



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Магнитогорский государственный технический университет им. Г.И. Носова»



УТВЕРЖДАЮ
Директор ИЭ и АС
В.Р. Храмшин

26.01.2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

ОСНОВЫ ТЕОРИИ МАШИННОГО ОБУЧЕНИЯ

Направление подготовки (специальность)
09.04.01 Информатика и вычислительная техника

Направленность (профиль/специализация) программы
Программное обеспечение для цифровизации предприятий и организаций

Уровень высшего образования - магистратура

Форма обучения
очная

Институт/ факультет	Институт энергетики и автоматизированных систем
Кафедра	Вычислительной техники и программирования
Курс	2
Семестр	3

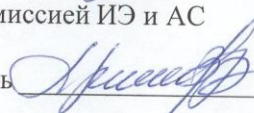
Магнитогорск
2022 год

Рабочая программа составлена на основе ФГОС ВО - магистратура по направлению подготовки 09.04.01 Информатика и вычислительная техника (приказ Минобрнауки России от 19.09.2017 г. № 918)

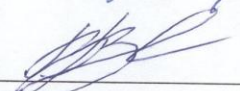
Рабочая программа рассмотрена и одобрена на заседании кафедры
Вычислительной техники и программирования
19.01.2022 г. протокол № 4

Зав. кафедрой  О.С. Логунова

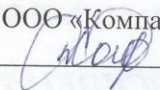
Рабочая программа одобрена методической комиссией ИЭ и АС
26.01.2022 г. протокол № 5

Председатель  В.Р. Храмшин

Рабочая программа составлена:
доцент кафедры ВТиП, д-р техн. наук

 В.С. Великанов

Рецензент:

Начальник отдела технологических платформ ООО «Компас Плюс», канд. техн. наук
 Д.С. Сафонов

Лист актуализации рабочей программы

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2022 - 2023 учебном году на заседании кафедры Вычислительной техники и программирования

Протокол от ____ 20__ г. № ____
Зав. кафедрой _____ О.С. Логунова

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2023 - 2024 учебном году на заседании кафедры Вычислительной техники и программирования

Протокол от ____ 20__ г. № ____
Зав. кафедрой _____ О.С. Логунова

1 Цели освоения дисциплины (модуля)

Целями освоения дисциплины «Основы теории машинного обучения» являются:

- формирование у студентов понимания современной методологии машинного обучения;
- формирование у студентов умения применять современные нечеткологические и нейросетевые методы;
- формирование у студентов навыков осознанного выбора и эффективного применения современных программных средств.

Для достижения поставленных целей в курсе «Основы машинного обучения» решаются задачи:

- изучение методологических основ машинного обучения;
- изучение алгоритмических основ машинного обучения;
- освоение современного программного обеспечения, реализующего методы машинного обучения.

2 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина Основы теории машинного обучения входит в часть учебного плана формируемую участниками образовательных отношений образовательной программы.

Для изучения дисциплины необходимы знания (умения, владения), сформированные в результате изучения дисциплин/ практик:

Интеллектуальные системы

Библиотеки языка программирования Python

Программное обеспечение для представления результатов научных исследований

Системный анализ в структурировании профессиональной информации

Современные проблемы цифровизации предприятий и организаций

Знания (умения, владения), полученные при изучении данной дисциплины будут необходимы для изучения дисциплин/практик:

Методы научного поиска

Выполнение и защита выпускной квалификационной работы

Информационно-управляющие системы предприятий и организаций

Методы оптимизации и их цифровая реализация

3 Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля) и планируемые результаты обучения

В результате освоения дисциплины (модуля) «Основы теории машинного обучения» обучающийся должен обладать следующими компетенциями:

Код индикатора	Индикатор достижения компетенции
ПК-4	Обладает способностью к разработке компонентов системы управления базами данных, отладке разрабатываемой системы управления базами данных, документированию разработанной системы управления базами данных в целом и ее компонентов и сопровождению созданной системы управления базами данных
ПК-4.1	Определяет необходимость разработки компонентов системы управления базами данных
ПК-4.2	Оценивает качество разработки компонентов системы управления базами данных

4. Структура, объём и содержание дисциплины (модуля)

Общая трудоемкость дисциплины составляет 2 зачетных единиц 72 академических часов, в том числе:

- контактная работа – 34,1 академических часов;
- аудиторная – 34 академических часов;
- внеаудиторная – 0,1 академических часов;
- самостоятельная работа – 37,9 академических часов;
- в форме практической подготовки – 0 академических часов;

Форма аттестации - зачет

Раздел/ тема дисциплины	Семестр	Аудиторная контактная работа (в акад. часах)			Самостоятельная работа студента	Вид самостоятельной работы	Форма текущего контроля успеваемости и промежуточной аттестации	Код компетенции
		Лек.	лаб. зан.	практ. зан.				
1. Введение в методы машинного обучения								
1.1 Основы работы с текстовой информацией	3		4		8	Самостоятельное изучение учебной и научной литературы	Беседа – обсуждение. Устный опрос.	ПК-4.1, ПК-4.2
1.2 Анализ информации на основе перцептральных моделей			4		2	Самостоятельное изучение учебной и научной литературы. Подготовка к лабораторному занятию. Выполнение лабораторной работы.	Беседа – обсуждение. Анализ программного кода. Устный опрос.	ПК-4.1, ПК-4.2
Итого по разделу			8		10			
2. Анализ текстовой информации								
2.1 Анализ «сырого» текста. Токенизация. Работа с корпусами текстов Программные средства для работы с текстами	3		4		6	Самостоятельное изучение учебной и научной литературы. Подготовка к лабораторному занятию. Выполнение лабораторной работы.	Беседа – обсуждение. Анализ программного кода. Устный опрос.	ПК-4.1, ПК-4.2

2.2 Аннотирование текстов. Задачи анализа текстовой информации. Специфика работы с информацией, размещенной в сети Интернет.			6		8	Самостоятельное изучение учебной и научной литературы. Подготовка к лабораторному занятию. Выполнение лабораторной работы.	Беседа – обсуждение. Анализ программного кода. Устный опрос.	ПК-4.1, ПК-4.2
Итого по разделу			10		14			
3. Нейросетевые методы извлечения информации. Нечеткологические методы извлечения информации								
3.1 Ассоциативные нейронные сети в задачах извлечения информации	3		6		4	Самостоятельное изучение учебной и научной литературы. Подготовка к лабораторному занятию. Выполнение лабораторной работы	Беседа – обсуждение. Анализ программного кода. Устный опрос	ПК-4.1, ПК-4.2
3.2 Нечеткологические методы в задачах извлечения информации			6		4	Самостоятельное изучение учебной и научной литературы. Подготовка к лабораторному занятию	Беседа – обсуждение. Анализ программного кода. Устный опрос	ПК-4.1, ПК-4.2
Итого по разделу			12		8			
4. Методы глубокого обучения								
4.1 Современные программные средства глубокого обучения	3		4		5,9	Самостоятельное изучение учебной и научной литературы. Подготовка к лабораторному занятию. Выполнение лабораторной работы	Беседа – обсуждение. Анализ программного кода. Устный опрос	ПК-4.1, ПК-4.2
Итого по разделу			4		5,9			
Итого за семестр			34		37,9		зачёт	
Итого по дисциплине			34		37,9		зачет	

5 Образовательные технологии

1. Традиционные образовательные технологии, ориентированные на организацию образовательного процесса и предполагающую прямую трансляцию знаний от преподавателя к студенту.

Формы учебных занятий с использованием традиционных технологий:

Информационная лекция – последовательное изложение материала в дисциплинарной логике, осуществляемое преимущественно вербальными средствами (монолог преподавателя).

Лабораторная работа – организация учебной работы с реальными материальными и информационными объектами, экспериментальная работа с аналоговыми моделями реальных объектов.

2. Технологии проблемного обучения – организация образовательного процесса, которая предполагает постановку проблемных вопросов, создание учебных проблемных ситуаций для стимулирования активной познавательной деятельности студентов.

3. Интерактивные технологии – организация образовательного процесса, которая предполагает активное и нелинейное взаимодействие всех участников, достижение на этой основе личностно значимого для них образовательного результата.

Формы учебных занятий с использованием специализированных интерактивных технологий:

Лекция «обратной связи» – лекция–провокация (изложение материала с заранее запланированными ошибками), лекция-беседа, лекция-дискуссия, лекция-конференция.

4. Информационно-коммуникационные образовательные технологии – организация образовательного процесса, основанная на применении программных сред и технических средств работы со знаниями в различных предметных областях.

6 Учебно-методическое обеспечение самостоятельной работы обучающихся

Представлено в приложении 1.

7 Оценочные средства для проведения промежуточной аттестации

Представлены в приложении 2.

8 Учебно-методическое и информационное обеспечение дисциплины (модуля)

а) Основная литература:

1. Рутковская, Д. Нейронные сети, генетические алгоритмы и нечеткие системы. [Электронный ресурс]. / Д. Рутковская, М. Пилиньский, Л. Рутковский. М.: Горячая линия – Телеком, 2013. – 384 с. – Режим доступа: <http://ibooks.ru/reading.php?productid=334029> . Заглавие с экрана ISBN 978-5-9912-0320-

2. Воронцов К.В. Машинное обучение. Курс лекций. [Электронный ресурс]. Режим доступа: <http://www.machinelearning.ru/wiki/index.php>

б) Дополнительная литература:

1. Галушкин, А.И. Нейронные сети: основы теории [Электронный ресурс]. / А.И. Галушкин М.: Горячая линия – Телеком, 2012. – 496 с. – Режим доступа: <http://ibooks.ru/reading.php?productid=333386> . Заглавие с экрана ISBN 978-5-9912-0082-0

2. Николенко С. Глубокое обучение. / С. Николенко, А. Кадури, Е. Архангельская. Пб.: Питер, 2018. – 480 с.

3. Асратян Р.Э. Интернет и распределенные многоагентные системы / Р.Э. Асратян, В.Н. Лебедев, Р.И. Дмитриев. М.: ЛЕНАНД, 2007. – 72 с

4. Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы / Н.Н. Леонтьева. М.: Издательский центр «Академия»; 2006. – 304 с.

5. Демидова, Л.А., Принятие решений в условиях неопределенности [Электронный ресурс]. / Л.А. Демидова, В.В. Кираковский, А.Н. Пылькин. М.: Горячая линия – Телеком, 2012. – 290 с. – Режим доступа: <http://ibooks.ru/reading.php?productid=333975> . Заглавие с экрана ISBN 978-5-9912-0112-7;

6. Башмаков, А.И. Интеллектуальные информационные технологии: Учеб. пособие [Текст] / А.И. Башмаков И.А.Башмаков. – М.: Изд во МГТУ им. Н.Э. Баумана, 2005. –304с.

7. Комарцова, Л.Г. Нейрокомпьютеры: Учеб. пособие для вузов.[Текст] / Л.Г. Комарцова, А.В. Максимов. – М.: Изд во МГТУ им. Н.Э. Баумана, 2002. –320с.

8. Хайкин, С. Нейронные сети: полный курс, 2 е изд., испр.: Пер. с англ. [Текст] / С. Хайкин. М.: ООО «И.Д. Вильямс», 2006. – 1104с.

9. Толкачев, С.Ф. Нейронное программирование диалоговых систем. [Текст] / С.Ф. Толкачев. СПб.: КОРОНА – Век, 2006. – 192 с.

в) Методические указания:

1. Жидьцов В.В. Практикум по нейросетевым технологиям. [Электронный ресурс]. Режим доступа: <http://bek.sibadi.org/fulltext/epd6.pdf>

г) Программное обеспечение и Интернет-ресурсы:

Программное обеспечение

Наименование ПО	№ договора	Срок действия лицензии
MS Windows 7 Professional(для классов)	Д-1227-18 от 08.10.2018	11.10.2021
JetBrains IDEA Community Edition	свободно распространяемое ПО	бессрочно
JetBrains PyCharm Community Edition	свободно распространяемое ПО	бессрочно
MS Windows 10 Professional (для классов)	Д-1227-18 от 08.10.2018	11.10.2021

Профессиональные базы данных и информационные справочные системы

Название курса	Ссылка
Информационная система - Единое окно доступа к информационным ресурсам	URL: http://window.edu.ru/
Поисковая система Академия Google (Google Scholar)	URL: https://scholar.google.ru/
Национальная информационно-аналитическая система – Российский индекс научного цитирования (РИНЦ)	URL: https://elibrary.ru/project_risc.asp
Электронная база периодических изданий East View Information Services, ООО «ИВИС»	https://dlib.eastview.com/

9 Материально-техническое обеспечение дисциплины (модуля)

Материально-техническое обеспечение дисциплины включает:

1. Лекционная аудитория Мультимедийные средства хранения, передачи и представления информации

2. Компьютерный класс. Персональные компьютеры с виртуальной машиной для установки серверного ПО, выходом в Интернет и с доступом в электронную информационно-образовательную среду университета.

3. Аудитории для самостоятельной работы: компьютерные классы; читальные залы библиотеки. Все классы УИТ и АСУ с персональными компьютерами, выходом в Интернет и с доступом в электронную информационно-образовательную среду университета.

4. Аудиторий для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Ауд. 282 и классы УИТ и АСУ.

5. Помещения для самостоятельной работы обучающихся, оснащенных компьютерной техникой с возможностью подключения к сети «Интернет» и наличием доступа в электронную информационно-образовательную среду организации. Классы УИТ и АСУ.

6. Помещения для хранения и профилактического обслуживания учебного оборудования Центр информационных технологий – ауд. 372.

Приложение 1. Учебно-методическое обеспечение самостоятельной работы обучающихся

Лабораторная работа №1

Обзор основных необходимых библиотек языка Python. Библиотека NumPy для оптимизированных вычислений над массивами данных. Введение в массивы библиотеки NumPy. Выполнение вычислений над массивами библиотеки NumPy, универсальные функции. Операции над данными в библиотеке Pandas. Обработка отсутствующих данных. Агрегирование и группировка. Визуализация с помощью библиотеки Matplotlib. Линейные графики, диаграммы рассеяния, гистограммы, трехмерные графики. Знакомство с библиотекой машинного обучения Scikit-Learn. Гиперпараметры и проверка качества модели

Лабораторная работа №2.

Задание 1. Построение и отбор признаков Извлечение признаков (Feature Extraction). Преобразования признаков (Feature transformations): кодирование нечисловых данных, нормировка и калибровка, заполнение пропусков. Выбор признаков (Feature selection): статистические подходы, визуализация, отбор с использованием моделей.

Задание 2. Решение задачи регрессии Разбор примера построения модели линейной регрессии для задачи предсказания велосипедного трафика Отбор и кодирование признаков. Визуальное сравнение общего и предсказанного моделью трафика. Проверка качества.

Задание 3. Решение задачи классификации. Разбор примера построения модели логистической регрессии для задачи предсказания оттока клиентов мобильного оператора. Отбор и кодирование признаков. Проверка качества модели с помощью перекрёстной проверки.

Лабораторная работа №3

Задание 1. Древоподобные модели: деревья решений, случайный лес Разбор примера построения модели дерева решений для задачи предсказания исхода футбольного матча. Анализ деревьев, полученных при использовании различных метрик. Построение модели случайного леса на примере задачи кредитного скоринга. Кодирование признаков и заполнение пропущенных данных.

Задание 2. Ансамбли моделей Бэггинг, бустинг, градиентный бустинг Разбор примера построения модели градиентного бустинга для задачи распознавания рукописных цифр из библиотеки MNIST.

Лабораторная работа № 4

Анализ текстовых данных Разбор примера построения модели анализа текстовых данных для задачи определения тональности киноотзывов.

Примеры тестовых заданий

Тест № 1

Задание 1: Что такое линейная регрессия?

Вариант 1 это функциональная зависимость, которая позволяет прогнозировать изменения непрерывных числовых параметров;

Вариант 2 модель зависимости непрерывной переменной y от объясняющих ее факторов, в которой функция зависимости является линейной

Вариант 3 модель зависимости дискретной переменной y от объясняющих ее факторов, в которой функция зависимости является линейной

Задание 2: Основная характеристика задачи бинарной классификации:

Вариант 1 классификация осуществляется по одному признаку

Вариант 2 зависимая переменная может принимать только два значения

Вариант 3 классификация осуществляется по двум признакам

Задание 3: Классификация относится к стратегии:

Вариант 1 обучения без учителя

Вариант 2 обучения с учителем

Вариант 3 оба ответа неверны

Задание 4: Явление переобучения характеризуется ...

Вариант 1 чрезмерно точным соответствием модели конкретному набору обучающих примеров, при котором модель теряет способность к обобщению

Вариант 2 возникновением, в случае слишком долгого обучения, недостаточного числа обучающих примеров или слишком сложной структуры модели

Вариант 3 возникновением, в случае слишком долгого обучения, слишком сложной структуры модели

Задание 5 Задача классификации сводится к ...

Вариант 1 нахождению частых зависимостей между объектами или событиями;

Вариант 2 определению класса объекта по его характеристикам;

Вариант 3 определению по известным характеристикам объекта значение некоторого его параметра;

Вариант 4 поиску независимых групп и их характеристик в всем множестве анализируемых данных.

Задание 6: Множество примеров, используемое для конструирования модели, называется...

Вариант 1 обучающим множеством

Вариант 2 тестовым множеством

Вариант 3 валидационным множеством

Перечень вопросов к зачету

1. Основные понятия машинного обучения. Основные постановки задач.

Примеры прикладных задач.

2. Линейные методы классификации и регрессии: функционалы качества, методы настройки, особенности применения.

3. Метрики качества алгоритмов регрессии и классификации.

4. Линейная регрессия. Простая многомерная регрессия. Регрессия с полиномиальными признаками. Методы регуляризации: Ridge, Lasso, ElasticNet.

5. Логистическая регрессия.

6. Деревья решений. Методы построения деревьев. Их регуляризация.

7. Композиции алгоритмов. Разложение ошибки на смещение и разброс.

8. Случайный лес, его особенности.

9. Градиентный бустинг, его особенности при использовании деревьев в качестве базовых алгоритмов.

10. Анализ текстов.

Приложение 2. Оценочные средства для проведения промежуточной аттестации

а) Планируемые результаты обучения и оценочные средства для проведения промежуточной аттестации:

Код индикатора	Индикатор достижения компетенции	Оценочные средства
ПК-4: Обладает способностью к разработке компонентов системы управления базами данных, отладке разрабатываемой системы управления базами данных, документированию разработанной системы управления базами данных в целом и ее компонентов и сопровождению созданной системы управления базами данных		
ПК-4.1	Определяет необходимость разработки компонентов системы управления базами данных	<i>Перечень теоретических вопросов</i> 1. Основные понятия машинного обучения. Основные постановки задач. Примеры прикладных задач. 2. Линейные методы классификации и регрессии: функционалы качества, методы настройки, особенности применения. 3. Метрики качества алгоритмов регрессии и классификации. 4. Линейная регрессия. Простая многомерная регрессия. Регрессия с полиномиальными признаками. Методы регуляризации: Ridge, Lasso, ElasticNet. 5. Логистическая регрессия. 6. Деревья решений. Методы построения деревьев. Их регуляризация. 7. Композиции алгоритмов. Разложение ошибки на смещение и разброс. 8. Случайный лес, его особенности. 9. Градиентный бустинг, его особенности при использовании деревьев в качестве базовых алгоритмов. 10. Анализ текстов. <i>Практические задания</i> <i>Лабораторная работа №1</i> Обзор основных необходимых библиотек языка Python. Библиотека NumPy для оптимизированных вычислений над массивами данных. Введение в массивы библиотеки NumPy. Выполнение вычислений над массивами библиотеки NumPy, универсальные функции. Операции над данными в библиотеке Pandas. Обработка отсутствующих данных. Агрегирование и группировка. Визуализация с помощью библиотеки Matplotlib. Линейные графики, диаграммы рассеяния, гистограммы, трехмерные графики. Знакомство с библиотекой машинного обучения Scikit-Learn. Гиперпараметры и проверка качества модели <i>Лабораторная работа №2.</i> Задание 1. Построение и отбор признаков Извлечение признаков (Feature Extraction).
ПК-4.2	Оценивает качество разработки компонентов системы управления базами данных	

Код индикатора	Индикатор достижения компетенции	Оценочные средства
		Преобразования признаков (Feature transformations): кодирование нечисловых данных, нормировка и калибровка, заполнение пропусков. Выбор признаков (Feature selection): статистические подходы, визуализация, отбор с использованием моделей. Задание 2. Решение задачи регрессии Разбор примера построения модели линейной регрессии для задачи предсказания велосипедного трафика Отбор и кодирование признаков. Визуальное сравнение общего и предсказанного моделью трафика. Проверка качества. Задание 3. Решение задачи классификации. Разбор примера построения модели логистической регрессии для задачи предсказания оттока клиентов мобильного оператора. Отбор и кодирование признаков. Проверка качества модели с помощью перекрёстной проверки. <i>Лабораторная работа №3</i> Задание 1. Древовидные модели: деревья решений, случайный лес Разбор примера построения модели дерева решений для задачи предсказания исхода футбольного матча. Анализ деревьев, полученных при использовании различных метрик. Построение модели случайного леса на примере задачи кредитного скоринга. Кодирование признаков и заполнение пропущенных данных. Задание 2. Ансамбли моделей Бэггинг, бустинг, градиентный бустинг Разбор примера построения модели градиентного бустинга для задачи распознавания рукописных цифр из библиотеки MNIST. <i>Лабораторная работа № 4</i> Анализ текстовых данных Разбор примера построения модели анализа текстовых данных для задачи определения тональности киноотзывов <i>Задания на решение задач из профессиональной области, комплексные задания</i> В рамках данного задания необходимо решить задачу бинарной классификации для предсказания уровня заработной платы по тексту объявления о вакансии на примере набора данных 1) Разбейте получившуюся выборку на обучающую и контрольную в соотношении 70/30 2) Создайте текстовое описание объектов обучающей и контрольной выборок, объединив значения всех признаков каждого объекта выборки через символы пробела. После этого получите признаковое описание объектов, осуществив векторизацию получившихся текстов при помощи CountVectorizer,

Код индикатора	Индикатор достижения компетенции	Оценочные средства
		обучив его на обучающей выборке и применив на тестовой. 3) Обучите логистическую регрессию из модуля sklearn с параметрами по умолчанию на обучающей выборке: 4) Вычислите значения ROC-AUC, F-меры, а также постройте матрицу ошибок на тестовой выборке. 5) Отсортируйте веса признаков для модели. Какие слова из встречающихся в выборке имеют наибольшее/наименьшее влияние на значение целевой переменной? Проинтерпретируйте полученный результат. 6) Создайте текстовое описание объектов обучающей и контрольной выборок, объединив значения всех признаков каждого объекта выборки через символы пробела. После этого получите признаковое описание объектов, вычислив вектор tf-idf для каждого объекта помощи TfidfVectorizer, обучив его на обучающей выборке и применив на тестовой. 7) Заново обучите модель 8) Вычислите значения ROC-AUC, F-меры, а также постройте матрицу ошибок на контрольной выборке. 9) Сравните значения метрик из п. 8 со значениями, полученными в п. 4, и сравните соответствующие модели по качеству из работы. 10. Отсортируйте веса признаков для модели логистической регрессии из scikit-learn, полученной в п. 7. Какие слова из встречающихся в выборке имеют наибольшее/наименьшее влияние на значение целевой переменной? Проинтерпретируйте полученный результат.

б) Порядок проведения промежуточной аттестации, показатели и критерии оценивания:

Промежуточная аттестация по дисциплине «Основы теории машинного обучения» включает теоретические вопросы, позволяющие оценить уровень усвоения обучающимися знаний, и практические задания, выявляющие степень сформированности умений и владений, проводится в форме зачета.

Зачет по дисциплине проводится по результатам отчетности за выполненные самостоятельные работы с опросом в устной форме по этапам выполнения в беседе-обсуждении на лекционных занятиях.

Показатели и критерии оценивания зачета:

- на оценку «зачтено» – обучающийся демонстрирует пороговый уровень сформированности компетенций;
- на оценку «не зачтено» – обучающийся демонстрирует знания не более 20% теоретического материала, допускает существенные ошибки, не может показать интеллектуальные навыки решения простых задач.