

**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:	Основы машинного обучения
по направлению:	Системный анализ и управление
профиль подготовки:	Системный анализ и управление в технических, экономических и социальных системах Физтех-школа Аэрокосмических Технологий кафедра логистических систем и технологий
курс:	4
квалификация:	бакалавр

Семестр, формы промежуточной аттестации: 8 (весенний) - Дифференцированный зачет

Аудиторных часов: 60 всего, в том числе:

лекции: 0 час.

семинары: 60 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 30 час.

Всего часов: 90, всего зач. ед.: 2

Количество контрольных работ, заданий: 2

Программу составил: А.В. Пушкарева

Программа обсуждена на заседании кафедры логистических систем и технологий 04.06.2020

Аннотация

В курсе рассматриваются основы машинного обучения и анализа данных, включая дискриминантный, кластерный и регрессионный анализ, освоение технологий извлечения информации из неструктурированных или слабоструктурированных систем. В процессе обучения студенты овладевают навыками практического решения задач интеллектуального анализа данных, а также осваивают методы обработки данных, подходов, методов и моделей извлечения информации с целью анализа динамики процессов различной природы и повышения точности построения прогнозов. У обучающихся формируются навыки по интерпретации полученных результатов для оптимизации моделей и объяснения возникающих в макросистемах эффектов.

Курс носит прикладной характер и ориентирован на практическую проработку студентами полученных знаний и навыков.

1. Цели и задачи

Цель дисциплины

- изучение основ машинного обучения и анализа данных, включая дискриминантный, кластерный и регрессионный анализ, освоение технологий извлечения информации из неструктурированных или слабоструктурированных систем;
- овладение навыками практического решения задач интеллектуального анализа данных.

Задачи дисциплины

- освоение студентами обработки данных, подходов, методов и моделей извлечения информации с целью анализа динамики процессов различной природы и повышения точности построения прогнозов;
- приобретение в условиях слабой структурированности данных практических навыков извлечения знаний;
- формирование навыков по интерпретации полученных результатов для оптимизации моделей и объяснения возникающих в макросистемах эффектов.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
ОПК-1 Способен анализировать задачи управления в технических системах на основе приобретенных знаний	ОПК-1.1 Осуществляет декомпозицию задачи управления, выделяет базовые составляющие задачи
	ОПК-1.2 Рассматривает возможные варианты решения задачи управления в технических системах, оценивает их достоинства и недостатки
ОПК-7 Способен принимать научно обоснованные решения в области системного анализа и автоматического управления на основе математических и естественно-научных дисциплин	ОПК-7.2 Способен обосновать причинно-следственные отношения используемых понятий и моделей
	ОПК-7.3 Умеет находить ключевые параметры, определяющие изучаемый процесс
ПК-1 Способен проводить исследование систем управления и их компонент	ПК-1.1 Владеет фундаментальными понятиями, законами и теориями системного анализа
	ПК-1.2 Имеет глубокое знание и понимание базовых математических дисциплин

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- базовые подходы, методы и модели представления и описания технологий машинного обучения;
- современное состояние исследований в области машинного обучения.

уметь:

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

владеть:

- теоретическим аппаратом базового интеллектуального анализа данных;
- методами реализации алгоритмов машинного обучения;
- навыками самостоятельной работы;
- практикой исследования и решения прикладных задач интеллектуального анализа данных.

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 и примеры прикладных задач		6		3
2	Метрические методы классификации		8		3
3	Логические методы классификации		8		4
4	Линейные методы классификации		4		2
5	Метод опорных векторов		4		2
6	Регрессионный анализ		4		2
7	Прогнозирование временных рядов		4		2
8	Критерии выбора моделей и методы отбора признаков		4		2
9	Кластеризация и частичное обучение		4		2
10	Нейронные сети		4		2
11	Линейные композиции, бустинг		4		2
12	Эвристические, стохастические, нелинейные композиции		4		2
13	Обучение с подкреплением		2		2
Итого часов			60		30
Подготовка к экзамену		0 час.			
Общая трудоёмкость		90 час., 2 зач.ед.			

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

1. Основные понятия и примеры прикладных задач

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. Методика экспериментального исследования и сравнения алгоритмов на модельных и реальных данных. Конкурсы по анализу данных [kaggle.com](https://www.kaggle.com/). Полигон алгоритмов классификации. CRISP-DM — межотраслевой стандарт ведения проектов интеллектуального анализа данных.

2. Метрические методы классификации

Гипотезы компактности и непрерывности. Обобщённый метрический классификатор. Метод ближайших соседей kNN и его обобщения. Подбор числа k по критерию скользящего контроля. Метод окна Парзена с постоянной и переменной шириной окна. Метод потенциальных функций и его связь с линейной моделью классификации. Непараметрическая регрессия. Локально взвешенный метод наименьших квадратов. Ядерное сглаживание. Оценка Надарая-Ватсона с постоянной и переменной шириной окна. Выбор функции ядра. Задача отсева выбросов. Робастная непараметрическая регрессия. Алгоритм LOWESS.

3. Логические методы классификации

Понятие логической закономерности. Параметрические семейства закономерностей: конъюнкции пороговых правил, синдромные правила, шары, гиперплоскости. Переборные алгоритмы синтеза конъюнкций: стохастический локальный поиск, стабилизация, редукция. Двухкритериальный отбор информативных закономерностей, парето-оптимальный фронт в (p, n) -пространстве. Решающее дерево. Жадная нисходящая стратегия «разделяй и властвуй». Алгоритм ID3. Недостатки жадной стратегии и способы их устранения. Проблема переобучения. Вывод критериев ветвления. Мера нечистоты (impurity) распределения. Энтропийный критерий, критерий Джини. Редукция решающих деревьев: предредукция и постредукция. Алгоритм C4.5. Деревья регрессии. Алгоритм CART. Небрежные решающие деревья (oblivious decision tree). Решающий лес. Случайный лес (Random Forest).

4. Линейные методы классификации

Линейный классификатор, модель МакКаллока-Питтса, непрерывные аппроксимации пороговой функции потерь. Метод стохастического градиента SG. Метод стохастического среднего градиента SAG. Эвристики: инициализация весов, порядок предъявления объектов, выбор величины градиентного шага, «выбивание» из локальных минимумов. Проблема мультиколлинеарности и переобучения, регуляризация или редукция весов (weight decay). Вероятностная постановка задачи классификации. Принцип максимума правдоподобия. Вероятностная интерпретация регуляризации, совместное правдоподобие данных и модели. Принцип максимума апостериорной вероятности. Гауссовский и лапласовский регуляризаторы. Логистическая регрессия. Принцип максимума правдоподобия и логарифмическая функция потерь. Метод стохастического градиента для логарифмической функции потерь. Многоклассовая логистическая регрессия. Регуляризованная логистическая регрессия. Калибровка Платта.

5. Метод опорных векторов

Оптимальная разделяющая гиперплоскость. Понятие зазора между классами (margin). Случаи линейной делимости и отсутствия линейной делимости. Связь с минимизацией регуляризованного эмпирического риска. Кусочно-линейная функция потерь. Задача квадратичного программирования и двойственная задача. Понятие опорных векторов. Рекомендации по выбору константы C . Функция ядра (kernel functions), спрямляющее пространство, теорема Мерсера. Способы конструктивного построения ядер. Примеры ядер. SVM-регрессия. Регуляризации для отбора признаков: LASSO SVM, Elastic Net SVM, SFM, RFM.

6. Регрессионный анализ

Задача регрессии, многомерная линейная регрессия. Метод наименьших квадратов, его вероятностный смысл и геометрический смысл. Сингулярное разложение. Проблемы мультиколлинеарности и переобучения. Регуляризация. Гребневая регрессия через сингулярное разложение. Методы отбора признаков: Лассо Тибширани, Elastic Net, сравнение с гребневой регрессией. Метод главных компонент и декоррелирующее преобразование Карунена-Лоэва, его связь с сингулярным разложением. Спектральный подход к решению задачи наименьших квадратов. Задачи и методы низкоранговых матричных разложений. Нелинейная регрессия.

7. Прогнозирование временных рядов

Задача прогнозирования временных рядов. Примеры приложений. Экспоненциальное скользящее среднее. Модель Хольта. Модель Тейла-Вейджа. Модель Хольта-Уинтерса. Адаптивная авторегрессионная модель. Следящий контрольный сигнал. Модель Тригга-Лича. Адаптивная селективная модель. Адаптивная композиция моделей. Локальная адаптация весов с регуляризацией.

8. Критерии выбора моделей и методы отбора признаков

Критерии качества классификации: чувствительность и специфичность, ROC-кривая и AUC, точность и полнота, AUC-PR. Внутренние и внешние критерии. Эмпирические и аналитические критерии. Скользящий контроль, разновидности эмпирических оценок скользящего контроля. Критерий непротиворечивости. Разновидности аналитических оценок. Регуляризация. Критерий Акаике (AIC). Байесовский информационный критерий (BIC). Оценка Вапника-Червоненкиса. Агрегированные и многоступенчатые критерии. Сложность задачи отбора признаков. Полный перебор. Метод добавления и удаления, шаговая регрессия. Поиск в глубину, метод ветвей и границ. Усечённый поиск в ширину, многорядный итерационный алгоритм МГУА. Генетический алгоритм, его сходство с МГУА.

9. Кластеризация и частичное обучение

Постановка задачи кластеризации. Примеры прикладных задач. 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. Графовые алгоритмы кластеризации. Выделение связанных компонент. Кратчайший незамкнутый путь. Алгоритм ФОРЭЛ. Алгоритм DBSCAN. Агломеративная кластеризация, Алгоритм Ланса-Вильямса и его частные случаи. Алгоритм построения дендрограммы. Определение числа кластеров. Свойства сжатия/растяжения, монотонности и редуктивности. 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.

10. Нейронные сети

Биологический нейрон, модель МакКаллока-Питтса как линейный классификатор. Функции активации. Проблема полноты. Задача исключаящего или. Полнота двухслойных сетей в пространстве булевых функций. Алгоритм обратного распространения ошибок. Быстрые методы стохастического градиента: Поляка, Нестерова, AdaGrad, RMSProp, AdaDelta, Adam, Nadam, диагональный метод Левенберга-Марквардта. Проблема взрыва градиента и эвристика gradient clipping. Метод случайных отключений нейронов (Dropout). Интерпретации Dropout. Обратный Dropout и L2-регуляризация. Функции активации ReLU и PReLU. Проблема «паралича» сети. Эвристики для формирования начального приближения. Метод послойной настройки сети. Подбор структуры сети: методы постепенного усложнения сети, оптимальное прореживание нейронных сетей (optimal brain damage).

Нейронные сети глубокого обучения. Свёрточные нейронные сети (CNN). Свёрточный нейрон. Pooling нейрон. Выборка размеченных изображений ImageNet. Идея обобщения CNN на любые структурированные данные. Рекуррентные нейронные сети (RNN). Обучение рекуррентных сетей: Backpropagation Through Time (BPTT). Сети долгой кратковременной памяти (Long short-term memory, LSTM). Автокодировщики. Векторные представления дискретных данных.

11. Линейные композиции, бустинг

Основные понятия: базовый алгоритм (алгоритмический оператор), корректирующая операция.

Взвешенное голосование. Алгоритм AdaBoost. Экспоненциальная аппроксимация пороговой функции потерь. Процесс последовательного обучения базовых алгоритмов. Теорема о сходимости бустинга. Обобщающая способность бустинга. Базовые алгоритмы в бустинге. Решающие пни. Градиентный бустинг. Стохастический градиентный бустинг. Простое голосование (комитет большинства). Алгоритм ComBoost. Идентификация нетипичных объектов (выбросов). Преобразование простого голосования во взвешенное. Обобщение на большое число классов.

12. Эвристические, стохастические, нелинейные композиции

Стохастические методы: бэггинг и метод случайных подпространств. Случайный лес. Анализ смещения и вариации для простого голосования. Смесь алгоритмов (квазилинейная композиция), область компетентности, примеры функций компетентности. Выпуклые функции потерь. Методы построения смесей: последовательный и иерархический. Построение смеси алгоритмов с помощью ЕМ-подобного алгоритма.

13. Обучение с подкреплением

Задача о многоруком бандите. Жадные и эpsilon-жадные стратегии. Метод UCB (upper confidence bound). Стратегия Softmax. Среда для экспериментов. 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 временных разностей TD. Метод Q-обучения. Градиентная оптимизация стратегии (policy gradient). Связь с максимизацией log-правдоподобия. Постановка задачи при наличии информации о среде в случае выбора действия. Контекстный многорукий бандит. Линейная регрессионная модель с верхней доверительной оценкой LinUCB. Оценивание новой стратегии по большим историческим данным.

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

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

6. Перечень рекомендуемой литературы

Основная литература

1. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: основы моделирования и первичная обработка данных. — М.: Финансы и статистика, 1983
2. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: исследование зависимостей. — М.: Финансы и статистика, 1985
3. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. — М.: Финансы и статистика, 1989
4. Вапник В. Н., Червоненкис А. Я. Теория распознавания образов. — М.: Наука, 1974
5. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. — М.: Наука, 1979
6. Журавлев Ю. И., Рязанов В. В., Сенько О. В. «Распознавание». Математические методы. Программная система. Практические применения. — М.: Фазис, 2006. ISBN 5-7036-0108-8
7. Загоруйко, Н. Г. Когнитивный анализ данных. — Новосибирск: Академическое издательство «ГЕО», 2012. ISBN 978-5-906284-04-4
8. Загоруйко Н. Г. Прикладные методы анализа данных и знаний. — Новосибирск: ИМ СО РАН, 1999. ISBN 5-86134-060-9
9. Мерков, А. Б. Распознавание образов. Введение в методы статистического обучения. — Москва: Едиториал УРСС, 2011. ISBN 978-5-354-01337-1
10. Jiawei Han, Micheline Kamber, Jian Pei Data Mining: Concepts and Techniques, Third Edition. — Morgan Kaufmann Publishers, 2012. ISBN 978-0123814791
11. Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning, 2nd edition. — Springer, 2009. ISBN 978-0387848570
12. Bishop, C.M. Pattern Recognition and Machine Learning. — Springer, 2006. ISBN 978-0-387-31073-2

Дополнительная литература

1. Шлезингер М., Главач В. Десять лекций по статистическому и структурному распознаванию. — Киев: Наукова думка, 2004. ISBN 966-00-0341 -2
2. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. — Springer, 2001. ISBN 0-387-95284-5
3. MacKay D. On-line book: Information Theory, Inference, and Learning Algorithms. — 2005
4. Mitchell T. Machine Learning. — McGraw-Hill Science/Engineering/Math, 1997. ISBN 0-07-042807-7
5. Шлезингер М., Главач В. Десять лекций по статистическому и структурному распознаванию. — Киев: Наукова думка, 2004. ISBN 966-00-0341 -2
- Schölkopf B., Smola A.J. Learning with Kernels. Support Vector Machines, Regularization, Optimization, and Beyond. — MIT Press, Cambridge, MA, 2002 ISBN 13-978-0-262-19475-4 [2]
6. Vapnik V.N. Statistical learning theory. — N.Y.: John Wiley & Sons, Inc., 1998
7. Witten I.H., Frank E. Data Mining: Practical Machine Learning Tools and Techniques (Second Edition). — Morgan Kaufmann, 2005 ISBN 0-12-088407-0

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

1. Федеральный портал «Российское образование»: <http://www.edu.ru>
2. 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 Anaconda с предустановленным пакетом Python 2.7, 3.6 или более поздней версией

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

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

Успешное освоение курса «Основы машинного обучения» 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).

ПРИЛОЖЕНИЕ

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Системный анализ и управление
профиль подготовки:	Системный анализ и управление в технических, экономических и социальных системах Физтех-школа Аэрокосмических Технологий кафедра логистических систем и технологий
курс:	<u>4</u>
квалификация:	бакалавр
Семестр, формы промежуточной аттестации: 8 (весенний) - Дифференцированный зачет	
Разработчик:	А.В. Пушкарева

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
ОПК-1 Способен анализировать задачи управления в технических системах на основе приобретенных знаний	ОПК-1.1 Осуществляет декомпозицию задачи управления, выделяет базовые составляющие задачи
	ОПК-1.2 Рассматривает возможные варианты решения задачи управления в технических системах, оценивает их достоинства и недостатки
ОПК-7 Способен принимать научно обоснованные решения в области системного анализа и автоматического управления на основе математических и естественно-научных дисциплин	ОПК-7.2 Способен обосновать причинно-следственные отношения используемых понятий и моделей
	ОПК-7.3 Умеет находить ключевые параметры, определяющие изучаемый процесс
ПК-1 Способен проводить исследование систем управления и их компонент	ПК-1.1 Владеет фундаментальными понятиями, законами и теориями системного анализа
	ПК-1.2 Имеет глубокое знание и понимание базовых математических дисциплин
	ПК-1.3 Владеет культурой постановки научной задачи и моделирования объектов и систем

2. Показатели оценивания компетенций

В результате изучения дисциплины «Основы машинного обучения» обучающийся должен:

знать:

- базовые подходы, методы и модели представления и описания технологий машинного обучения;
- современное состояние исследований в области машинного обучения.

уметь:

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

владеть:

- теоретическим аппаратом базового интеллектуального анализа данных;
- методами реализации алгоритмов машинного обучения;
- навыками самостоятельной работы;
- практикой исследования и решения прикладных задач интеллектуального анализа данных.

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

Текущий контроль по дисциплине «Основы машинного обучения» осуществляется в виде выполнения контрольной работы по отдельным разделам и в форме защиты самостоятельно выполненной работы.

Примерный перечень тем:

- обзор методов машинного обучения;
- методы отбора признаков;
- алгоритм back propagation;

Минимальный объем работы составляет 15 страниц, обязательно наличие заключения и выводов.

Самостоятельная работа 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 работы;
- соответствие содержания теме и плану работы;
- 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 работы;
- культура оформления: выделение абзацев.

Грамотность

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

Критерии оценивания работы:

9-10 баллов выставляется, если выполнены все требования к написанию и защите работы: 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 работе и его защите выполнены, но при этом допущены недочёты; в частности, имеются неточности в изложении материала; отсутствует логическая последовательность в суждениях; не выдержан объём работы; 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 работы или при ответе на дополнительные вопросы; во время защиты отсутствует вывод.

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

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

Примерный перечень контрольных вопросов:

- 1) Что такое обобщённый алгоритм классификации (надо помнить формулу)? Какие вы знаете частные случаи?
- 2) Метод k ближайших соседей. Преимущества и недостатки
- 3) Что такое окно переменной ширины, в каких случаях его стоит использовать?
- 4) Что такое метод потенциальных функций? Идея алгоритма настройки.
- 5) Приведите пример метрического алгоритма классификации, который одновременно является линейным классификатором.
- 6) Непараметрическая регрессия. Формула Надарая-Ватсона.
- 7) Как влияет выбор ядра и ширины окна сглаживания на результат аппроксимации?
- 8) Что такое логическая закономерность? Приведите примеры закономерностей в задаче кредитного скоринга.
- 9) Часто используемые типы логических закономерностей.
- 10) Дайте определение статистической закономерности (помнить формулы).
- 11) Сравните области статистических и логических закономерностей в (p,n) -плоскости.
- 12) Что такое решающее дерево? Достоинства и недостатки решающих деревьев.
- 13) Жадный метод обучения решающего дерева.
- 14) Зачем делается редукция решающих деревьев? Каким образом ее сделать?
- 15) Что такое случайный лес?
- 16) Что такое модель МакКаллока-Питтса (надо помнить формулу)?
- 17) Метод стохастического градиента. Описать алгоритм.
- 18) Недостатки метода SG и как с ними бороться?
- 19) Что такое правило Хэбба?
- 20) Что такое линейный адаптивный элемент ADALINE?
- 21) Что такое «сокращение весов»?
- 22) Принцип максимума правдоподобия. Логарифм правдоподобия для случая 2-х классов. Связь правдоподобия и аппроксимации эмпирического риска.
- 23) Оптимальная разделяющая гиперплоскость. Понятие зазора между классами (margin).
- 24) Случаи линейной разделимости и отсутствия линейной разделимости. Связь с минимизацией регуляризованного эмпирического риска. Кусочно-линейная функция потерь.
- 25) Задача квадратичного программирования и двойственная задача. Понятие опорных векторов.
- 26) Способы конструктивного построения ядер. Примеры ядер.
- 27) SVM-регрессия.
- 28) Регуляризации для отбора признаков: LASSO SVM, Elastic Net SVM, SFM, RFM.
- 29) Задача регрессии, многомерная линейная регрессия.
- 30) Метод наименьших квадратов, его вероятностный смысл и геометрический смысл.
- 31) Сингулярное разложение.
- 32) Проблемы мультиколлинеарности и переобучения.
- 33) Регуляризация. Гребневая регрессия через сингулярное разложение.
- 34) Elastic Net, сравнение с гребневой регрессией.
- 35) Задачи и методы низкоранговых матричных разложений.
- 36) Метод Ньютона-Рафсона, метод Ньютона-Гаусса.
- 37) Экспоненциальное скользящее среднее. Модель Хольта. Модель Тейла-Вейджа. Модель Хольта-Уинтерса.

- 38) Адаптивная авторегрессионная модель.
- 39) Следящий контрольный сигнал. Модель Тригга-Лича.
- 40) Адаптивная селективная модель. Адаптивная композиция моделей.
- 41) Локальная адаптация весов с регуляризацией.
- 42) Что такое точность, полнота, чувствительность, специфичность? Зачем они нужны?
- 43) Что такое ROC-кривая, как она определяется?
- 44) В чём отличия внутренних и внешних критериев? Какие есть критерии скользящего контроля?
- 45) Основная идея отбора признаков методом полного перебора. Действительно ли это полный перебор?
- 46) Основная идея отбора признаков методом добавлений и исключений.
- 47) Что такое МГУА?
- 48) Основная идея отбора признаков с помощью генетического алгоритма.
- 49) В чём отличия случайного поиска от случайного поиска с адаптацией?
- 50) Постановка задачи кластеризации. Примеры прикладных задач. Типы кластерных структур.
- 51) Постановка задачи Semisupervised Learning, примеры приложений.
- 52) Графовые алгоритмы кластеризации. Выделение связанных компонент. Кратчайший незамкнутый путь.
- 53) Алгоритм DBSCAN.
- 54) Алгоритм построения дендрограммы. Определение числа кластеров.
- 55) Простые эвристические методы частичного обучения: self-training, co-training, co-learning.
- 56) Трансдуктивный метод опорных векторов TSVM.
- 57) Приведите пример выборки, которую невозможно классифицировать без ошибок с помощью линейного алгоритма классификации. Какова минимальная длина выборки, обладающая данным свойством? Какие существуют способы модифицировать линейный алгоритм так, чтобы данная выборка стала линейно разделимой?
- 58) Почему любая булева функция представима в виде нейронной сети? Сколько в ней слоёв?
- 59) Метод обратного распространения ошибок. Основная идея. Основные недостатки и способы их устранения.
- 60) Функции активации ReLU и PReLU. Проблема «паралича» сети.
- 61) Свёрточные нейронные сети (CNN).
- 62) Рекуррентные нейронные сети (RNN).
- 63) Обучение рекуррентных сетей: Backpropagation Through Time (BPTT).
- 64) Взвешенное голосование. Алгоритм AdaBoost.
- 65) Процесс последовательного обучения базовых алгоритмов.
- 66) Базовые алгоритмы в бустинге. Решающие пни. Градиентный бустинг.
- 67) Стохастический градиентный бустинг.
- 68) Простое голосование (комитет большинства).
- 69) Идентификация нетипичных объектов (выбросов).
- 70) Бэггинг и метод случайных подпространств.
- 71) Анализ смещения и вариации для простого голосования.
- 72) Смесь алгоритмов (квазилинейная композиция), область компетентности, примеры функций компетентности.
- 73) Методы построения смесей: последовательный и иерархический.
- 74) Задача о многоруком бандите. Жадные и эpsilon-жадные стратегии. Метод UCB (upper confidence bound). Стратегия Softmax.
- 75) Адаптивные стратегии на основе скользящих средних. Метод сравнения с подкреплением. Метод преследования.
- 76) Постановка задачи при наличии информации о среде в случае выбора действия. Контекстный многорукий бандит.
- 77) Оценивание новой стратегии по большим историческим данным.

Примеры билетов на дифференцированном зачёте:

Билет 1.

- 1) Почему любая булева функция представима в виде нейронной сети? Сколько в ней слоёв?

2) В чём отличия внутренних и внешних критериев? Какие есть критерии скользящего контроля?

Билет 2.

- 1) В чём отличия случайного поиска от случайного поиска с адаптацией?
- 2) Жадный метод обучения решающего дерева.

Критерии оценивания

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

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

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

оценка «хорошо (7)» выставляется студенту, если он твердо знает материал билета, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе некоторые неточности;

оценка «хорошо (6)» выставляется студенту, если он знает материал билета, по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе много неточностей;

оценка «хорошо (5)» выставляется студенту, если он знает материал билета, излагает его, умеет применять полученные знания на практике, не допускает в ответе грубых ошибок;

оценка «удовлетворительно (4)» выставляется студенту если во время ответа билета 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 билета 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 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

Текущий контроль осуществляется в виде выполнения контрольной работы по отдельным разделам и выполнения самостоятельной работы. Для прохождения контроля студент должен 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 самостоятельной работы для представления полученных знаний.

Аттестация проводится в форме устного дифференцированного зачёта.

Оценка за промежуточный контроль учитывает оценку Ок/р за письменную контрольную работу и оценку Од/з самостоятельной работы студентов по текущим темам дисциплины; она рассчитывается следующим образом:

Опромежуточный= $0,5 \cdot \text{Ок/р} + 0,5 \cdot \text{Од/з}$ и округляется до целого числа арифметическим способом.

Итоговая оценка учитывает оценку за промежуточный контроль Опромежуточный и оценку за работу непосредственно на дифференцированном зачёте Одифзачёт и рассчитывается по формуле:

Оитоговая= $0,5 \cdot \text{Одифзачёт} + 0,5 \cdot \text{Опромежуточный}$.