

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

**УТВЕРЖДЕНО**

**Директор физтех-школы  
аэрокосмических технологий  
С.С. Негодяев**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Глубокое обучение в науках о Земле
<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Фундаментальная и прикладная физика природных систем Физтех-школа Аэрокосмических Технологий кафедра термогидромеханики океана
<b>курс:</b>	1
<b>квалификация:</b>	магистр

Семестры, формы промежуточной аттестации:

1 (осенний) - Зачет

2 (весенний) - Дифференцированный зачет

Аудиторных часов: 120 всего, в том числе:

лекции: 60 час.

семинары: 60 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 60 час.

Всего часов: 180, всего зач. ед.: 4

Программу составил: М.А. Криницкий, канд. техн. наук

Программа обсуждена на заседании кафедры термогидромеханики океана 22.03.2022

## Аннотация

Машинное обучение – область науки, объединяющая теорию вероятностей, математическую статистику, инженерный подход к решению сложных задач, подходы обработки данных, искусство программирования и многие другие базовые дисциплины. За последние несколько лет многие задачи наук о Земле были решены с применением машинного обучения намного более успешно, нежели с применением классических методов до этого. С другой стороны, в фундаментальных науках практика применения машинного обучения сталкивается со вполне объяснимым академическим скепсисом, связанным с пониженной интерпретируемостью получаемых результатов. В предлагаемом курсе рассматриваются вопросы применимости методов машинного обучения в различных исследовательских и инженерных задачах наук о Земле, а также вопросы интерпретации, достоверности и неопределенностей результатов.

В настоящем курсе рассматриваются базовые понятия машинного обучения, рассматриваются наиболее часто применяемые алгоритмы классов «обучение с учителем» и «обучение без учителя». Особое внимание уделяется оценке достоверности получаемых решений и оценке неопределенностей как параметров обучаемых моделей, так и аппроксимаций целевых переменных. Изучение методов класса «обучение с учителем» начинается с одной из наиболее изученных моделей – линейной регрессии. На основе этой модели изучаются все основные свойства, присущие большинству методов, и подходы повышения точности и достоверности решений: обобщающая способность, переобучение и недообучение, подходы порождения и отбора признаков, подходы оценки качества моделей в задачах регрессии и классификации. Особое место в курсе занимает комплекс вопросов, посвященных искусственным нейронным сетям и их применению в задачах обработки и анализа данных различной природы. Изучение методов класса «обучение без учителя» включает в себя как классические (метод главных компонент, статистические методы кластеризации), так и современные методы поиска и анализа структуры данных (нейросетевые автокодировщики и порождающие модели).

Курс формирует современное понимание научной деятельности как таковой, включая постановку и проверку гипотез, постановку задач анализа данных. Курс также будет полезен при разработке новых способов натурных наблюдений, при проведении измерений и обработке результатов эксперимента в науках о Земле.

## 1. Цели и задачи

### Цель дисциплины

формирование знаний о математических основах и общих принципах современных методов машинного обучения в применении к задачам наук о Земле; освоение общепринятых методик применения методов машинного обучения.

### Задачи дисциплины

- дать студентам знания об общих математических принципах современных методов машинного обучения;
- научить студентов самостоятельно формулировать задачу, планировать численный эксперимент, выбирать подходящий метод решения и эффективно его реализовывать в виде программы, а также анализировать результаты и оценивать качество получаемых моделей;
- выработать у студентов навыки эффективного применения методов машинного обучения с использованием доступных языков программирования и сред исполнения программного кода;
- выработать у студентов навык адаптации существующих методов машинного обучения с учетом специфики задач и с использованием результатов новейших публикаций;
- выработать у студентов навык визуального представления данных, представления промежуточных и конечных результатов исследования.

## 2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Имеет представление об актуальных проблемах науки и техники в области своей профессиональной деятельности, способен	ОПК-2.1 Имеет представление о современном состоянии исследований в рамках тематической области своей профессиональной деятельности
	ОПК-2.2 Способен оценивать актуальность исследований в области своей профессиональной деятельности и их практическую значимость

на научном языке формулировать профессиональные задачи	ОПК-2.3 Владеет профессиональной терминологией, используемой в современной научно-технической литературе, обладает навыками устного и письменного изложения результатов научной деятельности в рамках профессиональной коммуникации
ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области физико-математических наук и информационно-коммуникационных технологий	ОПК-4.2 Способен применять знания в области физико-математических наук для решения поставленной задачи, формулирования выводов и оценки полученных результатов
	ОПК-4.3 Способен аргументировано выбирать способ проведения научного исследования
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты

### 3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- основные понятия и математические принципы современных методов машинного обучения;
- типы задач, решаемые с применением методов машинного обучения;
- наиболее распространенные модели машинного обучения, области их применения, основные преимущества и недостатки;
- основные показатели качества моделей машинного обучения в задачах различного типа, их преимущества, недостатки, ограничения применимости и характерные значения, достигаемые по результатам современных исследований.

уметь:

- определять тип задачи с точки зрения машинного обучения, формулировать задачу в терминах методов машинного обучения, описывать исходные данные и целевые переменные, подбирать подходящий тип модели, формулировать и обосновывать метрики качества получаемого решения;
- реализовывать в виде программы цепь обработки данных и тренировки модели машинного обучения в рамках решения поставленной задачи;
- идентифицировать явления недообучения и переобучения для различных типов задач и для различных конкретных видов моделей машинного обучения, руководствуясь метриками качества решения и диагностическими показателями процесса обучения; принимать меры для купирования эффектов недообучения и переобучения;
- проводить исследование чувствительности модели к значениям гиперпараметров; проводить оптимизацию гиперпараметров;
- исследовать исходные данные в аспекте сформулированной задачи; оценивать границы применимости и возможные причины смещенности полученного решения.

владеть:

- навыками самостоятельной реализации алгоритмов машинного обучения по материалам современных исследований, изложенных в научных статьях;
- навыками адаптации существующих алгоритмов машинного обучения с учетом особенностей сформулированной задачи;
- навыками оптимизации процессов предобработки исходных данных и постобработки результатов численных экспериментов.

### 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	2	2		4
2	Построение ИНС	4	4		2
3	Современные инструменты для построения ИНС	4	4		4
4	Современные инструменты для построения СНС	4	4		2
5	СНС: современные архитектуры	4	4		4
6	Интерпретация СНС	4	4		2
7	Двумерные данные	2	2		4
8	Автокодировщики	2	2		4
9	Metric Learning	4	4		4
10	Задачи типа «обучение без учителя»: сокращение размерности	4	4		4
11	Нейросетевой автокодировщик	2	2		4
12	Задачи типа «обучение без учителя»: кластеризация	4	4		4
13	Кластеризация: иерархические алгоритмы	4	4		2
14	Рекуррентные нейронные сети (РНС)	4	4		4
15	Обработка естественного языка	4	4		2
16	Порождающие модели	2	2		4
17	Обучение порождающих состязательных сетей	4	4		2
18	Задачи, решаемые GAN	2	2		4
Итого часов		60	60		60
Подготовка к экзамену		0 час.			
Общая трудоёмкость		180 час., 4 зач.ед.			

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

Семестр: 1 (Осенний)

##### 1. Основные понятия

Машинное обучение: определения, взаимосвязь понятий, классификация задач и методов машинного обучения. Краткий обзор методов обучения «с учителем». Краткий обзор методов обучения «без учителя».

Вероятностная постановка задачи обучения по прецедентам. Линейная регрессия как вероятностная модель. Общая схема решения задач машинного обучения в подходе «обучение с учителем». Логистическая регрессия, обобщенные линейные модели, обобщенные аддитивные модели. Перцептрон, многослойный перцептрон.

Оптимизация моделей машинного обучения и настройка гиперпараметров. Ландшафт функции потерь, его свойства и особенности для различных моделей и задач.

Явление переобучения и недообучения моделей. Баланс между смещением и разбросом. Настройка гиперпараметров модели. Подход скользящего контроля. Стратегии скользящего контроля.

Методы классификации: метод опорных векторов, деревья решений, композиционные алгоритмы.

Искусственные нейронные сети (ИНС). Многослойный перцептрон. Сверточные нейронные сети (СНС). Функции потерь ИНС для различных задач. Обучение искусственных нейронных сетей методом стохастического градиентного спуска. Вычисление градиентов функции потерь по параметрам сети. Функции активации. Проблемы обучения глубоких нейронных сетей. Способы стабилизации обучения глубоких нейронных сетей. Варианты инициализации параметров. Пакетная нормализация. Прореживание. Регуляризации.

## 2. Построение ИНС

Построение, обучение и оценка качества полносвязной ИНС с применением базового инструментария (библиотека матричных вычислений numpy). Организация порождения обучающих данных и цепи вычислений процесса оптимизации ИНС.

## 3. Современные инструменты для построения ИНС

Построение, обучение и оценка качества полносвязной ИНС с применением современных библиотек глубокого обучения (tensorflow, pytorch). Организация порождения обучающих данных и цепи вычислений процесса оптимизации ИНС.

## 4. Современные инструменты для построения СНС

Построение, обучение и оценка качества СНС с применением современных библиотек глубокого обучения (tensorflow, pytorch). Организация порождения обучающих данных и цепи вычислений процесса оптимизации ИНС.

## 5. СНС: современные архитектуры

Обзор современных архитектур СНС: VGG, Resnet, U-net, Segnet, RetinaNet, DeepLab (V1,V2,V3). Виды задач, решаемых с их применением.

## 6. Интерпретация СНС

Визуализация и интерпретация СНС на этапе обучения и на этапе применения. Визуализация карт активации. Визуализация распределений активаций. Визуализация распределений градиентов. Saliency maps. Визуализация внутреннего состояния. Attention map. Adversarial examples.

## 7. Двумерные данные

Задачи преобразования изображения в изображение. Задача семантической сегментации. Задача регрессии двумерных данных.

## 8. Автокодировщики

Нейросетевые архитектуры типа кодировщик-декодер. Автокодировщик. Варианты автокодировщиков: разреженный, вариационный, понижающий, повышающий.

## 9. Metric Learning

Задача типа «обучение метрике». Современные нейросетевые архитектуры в задаче Metric Learning, функции потерь.

## Семестр: 2 (Весенний)

## 10. Задачи типа «обучение без учителя»: сокращение размерности

Метод главных компонент. t-SNE и UMAP. Самообучающиеся карты Кохонена.

## 11. Нейросетевой автокодировщик

Нейросетевой автокодировщик в задачах сокращения размерности. Вопрос эквивалентности линейного понижающего автокодировщика и метода главных компонент.

## 12. Задачи типа «обучение без учителя»: кластеризация

Постановка задачи. Мера близости событий и проклятие размерности в задачах кластеризации. Виды алгоритмов кластеризации. Графовые и эвристические алгоритмы. DBSCAN и HDBSCAN. Статистические алгоритмы. Метод разделения гауссовой смеси, метод К средних.

## 13. Кластеризация: иерархические алгоритмы

Метод Ланса-Уильямса. Вариации метода Ланса-Уильямса. Свойства иерархических алгоритмов.

Практика по методам «обучения без учителя».

## 14. Рекуррентные нейронные сети (РНС)

Краткий исторический обзор. Принцип локальности признаков. Кодирование последовательностей. One-hot, Word2Vec, GloVe, fastText. Рекуррентные нейронные сети: основные принципы. LSTM, двунаправленный LSTM, GRU.

## 15. Обработка естественного языка

Современные нейросетевые языковые модели. Типы задач, решаемые РНС. Классификация и регрессия, порождение последовательности на базе последовательности. Архитектура Transformer.

## 16. Порождающие модели

Нейросетевые порождающие модели. Принцип и статистические основы порождающих состязательных сетей (GAN). Перекрестная энтропия, правдоподобие, информационная энтропия. Дивергенции Кульбака-Лейблера и Шеннона-Йенсена.

## 17. Обучение порождающих состязательных сетей

Анализ обучения GAN. Стабилизация обучения GAN. Wasserstein GAN.

## 18. Задачи, решаемые GAN

DCGAN, LSGAN, WGAN. Другие варианты GAN. Задачи, решаемые современными GAN.

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

учебная аудитория, оснащенная компьютером и мультимедийным оборудованием (проектор, звуковая система).

## 6. Перечень рекомендуемой литературы

### Основная литература

1. Машинное обучение [Текст]/Х. Бринк, Дж. Ричардс, М. Феверолф, Real-World Machine Learning, -СПб., Питер, 2017

1. Флах П. «Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных.» / Флах П. М.: ДМК Пресс, 2015. 400 с.
2. Шай Шалев-Шварц, Шай Бен-Давид "Идеи машинного обучения." / - М.: ДМК Пресс, 2018. 432 с.
3. Гудфеллоу Я., Бенджио И., Курвилль А. «Глубокое обучение.» / М.: ДМК Пресс, 2017. 652 с.
4. Николенко С. И., Кадури А. А., Архангельская Е. О. «Глубокое обучение.» / СПб.: Питер. 2019. 480 с.

#### Дополнительная литература

1. Hastie T., Tibshirani R., Friedman J. «The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition» / T. Hastie, R. Tibshirani, J. Friedman, 2-е изд., New York: Springer-Verlag, 2009.
2. Bishop C. «Pattern Recognition and Machine Learning» / C. Bishop, New York: Springer-Verlag, 2006.

### **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)**

Интерактивная среда разработки программ PyCharm (производитель JetBrains, предоставляется бесплатно для студентов).

Язык программирования Python 3, клиент-серверная среда исполнения кода Jupyter

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

Настоящий курс ориентирован на аудиторию студентов, научная работа которых связана с анализом данных в науках о Земле. В последнее время направление статистического анализа данных динамично развивается, разрабатываются новые методы поиска и аппроксимации закономерностей в натурных данных. Несмотря на порой меньшую интерпретируемость по сравнению с классическими статистическими методами, современные подходы машинного обучения занимают свое место благодаря своей предиктивной способности, что ставит эту дисциплину в ряд обязательных для освоения студентом, планирующим успешную научную карьеру.

Основной упор курса сделан на приобретение практических навыков применения методов машинного обучения в реальных задачах геофизики при условии понимания математических оснований этих методов. Большинство рассматриваемых тем сопровождаются примерами задач, разбираемыми на семинарах. Кроме того, в процессе изучения курса студентам на основании полученных знаний предлагается выполнить ряд заданий по анализу данных – как синтетических, так и полученных в ходе различных экспедиций и экспериментов. Для закрепления получаемых знаний и навыков студенты получают как теоретические, так и практические домашние задания.

По итогу освоения курса студентам предлагается выполнить самостоятельную работу. Тему работы студент выбирает, руководствуясь собственными интересами в геофизике или выбирает из предложенных преподавателем. Для уточнения темы проводится консультация с преподавателем.

Успешное освоение курса возможно лишь при условии большой самостоятельной работы студента.

Самостоятельная работа включает в себя:

- решение задач при выполнении домашних заданий;
- чтение рекомендованной литературы;
- проработку учебного материала (по конспектам, учебной и научной литературе);

– самостоятельное освоение современных методов машинного обучения, не затронутых в курсе, на основе предложенного учебного материала.

Руководство и контроль за самостоятельной работой студента осуществляется анализом итогов домашних заданий, курсовой работы, а также индивидуальных консультаций.

Показателем владения материалом служит умение ставить и решать задачи. Для формирования умения применять знания на практике студенту необходимо решать как можно больше задач.



**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Фундаментальная и прикладная физика природных систем Физтех-школа Аэрокосмических Технологий кафедра термогидромеханики океана
<b>курс:</b>	<u>1</u>
<b>квалификация:</b>	магистр

Семестры, формы промежуточной аттестации:

1 (осенний) - Зачет

2 (весенний) - Дифференцированный зачет

**Разработчик:** М.А. Криницкий, канд. техн. наук

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Имеет представление об актуальных проблемах науки и техники в области своей профессиональной деятельности, способен на научном языке формулировать профессиональные задачи	ОПК-2.1 Имеет представление о современном состоянии исследований в рамках тематической области своей профессиональной деятельности
	ОПК-2.2 Способен оценивать актуальность исследований в области своей профессиональной деятельности и их практическую значимость
	ОПК-2.3 Владеет профессиональной терминологией, используемой в современной научно-технической литературе, обладает навыками устного и письменного изложения результатов научной деятельности в рамках профессиональной коммуникации
ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области физико-математических наук и информационно-коммуникационных технологий	ОПК-4.2 Способен применять знания в области физико-математических наук для решения поставленной задачи, формулирования выводов и оценки полученных результатов
	ОПК-4.3 Способен аргументировано выбирать способ проведения научного исследования
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты

## 2. Показатели оценивания компетенций

В результате изучения дисциплины «Глубокое обучение в науках о Земле» обучающийся должен:

### знать:

- основные понятия и математические принципы современных методов машинного обучения;
- типы задач, решаемые с применением методов машинного обучения;
- наиболее распространенные модели машинного обучения, области их применения, основные преимущества и недостатки;
- основные показатели качества моделей машинного обучения в задачах различного типа, их преимущества, недостатки, ограничения применимости и характерные значения, достигаемые по результатам современных исследований.

### уметь:

- определять тип задачи с точки зрения машинного обучения, формулировать задачу в терминах методов машинного обучения, описывать исходные данные и целевые переменные, подбирать подходящий тип модели, формулировать и обосновывать метрики качества получаемого решения;
- реализовывать в виде программы цепь обработки данных и тренировки модели машинного обучения в рамках решения поставленной задачи;
- идентифицировать явления недообучения и переобучения для различных типов задач и для различных конкретных видов моделей машинного обучения, руководствуясь метриками качества решения и диагностическими показателями процесса обучения; принимать меры для купирования эффектов недообучения и переобучения;
- проводить исследование чувствительности модели к значениям гиперпараметров; проводить оптимизацию гиперпараметров;
- исследовать исходные данные в аспекте сформулированной задачи; оценивать границы применимости и возможные причины смещенности полученного решения.

### владеть:

- навыками самостоятельной реализации алгоритмов машинного обучения по материалам современных исследований, изложенных в научных статьях;
- навыками адаптации существующих алгоритмов машинного обучения с учетом особенностей сформулированной задачи;
- навыками оптимизации процессов предобработки исходных данных и постобработки результатов численных экспериментов.

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

Текущий контроль осуществляется в форме домашних заданий по каждой теме, а также в форме оценки практики на семинарах.

Каждое домашнее задание оценивается согласно указанному весу. В рамках каждого домашнего задания предлагаются задачи, каждая из которых оценивается согласно указанному количеству баллов. Суммарное количество баллов за каждое домашнее задание - 100. Кроме того, для студентов, желающих более глубоко освоить материал, предлагаются опциональные задачи. Суммарное количество баллов по домашнему заданию с учетом веса этого домашнего задания идет в общий зачет за семестр.

За осенний семестр максимальное количество взвешенных баллов – 500 без учета опциональных заданий.

За весенний семестр максимальное количество взвешенных баллов – 600 без учета опциональных заданий. Условие допуска к дифференцированному зачету – набор не менее 150 взвешенных баллов за семестр.

Домашнее задание по теме «Построение ИНС»

Написать программу, реализующую построение, обучение и оценку качества полносвязной ИНС с применением базового инструментария (библиотека матричных вычислений `numpy`). В программе организовать порождение обучающих данных и цепь вычислений процесса оптимизации ИНС.

Домашнее задание по теме «Современные инструменты для построения ИНС»

Написать программу, реализующую построение, обучение и оценку качества полносвязной ИНС с применением современных библиотек глубокого обучения (`tensorflow/pytorch`). В программе организовать порождение обучающих данных и цепь вычислений процесса оптимизации ИНС.

Домашнее задание по теме «Современные инструменты для построения СНС»

Написать программу, реализующую построение, обучение и оценку качества СНС с применением современных библиотек глубокого обучения (`tensorflow/pytorch`). В программе организовать порождение обучающих данных и цепь вычислений процесса оптимизации ИНС.

Домашнее задание по теме «СНС: современные архитектуры»

Написать программу, реализующую построение, обучение и оценку качества СНС в подходе использования заранее обученных моделей СНС.

Домашнее задание по теме «Задачи преобразования изображения в изображение»

1. Написать программу, реализующую построение, обучение и оценку качества СНС в задаче семантической сегментации.
2. Написать программу, реализующую построение, обучение и оценку качества СНС в задаче регрессии действительной переменной на двумерной регулярной сетке.

Домашнее задание по теме «Автокодировщики»

1. Написать программу, реализующую фильтрацию шумов в двумерных данных с применением автокодировщиков
2. Написать программу, реализующую обнаружение аномалий с применением автокодировщиков

Домашнее задание по теме «Metric Learning»

Написать программу, реализующую обучение метрики в задаче сравнения геофизических объектов.

Домашнее задание по теме «Сокращение размерности»

1. Написать программу, реализующую понижение размерности многомерных данных с применением методов: PCA, t-SNE, UMAP.
2. Отобразить полученные данные на двумерной плоскости.

Домашнее задание по теме «Кластеризация»

Кластеризовать объекты предложенного набора данных с применением подходящего алгоритма кластеризации. Применить метод понижения размерности и отобразить полученные данные на двумерной плоскости с обозначением кластеров.

Домашнее задание по теме «Рекуррентные нейронные сети»

Решить задачу распознавания именованных сущностей с применением рекуррентных нейронных сетей.

Домашнее задание по теме «Порождающие модели»

Решить задачу аппроксимации распределения переменных скрытого представления объектов предложенного набора данных. Применить полученную модель для порождения новых экземпляров выборки. Проверить принадлежность порожденных экземпляров к распределению объектов предложенного набора данных.

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

##### **ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 1**

1. Теория: Машинное обучение и искусственный интеллект: основные понятия, определения, основные исторические события. Современные достижения в приложениях МО и ИИ. Современные достижения МО и ИИ в геофизике.
2. Практика: Реализация понижающего нейросетевого автокодировщика. Обучение на данных MNIST. Визуализация скрытых представлений с использованием t-SNE.

##### **ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 2**

1. Теория: Классификация задач МО с примерами задач геофизики.
2. Практика: Решение задачи классификации рукописных цифр из набора данных MNIST методом опорных векторов (опционально) с предварительным обучением нейросетевого генератора признаков с применением понижающего автокодировщика.

##### **ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 3**

1. Теория: Линейная регрессия как статистическая модель. Линейная регрессия как однослойный перцептрон.
2. Практика: Кластеризация данных - на выбор из предложенных вариантов. Выбор метода – на выбор студента.

##### **ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 4**

1. Теория: Принципы машинного обучения. Общая схема решения задач обучения по прецедентам.
2. Практика: Постановка и решение задачи классификации по предлагаемым наборам данных. Выбор набора данных и метода – на выбор студента.

##### **ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 5**

1. Теория: Логистическая регрессия как статистическая модель. Логистическая регрессия как однослойный перцептрон.
2. Практика: Классификация состояний стратосферного полярного вихря с применением сверточной нейронной сети.

##### **ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 6**

1. Теория: Методы оптимизации в МО. Ландшафт функции потерь. Методы первого и второго порядков.
2. Практика: Исследование и визуализация характеристик статистических признаков данных широкоугольных снимков видимой полусферы небосвода.

#### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 7

1. Теория: Переобучение и недообучение моделей МО. Смещение и разброс. Подходы балансирования смещения и разброса.
2. Практика: Визуализация данных, описывающих состояние стратосферного полярного вихря, с применением t-SNE и самоорганизующихся карт Кохонена.

#### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 8

1. Теория: Метод опорных векторов. Метод опорных векторов с ядром. Метод опорных векторов как двуслойный перцептрон.
2. Практика: Постановка и решение задачи классификации по предлагаемым наборам данных. Выбор набора данных и метода – на выбор студента. (Опционально) с предварительным обучением нейросетевого генератора признаков с применением понижающего вариационного автокодировщика.

#### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 9

1. Теория: Деревья решений. Композиционные алгоритмы на основе деревьев решений.
2. Практика: Постановка и решение задачи кластеризации по предлагаемым наборам данных. Выбор набора данных и метода – на выбор студента. Визуализация результатов (при необходимости) с применением методов сокращения размерности.

#### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 10

1. Теория: Искусственные нейронные сети. Математическая модель искусственного нейрона. Однослойный перцептрон. Многослойный перцептрон. Обучение ИНС методом обратного распространения ошибки. Обобщающая способность и методы регуляризации ИНС.
2. Практика: Постановка и решение задачи оценки балла общей облачности по данным широкоугольных оптических снимков видимой полусферы небосвода. Метод – композиционный на базе деревьев решений, на выбор.

#### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 11

3. Теория: Сверточные нейронные сети. Принципы, лежащие в основе СНС. Операция дискретной одномерной и двумерной свертки. Основные блоки СНС. Современные архитектуры СНС. Задачи, решаемые с применением СНС.
4. Практика: Исследование синтетического набора данных. Постановка и решение задачи кластеризации на этом наборе данных. Анализ решения и свойств полученной модели. Вид модели – на выбор студента.

#### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 12

5. Теория: Задача сокращения размерности. Метод главных компонент как статистическая модель. Самообучающиеся карты Кохонена. Понижающий нейросетевой автокодировщик.
6. Практика: Постановка и решение задачи классификации на синтетических данных с известными свойствами. Анализ решения и свойств полученной модели. Вид модели – многослойный перцептрон.

#### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 13

7. Теория: Задача кластеризации. Виды методов кластеризации. Отдельные методы кластеризации.
8. Практика: Постановка и решение задачи восстановления регрессии на синтетических данных с известными свойствами. Анализ решения и свойств полученной модели. Вид модели – многослойный перцептрон.

#### Критерии оценивания

оценка «отлично (10)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины при выполнении курсовой работы, домашних заданий, ответе экзаменационного билета и ответе на вопросы по программе дисциплины;

оценка «отлично (9)» выставляется студенту, показавшему систематизированные, глубокие знания учебной программы дисциплины при выполнении курсовой работы, домашних заданий, ответе экзаменационного билета и ответе на вопросы по программе дисциплины;

оценка «отлично (8)» выставляется студенту, показавшему систематизированные, знания учебной программы дисциплины при выполнении курсовой работы, домашних заданий, ответе экзаменационного билета и ответе на вопросы по программе дисциплины;

оценка «хорошо (7)» выставляется студенту, если он твердо знает материал экзаменационного билета, грамотно и по существу излагает его, демонстрирует умение применять полученные знания на практике при выполнении курсовой работы и домашних заданий, но допускает в ответе или в решении задач некоторые неточности;

оценка «хорошо (6)» выставляется студенту, если он знает материал экзаменационного билета, по существу излагает его, демонстрирует умение применять полученные знания на практике при выполнении курсовой работы и домашних заданий, но допускает в ответе или в решении задач много неточностей;

оценка «хорошо (5)» выставляется студенту, если он знает материал экзаменационного билета, излагает его, демонстрирует умение применять полученные знания на практике при выполнении курсовой работы и домашних заданий, не допускает в ответе грубых ошибок;

оценка «удовлетворительно (4)» выставляется студенту, если во время ответа экзаменационного билета, при выполнении курсовой работы и домашних заданий он показал фрагментарный, характер знаний, недостаточно правильные формулировки базовых понятий, но при этом он владеет основными разделами учебной программы, необходимыми для дальнейшего обучения;

оценка «удовлетворительно (3)» выставляется студенту, если во время ответа экзаменационного билета, при выполнении курсовой работы и домашних заданий он показал разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушение логической последовательности в изложении программного материала, но при этом он владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;

оценка «неудовлетворительно (2-1)» выставляется студенту, если во время ответа экзаменационного билета, он показал, что не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач.

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

Оценка за промежуточную аттестацию в виде зачёта выставляется по результатам текущего контроля успеваемости.

При проведении устного дифференцированного зачёта обучающемуся предоставляется 60 минут на подготовку. Опрос обучающегося по билету не должен превышать двух астрономических часов.

Во время проведения дифференцированного зачёта при подготовке ответов на билеты, обучающиеся могут пользоваться программой дисциплины, конспектами и любой другой литературой.

Во время проведения дифференцированного зачёта при ответе обучающегося на вопросы по билету или по программе дисциплины, он не может пользоваться конспектами и любой другой литературой.