия и SVM	<i>Лекция</i>	<i>Традиционная лекция, активное слушание.</i>
3. Функции потерь и оптимизация	<i>Лекция</i>	<i>Традиционная лекция, активное слушание.</i>
4. 2. Оценка точности модели, переобучение, регуляризация	<i>Лекция</i>	<i>Традиционная лекция, активное слушание.</i>
5. Проблема качества данных	<i>Лекция</i>	<i>Традиционная лекция, активное слушание.</i>
6. Работа с пропусками	<i>Лекция</i>	<i>Традиционная лекция, активное слушание.</i>
7. Работа с переменными	<i>Лекция</i>	<i>Традиционная лекция, активное слушание.</i>
8. Деревья решений	<i>Лекция</i>	<i>Традиционная лекция, активное слушание.</i>
9. Поиск выбросов и генерация новых признаков	<i>Лекция</i>	<i>Традиционная лекция, активное слушание.</i>
10. Feature Selection	<i>Лекция</i>	<i>Традиционная лекция, активное слушание.</i>
11. Машинное обучение. Введение в НС и библиотеку Keras	<i>Лекция</i>	<i>Традиционная лекция, активное слушание.</i>
12. Углубление в НС и библиотеку Keras	<i>Лекция</i>	<i>Традиционная лекция, активное слушание.</i>
13. Введение в сверточные НС	<i>Лекция</i>	<i>Традиционная лекция, активное слушание.</i>
14. Введение в рекуррентные НС	<i>Лекция</i>	<i>Традиционная лекция, активное слушание.</i>

Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

- тематика рефератов и методические рекомендации по их написанию;
- методические рекомендации по организации самостоятельной работы студентов;
- мультимедийные презентации.

IV. Оценочные материалы для проведения текущей и промежуточной аттестации

Для проведения текущей и промежуточной аттестации:

ОПК-1. Способен применять базовые знания в области физико-математических и (или) естественных наук в сфере своей профессиональной деятельности

Примеры задач к зачёту:

Задание 1:

Имеются данные `adult.csv` (см. в материалах к занятию).

Целевой переменной является уровень дохода `income` (крайний правый столбец).

Описание признаков можно найти по ссылке

<http://www.cs.toronto.edu/~dave/data/adult/adultDetail.html>

Вам необходимо построить модели логистической регрессии и SVM, которые предсказывает уровень дохода человека.

Вывести качество полученных моделей на тестовой выборке, используя функцию `score` у модели.

Способ аттестации: Проверка написанной программы и беседа со студентом.

Критерии оценки: Ориентируется в теории и методах решения задач подобного типа – зачет.

Задание 2: Прочитать про методы оптимизации для нейронных сетей <https://habr.com/post/318970/>. Взять код градиентного спуска переделать его для логической регрессии и обучить ее:

- Стандартная
- Методом `nesterov momentum`
- Методом `rmsprop`

Построить график зависимости точности от эпохи, и построить Accuracy, Precision, Recall для разных порогов. Построить график ROC.

Способ аттестации: Проверка написанной программы и беседа со студентом.

Критерии оценки: Ориентируется в теории и методах решения задач подобного типа – зачет.

Задание 3: Загрузить датасет недвижимости.

```
from sklearn.datasets import fetch_california_housing
fetch_california_housing().keys()
```

```
data = load_boston()
```

```
feature_names = data.feature_names
```

```
df = pd.DataFrame(data.data, columns = feature_names)
```

```
df.head()
```

```
data.DESCR
```

```
df['MEDV'] = data.target
```

```
df.head()
```

Попытайтесь обучить линейную регрессию наилучшим качеством.

Первоначально провести анализ данных:

1. Посмотреть статистические данные.
2. Построить тепловые карты корреляций данных
3. Построить парные отношения в наборе данных

Применить линейную регрессию и узнать первоначальную точность.

Произвести разное масштабирование данных и узнать, как влияет оно на точность предсказания.

С помощью модели «Деревьев» `model = RandomForestClassifier()` узнать какие признаки важны, взять только те которые выше определенного порога и проверить точность.

Сгенерировать новые признаки и также обучить, и проверить точность.

Способ аттестации: Проверка написанной программы и беседа со студентом.

УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач.

Примеры тем для самостоятельного изучения:

1. Дифференцируемый нейронный компьютер.
2. Порождающие модели.
3. Глубокое обучение с подкреплением.
4. Методы второго порядка.
5. Теорема Байеса и нейронные сети
6. Состязательные сети.
7. Бустинг.

Темы может предлагать студент самостоятельно, исходя из своих интересов.

Способ аттестации: беседа со студентом.

Критерии оценки: провел анализ литературы по выбранной тематике, показал знание материала и четко изложил результаты исследования – зачет.

V. Учебно-методическое и информационное обеспечение дисциплины

1) Рекомендуемая литература

а) Основная литература:

1. Джонс, Т. Программирование искусственного интеллекта в приложениях: практическое руководство / Т. Джонс ; пер. с англ. А. И. Осипов. - Москва : ДМК Пресс, 2018. - 312 с. - ISBN 978-5-97060-579-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2012525>.

2. Джонс, М. Т. Программирование искусственного интеллекта в приложениях [Электронный ресурс] / М. Тим Джонс ; Пер. с англ. А. И. Осипов. - 2-е изд. - М. : ДМК Пресс, 2011. - 312 с.: ил. - ISBN 978-5-94074-746-8. <http://znanium.com/bookread2.php?book=410211>.

2. Саттон, Р. С. Обучение с подкреплением: введение: руководство / Р. С. Саттон, Э. Д. Барто ; перевод с английского А. А. Слинкина. — Москва: ДМК

Пресс, 2020. — 552 с. — ISBN 978-5-97060-097-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/179453>.

3. Шапиро, Л. Компьютерное зрение: учебное пособие / Л. Шапиро, Д. Стокман ; перевод с английского А. А. Богуславского под редакцией С. М. Соколова. — 5-е изд. (эл.). — Москва : Лаборатория знаний, 2024. — 763 с. — ISBN 978-5-93208-725-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/417998>.

б) Дополнительная литература:

1. Информационные аналитические системы [Электронный ресурс] : учебник / Т. В. Алексеева, Ю. В. Амириди, В. В. Дик и др.; под ред. В. В. Дика. - М.: МФПУ Синергия, 2013. - 384 с. Режим доступа: <http://znanium.com/bookread2.php?book=451186>

2) Программное обеспечение

Google Chrome – бесплатное ПО.

Яндекс Браузер – бесплатное ПО.

Kaspersky Endpoint Security 10 – УПД № ПК 657 от 29.12.2023.

Многофункциональный редактор ONLYOFFICE – бесплатное ПО.

3) Современные профессиональные базы данных и информационные справочные системы

1.ЭБС«ZNANIUM.COM» www.znanium.com

2.ЭБС «Университетская библиотека онлайн» <https://biblioclub.ru/>

3.ЭБС «Лань» <http://e.lanbook.com>

VI. Методические материалы для обучающихся по освоению дисциплины

Методические рекомендации по организации самостоятельной работы студентов:

1) Изучить рекомендуемую литературу. Провести поиск дополнительной литературы в интернете.

2) Просмотреть задачи, разобранные на аудиторных занятиях. Задачи на зачете будут из тех, что рассматриваются на занятиях.

3) Разобрать задачи, рекомендованные преподавателем для самостоятельного решения (как правило, это задания, связанные с развитием рассмотренных на занятии задач), используя, при необходимости, примеры решения аналогичных задач.

4) Обсудить проблемы, возникшие при решении задач с преподавателем.

Для получения зачёта необходимо:

1) Решить задачи, рекомендуемые преподавателем на занятиях.

2) Подготовить исследовательскую реферативную работу по выбранной теме и доложить ее результаты преподавателю.

VII. Материально-техническое обеспечение

Учебная аудитория для проведения занятий лекционного типа, занятий семинарского типа, курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, Учебная аудитория № 28 (170002 Тверская обл., г. Тверь, Садовый пер., д. 35)	1 Экран настенный Screen Media 153x203 2. Комплект учебной мебели на 24 посадочных места 3. Меловая доска 4. Переносной ноутбук 5. Проектор EPSON EB-X05 с потолочным креплением
---	--

VIII. Сведения об обновлении рабочей программы дисциплины

№ п.п.	Обновленный раздел рабочей программы дисциплины	Описание внесенных изменений	Реквизиты документа, утвердившего изменения
1.			