

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

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

	Рабочая программа дисциплины (модуля)
по дисциплине:	Методы машинного обучения в логистических системах
по направлению:	Системный анализ и управление
профиль подготовки:	Системный анализ и управление в больших системах Физтех-школа Аэрокосмических Технологий кафедра логистических систем и технологий
курс:	1
квалификация:	магистр

Семестр, формы промежуточной аттестации: 2 (весенний) - Дифференцированный зачет

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 30 час.

Всего часов: 90, всего зач. ед.: 2

Программу составил: А.В. Пушкарева, ассистент

Программа обсуждена на заседании кафедры логистических систем и технологий 09.02.2022

Аннотация

В рамках учебной дисциплины «Методы машинного обучения в логистических системах» рассматриваются основные подходы и методы современной машинного обучения в применении к логистике.

Студент, изучающий дисциплину «Методы машинного обучения в логистических системах», должен овладеть современными подходами машинного обучения к решению задач логистики и получить навыки решения практических задач с использованием современных программных пакетов.

1. Цели и задачи

Цель дисциплины

- изучение подходов машинного обучения в логистике, освоение технологий извлечения информации из неструктурированных или слабоструктурированных систем;
- овладение навыками практического решения задач машинного анализа данных в логистике.

Задачи дисциплины

- освоение студентами подходов, методов и моделей извлечения информации с целью анализа процессов в логистике и повышения точности построения анализа и прогноза, а именно, прогнозирования запасов, поставок и спроса, сокращения расходов на перевозку, минимизации рисков, поиска факторов эффективности, улучшения планирования и управления, повышения прозрачности;
- приобретение в условиях слабой структурированности данных практических навыков извлечения знаний;
- формирование навыков по интерпретации полученных результатов для подбора оптимальных моделей и объяснения возникающих в микросистемах эффектов.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-3 Способен формулировать, формировать и применять критерии оценки эффективности полученных результатов разработки систем управления и их внедрения в производственной и непроизводственной сферах	ОПК-3.1 Проводит анализ этапов разработки и внедрения систем управления, и оценивает эффективность результатов
	ОПК-3.2 Разрабатывает и применяет критерии оценки эффективности полученных результатов на основании полученных знаний
ОПК-4 Способен применять методы математического, функционального и системного анализа для задач моделирования, анализа и синтеза автоматического управления техническими объектами	ОПК-4.2 Проводит анализ и моделирование при помощи методов математического, функционального и системного анализа при решении прикладных и теоретических задач автоматического управления техническими объектами
ОПК-5 Способен выбирать методы и разрабатывать алгоритмы решения задач автоматического управления сложными управляемыми объектами	ОПК-5.1 Анализирует и определяет оптимальные методы для решения задач автоматического управления
	ОПК-5.2 Разрабатывает алгоритмы для решения задач автоматического управления

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- базовые подходы, методы и модели представления и описания данных в логистике с помощью машинного обучения;
- современное состояние исследований в области машинного обучения в логистике.

уметь:

- проводить анализ предметной области;
- определять назначение, выбирать методы и средства для построения моделей машинного обучения;
- строить модели машинного обучения для решения логистических задач;
- интерпретировать результаты для дальнейшей оптимизации и использования результатов.

владеть:

- теоретическим аппаратом машинного анализа данных для логистики;
- методами реализации алгоритмов машинного обучения для логистики;
- навыками самостоятельной работы;
- практикой исследования и решения прикладных задач логистики с помощью машинного обучения.

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	Введение	2	2		1
2	Метод наименьших квадратов	4	4		2
3	Анализ временных рядов	2	2		2
4	Обучение с учителем	4	4		4
5	Кластеризация и частичное обучение	2	2		2
6	Обучение без учителя	2	2		2
7	Обучение с подкреплением	2	2		2
8	Глубокое обучение	4	4		4
9	Имитационное моделирование	2	2		2
10	Большие данные	2	2		3
11	Векторные представления текстов и графов	2	2		3
12	Распознавание изображений	2	2		3
Итого часов		30	30		30
Подготовка к экзамену		0 час.			
Общая трудоёмкость		90 час., 2 зач.ед.			

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

Семестр: 2 (Весенний)

1. Введение

Постановка задач обучения по прецедентам. Объекты и признаки. Типы шкал: бинарные, номинальные, порядковые, количественные. Типы задач: классификация, регрессия, прогнозирование, ранжирование. Основные понятия: модель алгоритмов, метод обучения, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль. Линейные модели регрессии и классификации. Метод наименьших квадратов. Полиномиальная регрессия. Примеры прикладных задач в логистике. Методика экспериментального исследования и сравнения алгоритмов на модельных и реальных данных. Критерии отбора моделей.

2. Метод наименьших квадратов

Основные типы данных. Метод наименьших квадратов. Регрессия на константу. Готовые формулы для парной регрессии. Множественная регрессия. Ошибка прогноза. Сумма квадратов остатков. Общая сумма квадратов. Объясненная сумма квадратов. Абсолютный ликбез по линейной алгебре. Геометрическая иллюстрация множественного МНК. Коэффициент детерминации. Многомерная линейная регрессия.

3. Анализ временных рядов

Задача прогнозирования временных рядов. Примеры приложений в логистике. Экспоненциальное скользящее среднее. Эконометрические, адаптивные, нейросетевые методы прогнозирования. Квантильная регрессия. Непараметрическая регрессия.

4. Обучение с учителем

Линейный, метрический, логический классификаторы. Метод опорных векторов. Ансамбли. Примеры в логистике.

5. Кластеризация и частичное обучение

Постановка задачи кластеризации. Типы кластерных структур. Постановка задачи Semisupervised Learning, примеры приложений. Оптимизационные постановки задач кластеризации и частичного обучения. Алгоритм k-средних и EM-алгоритм для разделения гауссовской смеси. Алгоритм DBSCAN. Агломеративная кластеризация, Алгоритм Ланса-Вильямса и его частные случаи. Алгоритм построения дендрограммы. Определение числа кластеров. Свойства сжатия/растяжения и монотонности. Простые эвристические методы частичного обучения: self-training, co-training, co-learning. Трансдуктивный метод опорных векторов TSVM. Алгоритм Expectation-Regularization на основе многоклассовой регуляризированной логистической регрессии.

6. Обучение без учителя

Нейронная сеть Кохонена. Конкурентное обучение, стратегии WTA и WTM. Самоорганизующаяся карта Кохонена. Применение для визуального анализа данных. Искусство интерпретации карт Кохонена. Автокодировщик. Линейный AE, SAE, DAE, CAE, RAE, VAE, AE для классификации, многослойный AE. Пред-обучение нейронных сетей (pre-training). Перенос обучения (transfer learning). Многозадачное обучение (multi-task learning). Самостоятельное обучение (self-supervised learning). Дистилляция моделей или суррогатное моделирование. Обучение с использованием привилегированной информации (learning using privileged information, LUPI). Генеративные состязательные сети (generative adversarial net, GAN).

7. Обучение с подкреплением

Задача о многоруком бандите. Жадные и эpsilon-жадные стратегии. Метод UCB (upper confidence bound). Адаптивные стратегии на основе скользящих средних. Метод сравнения с подкреплением. Метод преследования. Постановка задачи в случае, когда агент влияет на среду. Ценность состояния среды. Ценность действия. Жадные стратегии максимизации ценности. Уравнения оптимальности Беллмана. Метод SARSA. Метод Q-обучения. Типизация методов на on-policy и off-policy. Градиентная оптимизация стратегии (policy gradient). Связь с максимизацией log-правдоподобия. Постановка задачи при моделировании среды. Типизация методов на model-free и model-based. Контекстный многорукий бандит. Линейная регрессионная модель с верхней доверительной оценкой LinUCB. Оценивание новой стратегии по большим историческим данным, сформированным при старых стратегиях.

8. Глубокое обучение

Глубокие нейронные сети. Глубокое Q-обучение нейронной сети DQN на примере обучения играм Atari. Онлайнное глубокое обучение. Алгоритм Hedge BackProp. Глубокая структурированная семантическая модель DSSM (Deep Structured Semantic Model).

9. Имитационное моделирование

Постановки задачи имитационного моделирования. Основные этапы имитационного математического моделирования. Объекты с вероятностным характером поведения. Метод Монте-Карло.

10. Большие данные

Понятие больших данных. Принципы работы. MapReduce. Hadoop. Примеры задач. Парсинг данных.

11. Векторные представления текстов и графов

Векторные представления текста. Гипотеза дистрибутивной семантики. Модели CBOW и SGNS из программы word2vec. Иерархический SoftMax. Модель FastText. Векторные представления графов. Многомерное шкалирование (multidimensional scaling, MDS). Векторное представление соседства (stochastic neighbor embedding, SNE и tSNE). Матричные разложения (graph factorization). Модели случайных блужданий DeepWalk, node2vec. Обобщённый автокодировщик на графах GraphEDM. Представление о графовых нейронных сетях (graph neural network, GNN). Передача сообщений по графу (message passing).

12. Распознавание изображений

Современные модели для распознавания изображений. Практика.

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

Для обучения необходимо следующее мультимедийное оборудование:

Учебная компьютерная аудитория, персональные компьютеры, мультимедийное оборудование (проектор, экран, интерактивная доска).

6. Перечень рекомендуемой литературы

Основная литература

1. Эконометрика. Начальный курс [Текст] / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий ; Акад. народного хозяйства при Правительстве Рос. Федерации - М. Дело, 2007
2. Восстановление зависимостей по эмпирическим данным [Текст] / В. Н. Вапник, - М., Наука, 1979

1. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2014. — 739 p.
2. Bishop C. M. Pattern Recognition and Machine Learning. — Springer, 2006. — 738 p.
3. Мерков А. Б. Распознавание образов. Введение в методы статистического обучения. 2011. 256 с.
4. Мерков А. Б. Распознавание образов. Построение и обучение вероятностных моделей. 2014. 238 с.
5. Артамонов, Н.В. Введение в эконометрику: Учебник / 2-е изд., испр. и доп. - Москва : МЦНМО, 2014. - 222 с.: ISBN 978-5-4439-2010-8
6. Борзых, Д.А., Демешев Б.Б. Эконометрика в задачах и упражнениях : [более 300 задач]. — М.: ЛЕНАНД, 2014. — 202 с. : ил., табл.; 22 см.: ISBN 978-5-9710-1006-7
7. Коэльо Л.П., Ричарт В. Построение систем машинного обучения на языке Python. 2016. 302 с.
8. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: основы моделирования и первичная обработка данных. — М.: Финансы и статистика, 1983
9. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: исследование зависимостей. — М.: Финансы и статистика, 1985
10. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. — М.: Финансы и статистика, 1989

Дополнительная литература

1. Носко, В.П. Эконометрика [Текст] : учебник для студентов высших учебных заведений, обучающихся по экономическим специальностям. — М.: Дело, 2011. — 25 см. — (Академический учебник / Российская акад. нар. хоз-ва и гос. службы при Президенте Российской Федерации). Кн. 1: Кн. 1, ч. 1 : Основные понятия, элементарные методы, ч. 2 : Регрессионный анализ временных рядов. — 2011. — 671 с. : ил.; ISBN 978-5-7749-0654-3
2. Носко, В.П. Эконометрика [Текст] : учебник для студентов высших учебных заведений, обучающихся по экономическим специальностям. — М.: Дело, 2011. — 25 см. — (Академический учебник / Российская акад. нар. хоз-ва и гос. службы при Президенте Российской Федерации). Кн. 2: Кн. 2, ч. 3 : Системы одновременных уравнений, панельные данные, модели с дискретными и ограниченными объясняемыми переменными, ч. 4 : Временные ряды: дополнительные главы. Модель стохастической границы. — 2011. — 575 с. : ил., табл. ISBN 978-5-7749-0655-0
3. Носко, В.П. Эконометрика. Элементарные методы и введение в регрессионный анализ временных рядов. — М. : Ин-т экономики переход. периода, 2004. — 501 с. : ил., табл.; 23 см. ISBN 5-93255-141-0
4. Joshua D. Angrist, Jorn-Steffen Pischke. Mostly Harmless Econometrics: An Empiricist's Companion. — Princeton University Press, 2009.
5. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. — Springer, 2001. ISBN 0-387-95284-5
6. Vapnik V.N. Statistical learning theory. — N.Y.: John Wiley & Sons, Inc., 1998
7. Ширяев А.Н. Вероятность-1. Т.1, Изд. 7, стереотипное. — М.: МЦНМО, 2021. — 552 с. ISBN 978-5-4439-1557-9
8. Ширяев А.Н. Вероятность-2. Т.2, Изд. 7, стереотипное. — М.: МЦНМО, 2021. — 416 страниц. ISBN: 978-5-4439-1558-6
9. Умнов, А.Е. Аналитическая геометрия и линейная алгебра : учеб. Пособие. — 3-е изд., испр. и доп. — М.: МФТИ, 2011. — 544 с. ISBN: 978-5-7417-0387-6
10. Чернова Н.И. Математическая статистика: Учеб. пособие / Новосиб. гос. ун-т. Новосибирск, 2007. — 148 с. ISBN 978-5-94356-523-6

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

1. Электронная библиотека МФТИ: <http://lib.mipt.ru/>
2. Федеральный портал «Российское образование»: <http://www.edu.ru>
3. Библиотека по естественным наукам Российской академии наук: <http://benran.ru>

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

В процессе практических занятий используется программный пакет EViews 11 Student Version Lite.

В процессе обучения могут применяться технологии дистанционного тестирования в рамках системы дистанционного обучения МФТИ <http://moodle.physiech.edu/>

В образовательном процессе могут использоваться дистанционные занятия и вебинары с использованием коммуникационного программного обеспечения Zoom, сервиса видеотелефонной связи Google Meet, веб-сервиса Google Класс.

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

Успешное освоение курса требует большой самостоятельной работы студента. В программе курса приведено минимально необходимое время для работы студента над темой.

Самостоятельная работа включает в себя:

- чтение и конспектирование рекомендованной литературы;
- проработку учебного материала (по конспектам семинаров, учебной и научной литературе).

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Системный анализ и управление
профиль подготовки:	Системный анализ и управление в больших системах Физтех-школа Аэрокосмических Технологий кафедра логистических систем и технологий
курс:	<u>1</u>
квалификация:	магистр

Семестр, формы промежуточной аттестации: 2 (весенний) - Дифференцированный зачет

Разработчик: А.В. Пушкарева, ассистент

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-3 Способен формулировать, формировать и применять критерии оценки эффективности полученных результатов разработки систем управления и их внедрения в производственной и непроизводственной сферах	ОПК-3.1 Проводит анализ этапов разработки и внедрения систем управления, и оценивает эффективность результатов
	ОПК-3.2 Разрабатывает и применяет критерии оценки эффективности полученных результатов на основании полученных знаний
ОПК-4 Способен применять методы математического, функционального и системного анализа для задач моделирования, анализа и синтеза автоматического управления техническими объектами	ОПК-4.2 Проводит анализ и моделирование при помощи методов математического, функционального и системного анализа при решении прикладных и теоретических задач автоматического управления техническими объектами
ОПК-5 Способен выбирать методы и разрабатывать алгоритмы решения задач автоматического управления сложными управляемыми объектами	ОПК-5.1 Анализирует и определяет оптимальные методы для решения задач автоматического управления
	ОПК-5.2 Разрабатывает алгоритмы для решения задач автоматического управления

2. Показатели оценивания компетенций

В результате изучения дисциплины «Методы машинного обучения в логистических системах» обучающийся должен:

знать:

- базовые подходы, методы и модели представления и описания данных в логистике с помощью машинного обучения;
- современное состояние исследований в области машинного обучения в логистике.

уметь:

- проводить анализ предметной области;
- определять назначение, выбирать методы и средства для построения моделей машинного обучения;
- строить модели машинного обучения для решения логистических задач;
- интерпретировать результаты для дальнейшей оптимизации и использования результатов.

владеть:

- теоретическим аппаратом машинного анализа данных для логистики;
- методами реализации алгоритмов машинного обучения для логистики;
- навыками самостоятельной работы;
- практикой исследования и решения прикладных задач логистики с помощью машинного обучения.

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

Текущий контроль осуществляется в виде выполнения контрольной работы по отдельным разделам и выполнения домашних проектов. Для прохождения контроля студент должен, как минимум, продемонстрировать знания основных определений; умение решать стандартные задачи, разобранные на семинарских занятиях.

Текущий контроль включает 1 письменную контрольную работу в течение семестра, состоящую из нескольких вопросов и задач по пройденному материалу, а также выполнение домашних проектов для представления полученных знаний.

Реферат – форма изложения и интерпретации идей, содержащихся в нескольких источниках (рекомендуется использовать не менее 10), которая требует умения сопоставлять и анализировать различные точки зрения. Реферат, в отличие от конспекта, является новым, авторским текстом, что подразумевает новое изложение, систематизацию материала, особую авторскую позицию при сопоставлении различных точек зрения. Реферирование предполагает изложение какого-либо вопроса на основе классификации, обобщения, анализа и синтеза информации, полученной из нескольких литературных или интернет источников. Минимальный объем реферата 15 страниц, обязательно наличие заключения и выводов. Реферат оценивается по следующим критериям:

Авторская позиция:

- актуальность проблемы и темы;
- новизна и самостоятельность в постановке проблемы, в формулировании нового аспекта выбранной для анализа проблемы;
- наличие авторской позиции, самостоятельность суждений.

Степень раскрытия сущности проблемы:

- соответствие плана теме реферата;
- соответствие содержания теме и плану реферата;
- полнота и глубина раскрытия основных понятий проблемы;
- обоснованность способов и методов работы с материалом;
- умение работать с литературой, систематизировать и структурировать материал;
- наличие заключения и выводов;
- умение обобщать, сопоставлять различные точки зрения по рассматриваемому вопросу, аргументировать основные положения и выводы.

Обоснованность выбора источников:

- круг, полнота использования литературных источников по проблеме;
- привлечение новейших работ по проблеме (журнальные публикации, материалы сборников научных трудов и т.д.).

Соблюдение требований к оформлению:

- правильное оформление ссылок на используемую литературу;
- грамотность и культура изложения;
- владение терминологией и понятийным аппаратом проблемы;
- соблюдение требований к объему реферата;
- культура оформления: выделение абзацев.

Грамотность:

- отсутствие орфографических и синтаксических ошибок, стилистических погрешностей;
- отсутствие опечаток, сокращений слов, кроме общепринятых;
- литературный стиль.

Примерные темы рефератов:

1. Прогресс науки и роль машинного обучения в современном мире
2. Государственные и международные программы, направленные на реализацию принципов машинного обучения
3. Стратегии и подходы к использованию технологии графовых нейронных сетей
4. Свёрточные нейронные сети
5. Проблемы использования рекуррентных нейронных сетей
6. Методы Монте-Карло в логистике
7. Принципы работы алгоритма MapReduce
8. Гипотеза дистрибутивной семантики в анализе данных
9. Перспективы использования машинного обучения в современной логистике
10. Понятие больших данных в логистических системах
11. Простые эвристические методы частичного обучения в логистике

Критерии оценивания рефератов

9-10 баллов выставляется, если выполнены все требования к написанию и защите реферата: обозначена проблема и обоснована её актуальность, сделан краткий анализ различных точек зрения на рассматриваемую проблему и логично изложена собственная позиция, сформулированы выводы, тема раскрыта полностью, выдержан объём, соблюдены требования к внешнему оформлению, даны правильные ответы на дополнительные вопросы.

7-8 баллов выставляется, если основные требования к реферату и его защите выполнены, но при этом допущены недочёты; в частности, имеются неточности в изложении материала; отсутствует логическая последовательность в суждениях; не выдержан объём реферата; имеются упущения в оформлении; на дополнительные вопросы при защите даны неполные ответы.

4-6 баллов выставляется, если имеются существенные отступления от требований к реферированию; в частности: тема освещена лишь частично; допущены фактические ошибки в содержании реферата или при ответе на дополнительные вопросы; во время защиты отсутствует вывод.

1-3 балла выставляется, если тема реферата не раскрыта, выявлено существенное непонимание проблемы или же реферат не представлен вовсе.

Презентации должны быть подготовлены в редакторе Power Point или его аналогах и содержать не менее 10 слайдов (не включая титульный слайд), обязательным является наличие слайда с выводами. Слайды должны располагаться в логической последовательности. Информация должна быть грамотно и наглядно представлена с научной точки зрения в виде таблиц, графиков, схем и т.д., основана на объективных данных.

Перечень контрольных вопросов и заданий для самостоятельной работы и контрольных работ в течение семестра:

1. Основные типы данных. Метод наименьших квадратов. Формулы для оценки коэффициентов.
2. Ошибка прогноза. Сумма квадратов остатков. Общая сумма квадратов. Объясненная сумма квадратов. Коэффициент детерминации.
3. Множественная регрессия. Геометрическая иллюстрация множественного МНК.
4. Что такое обобщённый алгоритм классификации (надо помнить формулу)? Какие вы знаете частные случаи?
5. Метод k ближайших соседей. Преимущества и недостатки
6. Что такое окно переменной ширины, в каких случаях его стоит использовать?
7. Непараметрическая регрессия. Формула Надарая-Ватсона.
8. Что такое логическая закономерность? Приведите примеры в области логистики.
9. Сравните области статистических и логических закономерностей в (p,n) -плоскости.
10. Что такое решающее дерево? Достоинства и недостатки решающих деревьев.
11. Что такое случайный лес?
12. Что такое модель МакКаллока-Питтса (надо помнить формулу)?
13. Оптимальная разделяющая гиперплоскость. Понятие зазора между классами (margin).
14. Критерии отбора моделей.
15. Регуляризации для отбора признаков.

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

Перечень типовых вопросов и тем для проведения промежуточной аттестации обучающихся по итогам обучения:

1. Адаптивные модели
2. Нейросетевые модели анализа временных рядов

3. Постановка задачи кластеризации. Примеры прикладных задач. Типы кластерных структур.
4. Постановка задачи Semisupervised Learning, примеры приложений.
5. Алгоритм построения дендрограммы. Определение числа кластеров.
6. Свойства сжатия/растяжения и монотонности. Простые эвристические методы частичного обучения: self-training, co-training, co-learning.
7. Трансдуктивный метод опорных векторов TSVM. Алгоритм Expectation-Regularization на основе многоклассовой регуляризированной логистической регрессии.
8. Функции активации ReLU и PReLU. Проблема «паралича» сети.
9. Метод обратного распространения ошибок. Основная идея. Основные недостатки и способы их устранения.
10. Свёрточные нейронные сети (CNN).
11. Рекуррентные нейронные сети (RNN).
12. Задача о многоруком бандите. Жадные и эпсилон-жадные стратегии. Метод UCB (upper confidence bound). Стратегия Softmax.
13. Как устроена самоорганизующаяся карта Кохонена?
14. Постановки задачи имитационного моделирования. Основные этапы имитационного математического моделирования. Объекты с вероятностным характером поведения.
15. Метод Монте-Карло
16. Понятие больших данных. Принципы работы. MapReduce. Hadoop.
17. Оценивание новой стратегии по большим историческим данным.
18. Векторные представления текста. Гипотеза дистрибутивной семантики. Модели CBOW и SGNS из программы word2vec.
19. Постановка задачи при наличии информации о среде в случае выбора действия. Контекстный многорукий бандит.
20. Иерархический SoftMax. Модель FastText.
21. Векторные представления графов. Многомерное шкалирование (multidimensional scaling, MDS).
22. Векторное представление соседства (stochastic neighbor embedding, SNE и tSNE). Матричные разложения (graph factorization).
23. Представление о графовых нейронных сетях (graph neural network, GNN). Передача сообщений по графу (message passing).
24. Основные типы данных. Метод наименьших квадратов. Формулы для оценки коэффициентов.
25. Ошибка прогноза. Сумма квадратов остатков. Общая сумма квадратов. Объясненная сумма квадратов. Коэффициент детерминации.
26. Множественная регрессия. Геометрическая иллюстрация множественного МНК.
27. Что такое обобщённый алгоритм классификации (надо помнить формулу)? Какие вы знаете частные случаи?
28. Метод k ближайших соседей. Преимущества и недостатки
29. Что такое окно переменной ширины, в каких случаях его стоит использовать?
30. Непараметрическая регрессия. Формула Надарая-Ватсона.
31. Что такое логическая закономерность? Приведите примеры в области логистики.
32. Сравните области статистических и логических закономерностей в (p,n)-плоскости.
33. Что такое решающее дерево? Достоинства и недостатки решающих деревьев.
34. Что такое случайный лес?
35. Что такое модель МакКаллока-Питтса (надо помнить формулу)?
36. Оптимальная разделяющая гиперплоскость. Понятие зазора между классами (margin).
37. Критерии отбора моделей.
38. Регуляризации для отбора признаков.
39. Адаптивные модели
40. Нейросетевые модели анализа временных рядов
41. Постановка задачи кластеризации. Примеры прикладных задач. Типы кластерных структур.
42. Постановка задачи Semisupervised Learning, примеры приложений.
43. Алгоритм построения дендрограммы. Определение числа кластеров.
44. Свойства сжатия/растяжения и монотонности. Простые эвристические методы частичного обучения: self-training, co-training, co-learning.

45. Трансдуктивный метод опорных векторов TSVM. Алгоритм Expectation-Regularization на основе многоклассовой регуляризированной логистической регрессии.
46. Функции активации ReLU и PReLU. Проблема «паралича» сети.
47. Метод обратного распространения ошибок. Основная идея. Основные недостатки и способы их устранения.
48. Свёрточные нейронные сети (CNN).
49. Рекуррентные нейронные сети (RNN).

Примеры билетов для проведения дифференцированного зачёта:

Билет 1.

- 1) Метод Монте-Карло
- 2) Как устроена самоорганизующаяся карта Кохонена?

Билет 2.

- 1) Векторные представления текста. Гипотеза дистрибутивной семантики. Модели CBOW и SGNS из программы word2vec.
- 2) Оценивание новой стратегии по большим историческим данным.

Критерии оценивания

Отлично 10 - систематизированные, глубокие и полные знания по всем разделам учебной программы, а также по основным вопросам, выходящим за ее пределы;

- точное использование научной терминологии (в том числе на иностранном языке), стилистически грамотное, логически правильное изложение ответа на вопросы;
- безупречное владение инструментарием учебной дисциплины, умение его эффективно использовать в постановке и решении научных и профессиональных задач;
- выраженная способность самостоятельно и творчески решать сложные проблемы в нестандартной ситуации;
- полное и глубокое усвоение основной и дополнительной литературы, рекомендованной учебной программой дисциплины;
- умение ориентироваться в теориях, концепциях и направлениях по изучаемой дисциплине и давать им критическую оценку, использовать научные достижения других дисциплин;
- творческая самостоятельная работа на практических, лабораторных занятиях, активное участие в групповых обсуждениях, высокий уровень культуры исполнения заданий.

Отлично 9 - систематизированные, глубокие и полные знания по всем разделам учебной программы;

- точное использование научной терминологии (в том числе на иностранном языке), стилистически грамотное, логически правильное изложение ответа на вопросы;
- владение инструментарием учебной дисциплины, умение его эффективно использовать в постановке и решении научных и профессиональных задач;
- способность самостоятельно и творчески решать сложные проблемы в нестандартной ситуации в рамках учебной программы, полное усвоение основной и дополнительной литературы, рекомендованной учебной программой дисциплины;
- умение ориентироваться в основных теориях, концепциях и направлениях по изучаемой дисциплине и давать им критическую оценку;
- самостоятельная работа на практических, лабораторных занятиях, творческое участие в групповых обсуждениях, высокий уровень культуры исполнения заданий.

Отлично 8 - систематизированные, глубокие и полные знания по всем поставленным вопросам в объеме учебной программы;

- использование научной терминологии, стилистически грамотное, логически правильное изложение ответа на вопросы, умение делать обоснованные выводы;
- владение инструментарием учебной дисциплины (методами комплексного анализа, техникой информационных технологий), умение его использовать в постановке и решении научных и профессиональных задач; способность самостоятельно решать сложные проблемы в рамках учебной программы;

- усвоение основной и дополнительной литературы, рекомендованной учебной программой дисциплины;
- умение ориентироваться в основных теориях, концепциях и направлениях по изучаемой дисциплине и давать им критическую оценку с позиций государственной идеологии (по дисциплинам социально-гуманитарного цикла);
- активная самостоятельная работа на практических, лабораторных занятиях, систематическое участие в групповых обсуждениях, высокий уровень культуры исполнения заданий.

Хорошо 7 - систематизированные, глубокие и полные знания по всем разделам учебной программы;

- использование научной терминологии (в том числе на иностранном языке), лингвистически и логически правильное изложение ответа на вопросы, умение делать обоснованные выводы;
- владение инструментарием учебной дисциплины, умение его использовать в постановке и решении научных и профессиональных задач;
- усвоение основной и дополнительной литературы, рекомендованной учебной программой дисциплины;
- умение ориентироваться в основных теориях, концепциях и направлениях по изучаемой дисциплине и давать им критическую оценку;
- самостоятельная работа на практических, лабораторных занятиях, участие в групповых обсуждениях, высокий уровень культуры исполнения заданий.

Хорошо 6 - достаточно полные и систематизированные знания в объеме учебной программы;

- использование необходимой научной терминологии, стилистически грамотное, логически правильное изложение ответа на вопросы, умение делать обоснованные выводы;
- владение инструментарием учебной дисциплины, умение его использовать в решении учебных и профессиональных задач; способность самостоятельно применять типовые решения в рамках учебной программы;
- усвоение основной литературы, рекомендованной учебной программой дисциплины;
- умение ориентироваться в базовых теориях, концепциях и направлениях по изучаемой дисциплине и давать им сравнительную оценку; активная самостоятельная работа на практических, лабораторных занятиях, периодическое участие в групповых обсуждениях, высокий уровень культуры исполнения заданий.

Хорошо 5 - достаточные знания в объеме учебной программы;

- использование научной терминологии, стилистически грамотное, логически правильное изложение ответа на вопросы, умение делать выводы;
- владение инструментарием учебной дисциплины, умение его использовать в решении учебных и профессиональных задач;
- способность самостоятельно применять типовые решения в рамках учебной программы;
- усвоение основной литературы, рекомендованной учебной программой дисциплины;
- умение ориентироваться в базовых теориях, концепциях и направлениях по изучаемой дисциплине и давать им сравнительную оценку;
- самостоятельная работа на практических, лабораторных занятиях, участие в групповых обсуждениях, высокий уровень культуры исполнения заданий.

Удовлетворительно 4 - достаточный объем знаний в рамках образовательного стандарта;

- усвоение основной литературы, рекомендованной учебной программой дисциплины;
- использование научной терминологии, стилистическое и логическое изложение ответа на вопросы, умение делать выводы без существенных ошибок;
- владение инструментарием учебной дисциплины, умение его использовать в решении стандартных (типовых) задач;
- умение под руководством преподавателя решать стандартные (типовые) задачи;
- умение ориентироваться в основных теориях, концепциях и направлениях по изучаемой дисциплине и давать им оценку;
- работа под руководством преподавателя на практических, лабораторных занятиях, допустимый уровень культуры исполнения заданий.

Удовлетворительно 3 - недостаточно полный объем знаний в рамках образовательного стандарта;

- знание части основной литературы, рекомендованной учебной программой дисциплины;
- использование научной терминологии, изложение ответа на вопросы с существенными лингвистическими и логическими ошибками;
- слабое владение инструментарием учебной дисциплины, некомпетентность в решении стандартных (типовых) задач;
- неумение ориентироваться в основных теориях, концепциях и направлениях изучаемой дисциплины;
- пассивность на практических и лабораторных занятиях, низкий уровень культуры исполнения заданий.

Неудовлетворительно 2 - фрагментарные знания в рамках образовательного стандарта;

- знания отдельных литературных источников, рекомендованных учебной программой дисциплины;
- неумение использовать научную терминологию дисциплины, наличие в ответе грубых стилистических и логических ошибок;
- пассивность на практических и лабораторных занятиях, низкий уровень культуры исполнения заданий.

Неудовлетворительно 1 - отсутствие знаний и компетенций в рамках образовательного стандарта или отказ от ответа.

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

Дифференцированный зачет по дисциплине является заключительным этапом изучения всего курса и имеет целью проверку знаний студентов по теории и выявление навыков их применения при решении практических задач, а также навыков самостоятельной работы с рекомендованными основной и дополнительной литературой. Оценка студента за дифференцированный зачет рассчитывается как среднее арифметическое всех оценок, полученных студентами в течение семестра за сдачу контрольной работы, выполнению проектной работы, а также итоговое собеседование по всему пройденному курсу. Учитывается также работа студента на практических занятиях (подготовка презентаций и докладов) и написание реферата по заданной теме.

Для успешной подготовки к дифференцированному зачету обучающемуся следует тщательно разобрать конспекты лекций, источники литературы и также материалы практических занятий. Основное в подготовке - повторение всего учебного материала дисциплины, по которому необходимо сдавать зачет. Для подготовки к дифференцированному зачету рекомендуется опираться на вопросы, представленные в рабочей учебной программе.

Во время проведения дифференцированного зачета обучающиеся могут пользоваться программой дисциплины, а также справочной литературой.

Обучающийся должен проявить всестороннее, систематическое и глубокое знание учебно-программного материала, умение свободно выполнять задания, предусмотренные программой, усвоить основную литературу и быть знакомым с дополнительной литературой, рекомендованной программой дисциплины, усвоить взаимосвязь основных понятий дисциплины в их значении для приобретаемой профессии.