



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение высшего образования  
**«Дальневосточный федеральный университет»**  
(ДФУ)  
ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ (ШКОЛА)

«СОГЛАСОВАНО»

Руководитель ОП

Подпись

Дремлюга Р.И.

УТВЕРЖДАЮ

И.о. директора Академии цифровой  
трансформации

Еременко А.С.

«26» января 2022 г.



### РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Машинное обучение (Machine Learning)

**Направление подготовки - 09.04.01 Информатика и вычислительная техника**  
(Искусственный интеллект и большие данные (совместно с ПАО Сбербанк))

#### Форма подготовки очная

курс 1 семестр 1,2  
лекции 54 час.  
практические занятия 72 час.  
лабораторные работы 00 час.  
в том числе с использованием МАО практические занятия 63 час  
всего часов аудиторной нагрузки 126 час.  
самостоятельная работа 162 час.  
в том числе на подготовку к экзамену 63 час.  
контрольные работы (количество) не предусмотрены  
курсовая работа / курсовой проект не предусмотрены  
зачет не предусмотрен  
экзамен 1 и 2 семестр

Рабочая программа дисциплины разработана в соответствии с федеральным государственным образовательным стандартом (ФГОС) высшего образования (ВО) – магистратура по направлению подготовки 09.04.01 Информатика и вычислительная техника, утвержденного приказом Министерства образования и науки Российской Федерации от 19.09.2017 г. № 918 (с изменениями и дополнениями).

Рабочая программа обсуждена на заседании Академии цифровой трансформации протокол № от 25 января 2022 г.

И.о. директора Академии цифровой трансформации

Еременко А.С.

Составители: к.т.н. Еременко А.С., Кленин А.С.

Владивосток  
2022

**Оборотная сторона титульного листа РПУД**

**I. Рабочая программа пересмотрена на заседании департамента:**

Протокол от « \_\_\_\_\_ » \_\_\_\_\_ 20\_\_ г. № \_\_\_\_\_

Директор департамента \_\_\_\_\_  
(подпись) (И.О. Фамилия)

**II. Рабочая программа пересмотрена на заседании департамента:**

Протокол от « \_\_\_\_\_ » \_\_\_\_\_ 20\_\_ г. № \_\_\_\_\_

Директор департамента \_\_\_\_\_  
(подпись) (И.О. Фамилия)

**III. Рабочая программа пересмотрена на заседании департамента:**

Протокол от « \_\_\_\_\_ » \_\_\_\_\_ 20\_\_ г. № \_\_\_\_\_

Директор департамента \_\_\_\_\_  
(подпись) (И.О. Фамилия)

**IV. Рабочая программа пересмотрена на на заседании департамента:**

Протокол от « \_\_\_\_\_ » \_\_\_\_\_ 20\_\_ г. № \_\_\_\_\_

Директор департамента \_\_\_\_\_  
(подпись) (И.О. Фамилия)

## I. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

**Цель** – изучение основных разделов теории машинного обучения (Machine Learning) и овладение навыками практического решения задач интеллектуального анализа данных - майнинга данных (Data Mining).

### **Задачи:**

- Изучить основные инструменты математического анализа, линейной алгебры, методов оптимизации и теории вероятностей;
- Получить базовые навыки программирования на языках C++ и Python применительно к работе с большими объемами данных;
- Изучить основные модели машинного обучения и методики оценки их качества;
- Изучить основные способы организации искусственных нейронных сетей;
- Овладеть методологией управления data-science проектами;
- Научиться строить модели машинного обучения для решения профессиональных задач.

В результате освоения дисциплины обучающийся должен:

### **Знать:**

- современное состояние исследований в области машинного обучения;
- принципы построения систем машинного обучения;
- модели представления и описания технологий машинного обучения.

### **Уметь:**

- проводить анализ предметной области;
- определять назначение, выбирать методы и средства для построения систем машинного обучения;
- строить системы машинного обучения.

### **Иметь навыки и (или) опыт деятельности (владеть):**

- использования аппарата простейшего анализ данных;
- применения методов классификации информации;
- реализации алгоритмов машинного обучения.

### *Связь курса с другими дисциплинами*

Для успешного изучения дисциплины «Машинное обучение» необходимы знания базовой программы курса «Высшая математика» и основ программирования (желательно Python).

В результате данной дисциплины у обучающихся формируются следующие общепрофессиональные компетенции (элементы компетенций):

## 1. Универсальные компетенции и индикаторы их достижения:

Наименование категории (группы) компетенций	Код и наименование компетенции выпускника	Код и наименование индикатора достижения компетенции
Системное и критическое мышление	<b>УК-1</b> Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	<b>УК-1.1</b> Анализирует проблемную ситуацию с применением системного подхода и критического анализа, используя достоверные данные и надежные источники информации
		<b>УК-1.2</b> Формирует обоснованную и логически последовательную позицию, аргументирует свою точку зрения, предлагает возможные варианты решения поставленной задачи с учетом возможной критики и ограничений
		<b>УК-1.3</b> Разрабатывает сценарий реализации оптимальной стратегии решения проблемной ситуации с учетом необходимых ресурсов, достижимых результатов, возможных рисков и последствий
Коммуникация	<b>УК-4</b> Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия	<b>УК-4.1</b> Создает различные типы письменных и устных текстов на русском и иностранном языке для академического и профессионального взаимодействия

Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения)
<b>УК-1.1</b> Анализирует проблемную ситуацию с применением системного подхода и критического анализа, используя достоверные данные и надежные источники информации	Знает методы поиска информации, требуемой для выполнения исследований
	Умеет производить отбор и систематизацию информации, требуемой для выполнения исследований и решения проблемы
	Владеет методологией системного и критического анализа проблемных ситуаций
<b>УК-1.2</b> Формирует обоснованную и логически последовательную позицию, аргументирует свою точку зрения, предлагает возможные варианты решения поставленной задачи с учетом возможной критики и ограничений	Знает методы поиска информации, требуемой для решения поставленной задачи
	Умеет аргументировать свою точку зрения
	Владеет навыками определения альтернативных вариантов решений поставленной задачи
<b>УК-1.3</b> Разрабатывает сценарий реализации оптимальной стратегии решения проблемной ситуации с учетом необходимых ресурсов, достижимых результатов, возможных рисков и последствий	Знает правила формулирования стратегических целей
	Умеет разрабатывать сценарий реализации оптимальной стратегии решения проблемной ситуации
	Владеет методологией системного и критического анализа проблемных ситуаций; методиками постановки цели, определения способов ее достижения, разработки стратегий

Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения)
	действий
УК-4.3 Представляет результаты исследовательской и проектной деятельности на различных публичных мероприятиях, участвует в академических и профессиональных дискуссиях, в том числе на иностранном языке	Знает основные специальные термины и грамматические конструкции для работы с оригинальными текстами академического и профессионального характера
	Умеет строить лексически правильно, грамотно, логично и последовательно устные и письменные высказывания в ситуациях академического и профессионального взаимодействия
	Владеет навыками построения лексически правильного, грамотного, логичного и последовательного устного и письменного высказывания в ситуациях академического и профессионального взаимодействия на английском языке

## 1. Общепрофессиональные компетенции и индикаторы их достижения:

Код и наименование компетенции выпускника	Код и наименование индикатора достижения компетенции
<b>ОПК-1</b> Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте	ОПК-1.1. Применяет при решении профессиональных задач математические, естественнонаучные, социально-экономические и профессиональные знания
	ОПК-1.2. Применяет методы теоретического и экспериментального исследования объектов профессиональной деятельности
	ОПК-1.3. Выбирает современные информационно-коммуникационные технологии при постановке и решении задач профессиональной деятельности, в том числе в новой или незнакомой среде и в междисциплинарном контексте
<b>ОПК-2</b> Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач	ОПК-2.1. Применяет знания современных интеллектуальных технологий, инструментальных сред, программно-технических платформ для решения профессиональных задач
	ОПК-2.2. Разрабатывает оригинальные алгоритмы и программные средства для решения профессиональных задач
<b>ОПК-6</b> Способен разрабатывать компоненты программно-аппаратных комплексов обработки информации и автоматизированного проектирования	ОПК-6.1. Применяет знания аппаратных средств и платформ инфраструктуры информационных технологий, методов разработки и администрирования программно-аппаратных комплексов для решения профессиональных задач

Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения)
ОПК-1.1. Применяет при решении профессиональных задач математические, естественнонаучные, социально-экономические и профессиональные знания	Знает математические, естественнонаучные и социально-экономические методы для использования в профессиональной деятельности
	Умеет адаптировать существующие математические, естественнонаучные и социально-экономические методы для решения основных,

Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения)
	нестандартных задач создания и применения искусственного интеллекта
	Владеет навыками построения и реализации основных математических алгоритмов, навыками анализа математических проблем
ОПК-1.2. Применяет методы теоретического и экспериментального исследования объектов профессиональной деятельности	Знает методологию исследования объектов профессиональной деятельности
	Умеет выполнять анализ существенных свойств объектов профессиональной деятельности
	Владеет методами формального описания результатов анализа свойств объектов профессиональной деятельности
ОПК-1.3. Выбирает современные информационно-коммуникационные технологии при постановке и решении задач профессиональной деятельности, в том числе в новой или незнакомой среде и в междисциплинарном контексте	Знает направления современного развития информационно-коммуникационных технологий
	Умеет решать нестандартные профессиональные задачи, в том числе в новой или незнакомой среде и в междисциплинарном контексте, с применением математических, естественнонаучных, социально-экономических и профессиональных знаний
	Владеет методами создания программных средств для решения нестандартных задач
ОПК-2.1. Применяет знания современных интеллектуальных технологий, инструментальных сред, программно-технических платформ для решения профессиональных задач	Знает пакеты прикладных программ, относящиеся к профессиональной сфере
	Умеет разрабатывать оригинальные программные средства, в том числе с использованием современных информационно-коммуникационных и интеллектуальных компьютерных технологий
	Владеет методами проектирования и программирования интеллектуальных технологий
ОПК-2.2. Разрабатывает оригинальные алгоритмы и программные средства для решения профессиональных задач	Знает пакеты прикладных программ, относящиеся к профессиональной сфере
	Умеет разрабатывать оригинальные программные средства, в том числе с использованием современных информационно-коммуникационных и интеллектуальных компьютерных технологий
	Владеет навыками создания математических моделей, алгоритмов, методов, инструментальных средств по тематике проводимых проектов
ОПК-6.2. Анализирует техническое задание, разрабатывает и оптимизирует программный код для решения задач обработки информации и автоматизированного проектирования	Знает методы разработки компонентов программно-аппаратных комплексов обработки информации и автоматизированного проектирования
	Умеет разрабатывать компоненты программно-аппаратных комплексов обработки информации и автоматизированного проектирования
	Владеет навыками администрирования программно-аппаратных комплексов для решения профессиональных задач

## II. ТРУДОЁМКОСТЬ ДИСЦИПЛИНЫ И ВИДОВ УЧЕБНЫХ ЗАНЯТИЙ ПО ДИСЦИПЛИНЕ

Общая трудоемкость дисциплины составляет 3 зачётных единицы (108 академических часов).

Видами учебных занятий и работы обучающегося по дисциплине могут являться:

Обозначение	Виды учебных занятий и работы обучающегося
Лек	Лекции
Пр	Практические занятия
СР	Самостоятельная работа обучающегося в период теоретического обучения
Контроль	Самостоятельная работа обучающегося и контактная работа обучающегося с преподавателем в период промежуточной аттестации

### Структура дисциплины:

Форма обучения – очная.

№	Наименование раздела (темы) дисциплины	Семестр	Количество часов по видам учебных занятий и работы обучающегося				Контроль	Формы промежуточной аттестации, текущего контроля
			Лекции	Практ. занятие	СРС	Контроль		
1	Раздел 1	1	4	4	4	4	Контрольная работа, разноуровневые задачи и задания, собеседование	
2	Раздел 2	1	8	8	8	8	Контрольная работа, разноуровневые задачи и задания, собеседование	
3	Раздел 3	1	8	8	8	8	Контрольная работа, разноуровневые задачи и задания, собеседование	
4	Раздел 4	1	8	8	8	8	Реферат, разноуровневые задачи и задания, собеседование	
5	Раздел 5	1	8	8	8	8	Контрольная работа, разноуровневые задачи и задания, собеседование	
6	Раздел 6	2	4	9	15	6	Тест, разноуровневые задачи и задания, собеседование	
7	Раздел 7	2	5	9	16	7	Контрольная работа, разноуровневые задачи и задания, собеседование	

8	Раздел 8	2	4	9	16	7	Письменная работа, разноуровневые задачи и задания, собеседование
9	Раздел 9	2	5	9	16	7	Контрольная работа, разноуровневые задачи и задания, собеседование
	Итого:		54	72	99	63	Экзамен

### III. СТРУКТУРА И СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА

(54 часа)

#### **Занятие 1. Основы машинного обучения (4 час.)**

Введение. Основы машинного обучения. Задача обучения с учителем и без учителя. Классификация и регрессия. Линейные модели. Обработка данных. Кросс-валидация. Подбор гиперпараметров. Визуализация данных.

#### **Занятие 2. Методы оптимизации. Градиентный спуск (8 час.)**

Задача регрессии, классификации. Функция потерь. Оптимизация. Перебор по сетке. Производная, частные производные, градиент. Градиентный спуск, проблема выбора шага. Стохастический градиентный спуск. Использование момента. Adagrad, Adadelata, Adam. RMSProp\*.

#### **Занятие 3 Линейная регрессия. Глобальная оптимизация (8 час.)**

Постановка задачи линейной регрессии. Метод наименьших квадратов. Ковариация, корреляция. Критерий R<sup>2</sup>. Анализ остатков.

#### **Занятие 4. Глобальная оптимизация (8 час.)**

Генетический алгоритм. Многопараметрическая оптимизация. Доминанция и оптимальность по Парето. Функция качества (fitness). Аппроксимация качества. Общая идея генетического алгоритма. Представление генома. Методы селекции: пропорционально качеству, универсальная выборка (stochastic universal sampling), с наследием (reward-based), турнир. Стратегия элитизма. Методы кроссовера. Двух и многоточечный, равномерный (по подмножествам), для перестановок. Мутация. Влияние на скорость обучения. Управление популяцией. Сегрегация, старение, распараллеливание. Генетическое программирование.

#### **Занятие 5. Метод ближайших соседей (k-NN) (8 час.)**

Понятие и свойства метрики. Ослабление требования к неравенству треугольника. Базовый алгоритм классификации методом 1-NN и k-NN. Преимущества и недостатки. Метрики L1, L2, Хемминга, Левенштейна, косинусное расстояние. Потеря точности нормы в высоких размерностях.

Нормализация координат. Предварительная трансформация пространства признаков. Метрика Махаланобиса. Кросс-валидация методом "без одного" (leave one out). Определение границ, показатель пограничности. Сжатие по данным. Понятия выброса, прототипа, усвоенной точки. Алгоритм Харта (Hart). Регрессия методом k-NN. Взвешенные соседи. Связь с градиентным спуском. Стохастическая формулировка, softmax. Метод соседних компонент (neighbour component analysis)\*. Связь с выпуклой оптимизацией. Метод большого запаса (Large margin NN). Оптимизация классификатора, k-d деревья Хеши чувствительные к локальности, хеши сохраняющие локальность\*.

#### **Занятие 6. Наивный байесов классификатор (4 час.)**

Условная вероятность. Байесово решающее правило. Обновление вероятностей. Наивный классификатор, предположение о независимости признаков. Оценка плотности распределения для числовых признаков. Алгоритмические оптимизации. Алгоритм EM.

#### **Занятие 7. Логистическая регрессия (5 час.)**

Сигмоид. Метод наибольшего правдоподобия. Логистическая регрессия для меток  $-1, 1$ .

#### **Занятие 8. Деревья решений (4 час.)**

Понятие дерева решений. Борьба с оверфиттингом: bagging, выборки признаков. Ансамбли, случайный лес (Random Forest). Понятие энтропии, определение информации по Шеннону. Метрики: примеси Джини (Gini impurity), добавленная информация (information gain). Деревья регрессии. Метрика вариации. Непрерывные признаки. Использование главных компонент вместо признаков. Сокращение дерева (pruning).

#### **Занятие 9. Кластеризация (5 час.)**

Задача обучения без учителя, применения при эксплораторном анализе. Неметрическая кластеризация: функция схожести, компоненты связности и остовные деревья, иерархическая кластеризация снизу вверх Метрики, понятие центроида и представителя класса Центроидные алгоритмы: k-means, k-medoid. Алгоритмы, основанные на плотности: DBSCAN, OPTICS. Алгоритмы, основанные на распределении: сумма гауссиан. Нечёткая кластеризация, алгоритм c-means. Метрики качества: leave-one-out, силуэт, индекс Дэвиса-Болдина (Davies-Bouldin), индекс Данна (Dunn).

## **IV. СТРУКТУРА И СОДЕРЖАНИЕ ПРАКТИЧЕСКОЙ ЧАСТИ КУРСА (72 часа)**

### **Занятие 1. Работа с текстом (12 час.)**

Задачи обработки текста: извлечение, поиск, классификация (тематическая, эмоциональная), перевод. Разбиение на слова, пунктуация, лексический и морфологический анализ. Определение частей речи, имён, основ слов. Частотный анализ, представление bag-of-words, TF-IDF и его варианты. N-граммы, byte-pair encoding. Векторные представления, семантическая интерпретация алгебраических операций. Унитарный код (One-hot encoding). Алгоритмы Word2Vec и FastText. Алгоритм GloVe\*.

### **Занятие 2. Снижение размерности (12 час.)**

Постановка задачи, причины и цели снижения размерности. Выбор и извлечение признаков. Подходы к выбору признаков: filtering, wrapping, embedding. Расстояние между распределениями. Расстояние Кульбака-Лейблера. Взаимная информация. Алгоритмы выбора признаков: на основе корреляции (CFS), взаимной информации, Relief. Метод главных компонент (PCA). Нелинейные обобщения метода главных компонент. Kernel PCA. Неотрицательное матричное разложение (NMF). Стохастическое вложение соседей с t-распределением (t-SNE).

### **Занятие 3. Метод опорных векторов (12 час.)**

Постановка задачи линейного SVM для линейно разделимой выборки. Задача оптимизации с ограничениями. Двойственная задача Лагранжа. Условия Каруша-Куна-Такера. Функция Лагранжа для линейного SVM. Опорный вектор. Типы опорных векторов. Kernel trick. Полиномиальное ядро. Радиально-базисное ядро (RBF). SVM для задачи регрессии.

### **Занятие 4. Работа с изображениями (12 час.)**

Сверхточные фильтры, непрерывное и дискретное определение свёртки. Сглаживающие фильтры. Фильтр Гаусса. Дифференцирующие фильтры: Roberts cross, Sobel, Prewitt, Scharr.

### **Занятие 5. Поиск границ (12 час.)**

Алгоритм Кенни (Canny). Адаптивное сглаживание. Определение порога методом Отцу (Otsu).

### **Занятие 6. Оптимизация (12 час.)**

Оптимизация с учётом направления градиента. Преобразование Hough. Обобщения на многопараметрический и многомерный случай. Извлечение признаков. Признаки Хаара (Haar).

## **V. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ**

## План-график выполнения самостоятельной работы по дисциплине

№ п/п	Дата / сроки выполнения	Вид самостоятельной работы	Примерные норма времени на выполнение	Форма контроля
1.	3 неделя 1 семестра	Индивидуальное задание. Работа с текстом	3 недели	Проверка программы
2.	9 неделя 1 семестра	Индивидуальное задание. Снижение размерности	3 недели	Проверка программы
3.	12 неделя 1 семестра	Индивидуальное задание. Метод опорных векторов	3 недели	Проверка программы
4.	2 неделя 2 семестра	Индивидуальное задание. Работа с изображениями	4 недели	Проверка программы
5.	10 неделя 2 семестра	Индивидуальное задание. Поиск границ	4 недели	Проверка программы
6.	15 неделя 2 семестра	Индивидуальное задание. Оптимизация	3 недели	Проверка программы
7.	16-19 неделя	Подготовка к экзамену	-	Экзамен

Самостоятельная работа студентов состоит из подготовки индивидуального задания, а также подготовки к тестированию и практическим занятиям, подготовки презентаций и докладов по заданным темам.

Преподаватель даёт каждому студенту индивидуальные и дифференцированные задания. Некоторые из них могут осуществляться в группе (например, подготовка доклада и презентации по одной теме могут делать несколько студентов с разделением своих обязанностей – один готовит научно-теоретическую часть, а второй проводит анализ практики).

Типовые индивидуальные задания:

1. Задача А. Градиентный спуск

Входной файл: Стандартный вход – Ограничение времени: 1 сек

Выходной файл: Стандартный выход – Ограничение памяти: 512 Мб

Условие: требуется реализовать класс на языке Python, который соответствует следующему интерфейсу:

```
class GradientOptimizer:
```

```
def __init__(self, oracle, x0):
```

```
self.oracle = oracle
```

```
self.x0 = x0
```

```
def optimize(self, iterations, eps, alpha):
```

```
pass
```

В конструктор принимаются два аргумента – оракул, с помощью которого можно получить градиент оптимизируемой функции, а также точку, с которой необходимо начать градиентный спуск.

Метод `optimize` принимает максимальное число итераций для критерия остановки, L2-норму градиента, которую можно считать оптимальной, а также `learning rate`. Метод возвращает оптимальную точку.

Оракул имеет следующий интерфейс:

```
class Oracle:
```

```
def get_func(self, x)
```

```
def get_grad(self, x)
```

`x` имеет тип `np.array` вещественных чисел.

Формат выходных данных: код должен содержать только класс и его реализацию. Он не должен ничего выводить на экран.

2. Задача В. Линейная регрессия. Основы

Входной файл: Стандартный вход – Ограничение времени: 1 сек

Выходной файл: Стандартный выход – Ограничение памяти: 512 Мб

Условие: требуется реализовать следующие функции на языке Python.

```
def linear_func(theta, x) # function value
def linear_func_all(theta, X) # 1-d np.array of function values of all
                             # rows of the matrix X
```

```
def mean_squared_error(theta, X, y) # MSE value of current regression
```

```
def grad_mean_squared_error(theta, X, y) # 1-d array of gradient by theta
```

`theta` – одномерный `np.array`

`x` – одномерный `np.array`

`X` – двумерный `np.array`. Каждая строка соответствует по размерности вектору `theta`

`y` – реальные значения предсказываемой величины

Матрица `XX` имеет размер  $M \times N M \times N$ .  $MM$  строк и  $NN$  столбцов.

Используется линейная функция вида:

$$h\theta(x) = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

Mean squared error (MSE) как функция от  $\theta$ :

$$J(\theta) = \frac{1}{M} \sum_{i=1}^M (y_i - h\theta(x(i)))^2$$

Где  $x(i)$  —  $i$ -я строка матрицы `XX`

Градиент функции MSE:

$$\nabla J(\theta) = \{\partial J / \partial \theta_1, \partial J / \partial \theta_2, \dots, \partial J / \partial \theta_N\}$$

Пример:

```
X = np.array([[1,2],[3,4],[4,5]])
```

```
theta = np.array([5, 6])
```

```
y = np.array([1, 2, 1])
```

```
linear_func_all(theta, X) # —> array([17, 39, 50])
```

```
mean_squared_error(theta, X, y) # —> 1342.0
```

```
grad_mean_squared_error(theta, X, y) # —> array([215.33333333, 283.33333333])
```

Формат выходных данных: код должен содержать только реализацию функций.

### 3. Задача С. Найти линейную регрессию

Входной файл: Стандартный вход – Ограничение времени: 10 сек

Выходной файл: Стандартный выход – Ограничение памяти: 512 Мб

Условие: требуется реализовать функцию на языке Python, которая находит линейную регрессию заданных векторов, используя метрику MSE.

```
def fit_linear_regression(X, y) # np.array of linear regression coefs
```

X – двумерный np.array. Каждая строка соответствует отдельному примеру.

y – реальные значения предсказываемой величины

Формат выходных данных: код должен содержать только реализацию функций.

### 4. Задача D. Логистическая регрессия. Основы

Входной файл: Стандартный вход – Ограничение времени: 1 сек

Выходной файл: Стандартный выход – Ограничение памяти: 512 Мб

Условие: требуется реализовать следующие функции на языке Python.

```
def logistic_func(theta, x) # function value
```

```
def logistic_func_all(theta, X) # 1-d np.array of function values of all  
rows of the matrix X
```

```
def cross_entropy_loss(theta, X, y) # cross entropy loss value of current  
regression
```

```
def grad_cross_entropy_loss(theta, X, y) # 1-d array of gradient by theta
```

theta – одномерный np.array

x – одномерный np.array

X – двумерный np.array. Каждая строка соответствует по размерности вектору theta

y – реальные значения предсказываемой величины

Матрица XX имеет размер  $M \times NM \times N$ . MM строк и NN столбцов.

Используется линейная функция вида:

$$h\theta(x) = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

Формат выходных данных: код должен содержать только реализацию функций.

### 5. Задача В. Найти логистическую регрессию

Входной файл: Стандартный вход – Ограничение времени: 10 сек

Выходной файл: Стандартный выход – Ограничение памяти: 512 Мб

Условие: требуется реализовать функцию на языке Python, которая находит логистическую регрессию заданных векторов, используя метрику cross entropy loss.

`def fit_logistic_regression(X, y)` # np.array of logistic regression coefs  
X – двумерный np.array. Каждая строка соответствует отдельному примеру.

y – реальные значения предсказываемой величины

Формат выходных данных: код должен содержать только реализацию функций.

Требования к представлению и оформлению результатов самостоятельной работы

– Подготовка к тестированию

Студентам предлагается самостоятельно подготовиться к тестированию. При подготовке необходимо использовать конспект лекций, презентации лекций, которые передаются преподавателем студентам, рекомендуемую литературу.

Самостоятельная работа по подготовке к тестированию считается выполненной и зачтённой в случае более 70% правильных ответов на вопросы тестов (10-6 баллов).

– Подготовка к практическим занятиям

Студентам предлагается самостоятельно подготовиться к выполнению практических занятий. Для этого студент должен проработать теоретическую основу практической работы и методику её выполнения. Самостоятельная работа для подготовки к практическому занятию считается выполненной и зачтённой в случае аргументированного обоснования результата практической работы при её защите. Каждое практическое или контрольное задание включает краткие методические указания, задания для решения, контрольные вопросы для подготовки и защиты. По основным разделам приведены тесты. При выполнении сложных самостоятельных заданий приведены примеры выполнения. Студенты могут скачать пособие на сайте ДВФУ.

– Подготовка к докладам и презентациям

Презентации к докладам должны быть выполнены в программе PowerPoint. Первый слайд обязательно содержит выходные сведения: ФИО автора, ФИО руководителя, название профильной кафедры, тему доклада, год, мест создания, все слайды (кроме титульного) должны быть пронумерованы.

Последовательность подготовки презентации:

1. Чётко сформулировать цель презентации: вы хотите свою аудиторию мотивировать, убедить, заразить какой-то идеей или просто формально отчитаться.

2. Определить каков будет формат презентации: живое выступление (тогда, сколько будет его продолжительность) или электронная рассылка (каков будет контекст презентации).

3. Отобрать всю содержательную часть для презентации и выстроить логическую цепочку представления.

4. Определить ключевые моменты в содержании текста и выделить их.

5. Определить виды визуализации (картинки) для отображения их на слайдах в соответствии с логикой, целью и спецификой материала.

6. Подобрать дизайн и форматировать слайды (количество картинок и текста, их расположение, цвет и размер).

7. Проверить визуальное восприятие презентации.

– Подготовка к экзамену

Студент должен самостоятельно проработать информацию, используя все лекции, глоссарий, рекомендованную учебно-методическую литературу и информацию из иных источников для ответов по контрольным вопросам к экзамену.

Рекомендуется регулярное посещение всех учебных занятий в течение всего семестра: лекций, консультаций и прочее, а также активное изучение рекомендованной литературы, и выполнение в установленные сроки всех индивидуальных заданий.

При ответе на каждый вопрос экзамена студент должен продемонстрировать знание определения указанного понятия, связанных с ним особенностей реализации и применения, умение реализовать указанную операцию, а также навыки иллюстрации теоретических принципов на предложенных простых примерах.

Итоговая оценка выставляется с использованием системы «Рейтинг» ДВФУ по текущей успеваемости.

Суммарно по дисциплине можно получить 100 баллов.

В течение каждого семестра студентам последовательно выдается набор практических работ, каждая из которых имеет вес от 10% до 20%. Во втором семестре также предлагается для выполнения набор дополнительных практических заданий, каждое из задание имеет вес от 10% до 15%.

Посещаемость занятий имеет вес 2%.

Для получения экзамена в 1 и 2 семестре необходимо набрать не менее 60%. Критерии оценки:

86 % и более – «отлично»;

71-85 % – «хорошо»;

56-70 % – «удовлетворительно»;

55 % и менее – «неудовлетворительно».



## VI. КОНТРОЛЬ ДОСТИЖЕНИЯ ЦЕЛЕЙ КУРСА

Для текущей аттестации при изучении дисциплины «Машинное обучение (Machine Learning)» используются следующие оценочные средства:

1) Устный опрос (УО):

Собеседование (консультация с преподавателем) (УО-1)

2) Письменные работы (ПР):

Кейс-задача (ПР-2)

Коллоквиум (ПР-3)

Контрольная работа (ПР-4)

Конспект (ПР-7)

Разноуровневые задачи и задания (ПР-9)

Расчетно-графическая работа (ПР-10)

№ п/п	Контролируемые разделы / темы дисциплины	Коды и этапы формирования компетенций		Оценочные средства	
				текущий контроль	промежуточная аттестация
1	Раздел № 1-8 практической части курса.	УК-1	знает	УО-1, УО-3, ПР-2, ПР-10	Устный опрос
			умеет	ПР-7	Устный опрос
2	Раздел № 1-4 практической части курса.	УК-4	знает	УО-1, УО-3, ПР-2, ПР-10	Устный опрос
			умеет	ПР-7	Устный опрос
3	Раздел № 1-9 практической части курса.	ОПК-1	знает	УО-2, УО-3, ПР-2, ПР-3	Коллоквиум
			умеет	ПР-11	Устный опрос
4	Раздел № 1-9 практической части курса.	ОПК-2	знает	УО-2, УО-3, ПР-2, ПР-4	Коллоквиум
			умеет	ПР-11	Устный опрос
5	Раздел № 1-6 практической части курса.	ОПК-6	знает	УО-2, УО-3, ПР-2, ПР-4, ПР-3	Коллоквиум
			умеет	ПР-7	Устный опрос

## VII. СПИСОК УЧЕБНОЙ ЛИТЕРАТУРЫ И ИНФОРМАЦИОННО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### Основная литература

*(электронные и печатные издания)*

1. Анализ данных / В.С. Мхитарян и др. – М: Издательство Юрайт, 2019. – Текст : электронный // ЭБС Юрайт [сайт]. – URL: <https://bibli-online.ru/bcode/432178>
2. Маккинни, У. Python и анализ данных / У. Маккинни; перевод с английского А. А. Слинкина. – 2-ое изд., испр. и доп. – Москва : ДМК

Пресс, 2020. – 540 с. – ISBN 978-5-97060-590-5. – Текст: электронный // Лань: электронно-библиотечная система. – Режим доступа: <https://e.lanbook.com/book/131721>

3. Шарден, Б. Крупномасштабное машинное обучение вместе с Python [Электронный ресурс]: учебное пособие / Б. Шарден, Л. Массарон, А. Боскетти; пер. с англ. А. В. Логунова. – Электрон. дан. – Москва: ДМК Пресс, 2018. – 358 с. – Режим доступа: <https://e.lanbook.com/book/105836>. – Загл. с экрана.
4. Кук, Д. Машинное обучение с использованием библиотеки H2O [Электронный ресурс] / Д. Кук; пер. с англ. Огурцова А.Б. – Электрон. дан. – Москва: ДМК Пресс, 2018. – 250 с. – Режим доступа: <https://e.lanbook.com/book/97353>. – Загл. с экрана.
5. Коэльо, Л.П. Построение систем машинного обучения на языке Python [Электронный ресурс] / Л.П. Коэльо, В. Ричарт; пер. с англ. Слинкин А.А. — Электрон. дан. — Москва: ДМК Пресс, 2016. — 302 с. — Режим доступа: <https://e.lanbook.com/book/82818>. — Загл. с экрана.
6. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных [Электронный ресурс] / П. Флах. — Электрон. дан. — Москва: ДМК Пресс, 2015. — 400 с. — Режим доступа: <https://e.lanbook.com/book/69955>. — Загл. с экрана.
7. Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения [Электронный ресурс]: руководство / С. Рашка; пер. с англ. Логунова А.В. — Электрон.дан. — Москва: ДМК Пресс, 2017. — 418 с. — Режим доступа:<https://e.lanbook.com/book/100905>. — Загл. с экрана.

### **Дополнительная литература** (печатные и электронные издания)

1. SPSS 19: профессиональный статистический анализ данных : [практическое руководство] / А. Наследов. - Санкт-Петербург : Питер, 2011. - 399 с.
2. Макаров, А.А. Анализ данных на компьютере / Макаров А.А., Тюрин Ю.Н. – М.: МЦНМО, 2016.
3. Шитиков, В.К. Классификация, регрессия, алгоритмы Data Mining с использованием R [Электронный ресурс] / Шитиков В. К., Мастицкий С.

- Э. – URL: <https://github.com/ranalytics/data-mining> (дата обращения: 19.01.2020), <https://ranalytics.github.io/data-mining/index.html> (дата обращения: 19.01.2020).
4. MachineLearning: профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. [Электронный ресурс]. – URL: <http://www.machinelearning.ru> (дата обращения: 19.01.2020).
  5. Каталог визуализации данных. [Электронный ресурс]. – URL: <https://datavizcatalogue.com/RU/index.html> (дата обращения: 19.01.2020).
  6. Машинное обучение и анализ данных [Электронный ресурс]. – URL: <https://www.coursera.org/specializations/machine-learning-data-analysis> (дата обращения: 19.01.2020).
  7. Математика и Python для анализа данных [Электронный ресурс]. – URL: <https://www.coursera.org/learn/mathematics-and-python?specialization=machine-learning-data-analysis#syllabus> (дата обращения: 19.01.2020).
  8. Анализ данных на Python [Электронный ресурс]. – URL: <https://www.pvsm.ru/python/310645> (дата обращения: 19.01.2020).
  9. Анализ данных на Python в примерах и задачах [Электронный ресурс]. – URL: <https://compscicenter.ru/courses/data-mining-python/2018-spring/classes/>
  10. Крылов С.В. Основы интеллектуального анализа данных. – Н. Новгород: Изд-во ННГУ им. Н.И. Лобачевского, 2015.

#### **Перечень ресурсов информационно-телекоммуникационной сети «Интернет»**

1. Электронная библиотека и базы данных ДВФУ. <http://dvfu.ru/web/library/elib>
2. Электронно-библиотечная система «Лань» <http://e.lanbook.com>
3. Электронно-библиотечная система «Научно-издательского центра ИНФРА-М» <http://znanium.com>
4. Электронно-библиотечная система БиблиоТех. <http://www.bibliotech.ru>
5. Электронный каталог научной библиотеки ДВФУ. <http://ini-fb.dv-gu.ru:8000/cgi-bin/gw/chameleon>
6. Научная библиотека КиберЛенинка

## **Перечень информационных технологий и программного обеспечения**

При осуществлении образовательного процесса по дисциплине используется свободно распространяемое программное обеспечение MS Excel, GNU R, Python, а также автоматическая тестирующая система CATS ДВФУ <https://imcs.dvfu.ru/cats/>.

## **VIII.МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ**

### **Описание последовательности действий обучающихся: алгоритм изучения дисциплины**

Студентам, при освоении дисциплины «Машинное обучение (Machine Learning)», необходимо ознакомиться: с содержанием рабочей учебной программы дисциплины (далее – РПУД), с целями и задачами дисциплины, методическими разработками по данной дисциплине, имеющимся на образовательном портале и сайте кафедры, с графиком консультаций преподавателей кафедры.

При изучении дисциплины студентам необходимо дополнительно самостоятельно изучать материал. Для понимания материала и качественного его усвоения рекомендуется такая последовательность действий:

1. В течение недели выбрать время для работы с рекомендованной литературой в библиотеке или ресурсами Интернет.

2. При подготовке к практическим занятиям следующего дня необходимо сначала повторить пройденный практический материал предыдущего занятия по теме домашнего задания.

### **Методические рекомендации для подготовки к практическим занятиям**

Практическое занятие - одна из основных форм организации учебного процесса, заключающаяся в выполнении студентами под руководством преподавателя комплекса учебных заданий, с целью усвоения научно-теоретических основ учебной дисциплины, приобретения умений и навыков, опыта исследовательской деятельности.

В системе подготовки практические занятия позволяют студентам приобретать и совершенствовать универсальные и профессиональные компетенции.

Содержание практических занятий и методика их проведения должны обеспечивать развитие творческой, научно-исследовательской активности учащихся. В ходе их проведения создаются условия для развития научного мышления, аналитических умений и навыков обучающихся. Практические занятия позволяют проверить знания студентов, в связи с чем они выступают

важным средством достаточно оперативной обратной связи.

Цели практических занятий:

– помочь обучающимся сформировать, систематизировать, закрепить и углубить теоретические знания и практические умения и навыки в области математического анализа, линейной алгебры, методов оптимизации и теории вероятностей.

– научить студентов интегрировать разрозненную информацию, анализировать предметную область информационно-аналитических систем и цифровых технологий, способствовать овладению навыками и умениями применения полученных знаний в области математического анализа, линейной алгебры, методов оптимизации и теории вероятностей, в практической и научно-исследовательской деятельности, аргументировать собственную точку зрения.

При подготовке к практическим занятиям студентам необходимо придерживаться следующих рекомендаций:

– ознакомиться с тематическим планом дисциплины «Машинное обучение (Machine Learning)»;

– изучить содержание темы предстоящего практического занятия по предложенным источникам;

– составить конспект предстоящего семинара, используя предлагаемый план и рекомендованные источники;

– соотнести теоретический материал с актуальными правовыми нормами;

– зафиксировать вопросы, возникшие в процессе подготовки к практическому занятию.

Подготовка к практическим занятиям предполагает работу с учебной и научной литературой.

Студенты, не отчитавшиеся по каждой не проработанной ими на занятиях теме к началу зачётной сессии, упускают возможность получить положенные баллы за работу в соответствующем семестре.

### **Методические рекомендации по работе с литературой и составлению конспектов**

Основным методом самостоятельного овладения знаниями является работа с литературой. Это сложный процесс, требующий выработки определенных навыков, поэтому студенту нужно обязательно научиться работать с книгой.

Конспектирование источников по дисциплине «Машинное обучение (Machine Learning)» осуществляется при подготовке к практическим занятиям № 1-6.

Студенты конспектируют учебную и научную литературу. Конспектирование учебной литературы осуществляется в соответствии с темами (вопросами) практических занятий. Перечень источников для конспектирования представлен в списке основной и дополнительной литературы. В процессе подготовки конспекта возможно использование ресурсов информационно-телекоммуникационной сети «Интернет».

Осмысление литературы требует системного подхода к освоению материала. В работе с литературой системный подход предусматривает не только внимательное чтение текста и изучение специальной литературы, но и обращение к дополнительным источникам – правовым актам, законам, справочникам, энциклопедиям, словарям, которые являются основными помощниками в самостоятельной работе студента, так как глубокое изучение именно их материалов позволит студенту освоить новую научную терминологию, а затем самостоятельно оперировать теоретическими категориями и понятиями. Такого рода работа с литературой обеспечивает решение студентом поставленной перед ним задачи (подготовка к семинарскому занятию, выполнение практических заданий и т. д.).

Литература для изучения обычно выбирается из списка литературы, выданного преподавателем, либо путем самостоятельного отбора материалов. После этого непосредственно начинается изучение материала, изложенного в источнике.

При изучении материала источника необходимо обращать особое внимание на комментарии и примечания, которыми сопровождается текст. Они разъясняют отдельные места текста, дополняют изложенный материал, указывают ссылки на цитируемые источники, объясняют малоизвестные или иностранные слова.

Во время изучения литературы следует конспектировать и составлять рабочие записи прочитанного, которые могут быть сделаны и в виде простого и развернутого плана, цитирования, тезисов, резюме, аннотации, конспекта. Такие записи удлиняют процесс проработки, изучения книги, но способствуют ее лучшему осмыслению и усвоению, выработке навыков кратко и точно излагать материал.

Наиболее надежный способ собрать нужный материал - составить конспект, то есть краткое изложение своими словами содержания книги. Конспекты позволяют восстановить в памяти ранее прочитанное без дополнительного обращения к самой книге. При их составлении следует пользоваться различными приемами выделения отдельных частей текста, ключевых выражений, терминов, основных понятий (выделение абзацев, подчеркивание, написание жирным шрифтом, курсивом, использование

цветных чернил и т.п.). Желательно оставлять поля для внесения дополнений, поправок или фиксации собственных мыслей по данной записи, возможно несовпадающих с авторской точкой зрения.

При изучении литературы особое внимание следует обращать на новые термины и понятия. Понимание сущности и значения терминов способствует формированию способности логического мышления, приучает мыслить абстракциями, что важно при усвоении дисциплины. Поэтому при изучении темы курса студенту следует активно использовать универсальные и специализированные энциклопедии, словари, иную справочную литературу.

Вся рекомендуемая для изучения курса литература подразделяется на основную и дополнительную. К основной литературе относятся источники, необходимые для полного и твердого усвоения учебного материала. Необходимость изучения дополнительной литературы диктуется, прежде всего тем, что в учебной литературе нередко остаются неосвещенными современные проблемы, а также не находят отражение новые документы, события, явления, научные открытия последних лет. Поэтому дополнительная литература рекомендуется для более углубленного изучения программного материала.

#### **Методические рекомендации по выполнению различных форм самостоятельных домашних заданий**

Самостоятельная работа студентов включает в себя выполнение различного рода заданий, которые ориентированы на более глубокое усвоение материала изучаемой дисциплины. По каждой теме учебной дисциплины студентам предлагается перечень заданий для самостоятельной работы. К выполнению заданий для самостоятельной работы предъявляются следующие требования: задания должны исполняться самостоятельно и представляться в установленный срок, а также соответствовать установленным требованиям по оформлению.

Студентам следует руководствоваться графиком самостоятельной работы, определённым РПУД, выполнять все плановые задания, выдаваемые преподавателем для самостоятельного выполнения, и разбирать на консультациях неясные вопросы.

#### **Методические рекомендации по подготовке к экзамену**

Экзамен – это заключительный этап изучения дисциплины «Машинное обучение (Machine Learning)», имеющий целью проверить теоретические знания студента, его навыки и умение применять полученные знания при решении практических задач. Экзамен проводится в устной форме.

Подготовка начинается с первого занятия по дисциплине, на котором студенты получают общую установку преподавателя и перечень основных

требований к текущей и промежуточной аттестации. При этом важно с самого начала планомерно осваивать материал, руководствуясь, прежде всего, перечнем вопросов, конспектировать важные для решения учебных задач источники. В течение семестра происходят пополнение, систематизация и корректировка студенческих работ, освоение нового и закрепление уже изученного материала.

## **IX. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

Учебные занятия по дисциплине проводятся в помещениях, оснащенных соответствующим оборудованием и программным обеспечением.

Перечень материально-технического и программного обеспечения дисциплины приведен в таблице.

### **Материально-техническое и программное обеспечение дисциплины**

Наименование специальных помещений и помещений для самостоятельной работы	Оснащенность специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа
Мультимедийная аудитория: G467	Проектор DLP, 3000 ANSI Lm, WXGA 1280x800, 2000:1 EW330U Mitsubishi;; Моноблок HP ProOne 440 G3 23.8" All-in-One, диагональ экрана 23.8", разрешение экрана 1920x1080, Bluetooth, Wi-Fi, операционная система: Windows 10 Enterprise, оптический привод DVD, процессор: Intel Core i5-7500T, размер оперативной памяти: 8 ГБ, видеопроцессор: Intel HD Graphics 630, объем жесткого диска: 1Тб. Беспроводные ЛВС для обучающихся обеспечены системой на базе точек доступа 802.11a/b/g/n 2x2 MIMO(2SS). AfterEffects	Techdesigner, MAX8, VVVV, Adobe Photoshop, Adobe Premier, Adobe
Мультимедийная аудитория: G469	Проектор DLP, 4000 ANSI Lm, 1920x1080, 2000:1 FD630u Mitsubishi;; Проектор DLP, 2800 ANSI Lm, 1920x1080, 2000:1 GT1080 Optoma; Проектор DLP, 3000 ANSI Lm, WXGA 1280x800, 2000:1 EW330U Mitsubishi;; Беспроводные ЛВС для обучающихся обеспечены системой на базе точек	Techdesigner, MAX8, VVVV, Adobe Photoshop, Adobe Premier, Adobe

	доступа 802.11a/b/g/n 2x2 MIMO(2SS). Специализированное оборудование: Платформа Arduino UNO, Бесконтактный сенсорный Microsoft Kinect 2.0, Аудио система Dialog 2.0, MIDI контроллер Playtron, Одноплатный компьютер Raspberry PI	
--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

Рабочие места для людей с ограниченными возможностями здоровья оснащены дисплеями и принтерами Брайля; оборудованы: портативными устройствами для чтения плоскочечатных текстов, сканирующими и читающими машинами, видеоувеличителем с возможностью регуляции цветовых спектров; увеличивающими электронными лупами и ультразвуковыми маркировщиками.

В целях обеспечения специальных условий обучения инвалидов и лиц с ограниченными возможностями здоровья в ДВФУ все здания оборудованы пандусами, лифтами, подъемниками, специализированными местами, оснащенными туалетными комнатами, табличками информационно-навигационной поддержки.

## Х. ФОНДЫ ОЦЕНОЧНЫХ СРЕДСТВ

### Критерии оценочных средств

Для текущей аттестации при изучении дисциплины «Машинное обучение (Machine Learning)» используются следующие оценочные средства:

1) Устный опрос (УО):

Собеседование (консультация с преподавателем) (УО-1)

2) Письменные работы (ПР):

Кейс-задача (ПР-2)

Коллоквиум (ПР-3)

Контрольная работа (ПР-4)

Конспект (ПР-7)

Разноуровневые задачи и задания (ПР-9)

Расчетно-графическая работа (ПР-10)

№ п/п	Контролируемые разделы / темы дисциплины	Коды и этапы формирования компетенций	Оценочные средства		
			текущий контроль	промежуточная аттестация	
1	Раздел № 1-8 практической части курса.	УК-1	знает	УО-1, УО-3, ПР-2, ПР-10	Устный опрос
			умеет	ПР-7	Устный опрос
2	Раздел № 1-4 практической части курса.	УК-4	знает	УО-1, УО-3, ПР-2, ПР-10	Устный опрос
			умеет	ПР-7	Устный опрос
3	Раздел № 1-9 практической части курса.	ОПК-1	знает	УО-2, УО-3, ПР-2, ПР-3	Коллоквиум
			умеет	ПР-11	Устный опрос
4	Раздел № 1-9 практической части курса.	ОПК-2	знает	УО-2, УО-3, ПР-2, ПР-4	Коллоквиум
			умеет	ПР-11	Устный опрос
5	Раздел № 1-6 практической части курса.	ОПК-6	знает	УО-2, УО-3, ПР-2, ПР-4, ПР-3	Коллоквиум
			умеет	ПР-7	Устный опрос

### Текущая аттестация студентов

Текущая аттестация студентов по дисциплине «Машинное обучение (machine learning)» проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной.

Текущая аттестация по дисциплине «Машинное обучение (machine learning)» проводится в форме контрольных мероприятий (устного опроса, коллоквиума, конспекта первоисточника, решения разноуровневых практических задач) по оцениванию фактических результатов обучения студентов и осуществляется ведущим преподавателем.

Объектами оценивания выступают:

- учебная дисциплина (активность на занятиях, своевременность выполнения различных видов заданий, посещаемость всех видов занятий по аттестуемой дисциплине);
- степень усвоения теоретических знаний;
- уровень овладения практическими умениями и навыками по всем видам учебной работы;
- результаты самостоятельной работы.

Каждому объекту оценивания присваивается конкретный балл. Составляется календарный план контрольных мероприятий по дисциплине и внесения данных в АРС. По окончании семестра студент набирает определенное количество баллов, которые переводятся в пятибалльную систему оценки.

#### **Критерии оценки устного опроса:**

Результат работы студента в ходе устного ответа в виде собеседования с преподавателем оценивается по следующим критериям: полнота раскрытия вопросов; степень самостоятельности выполнения задания; и его презентация; исполнение сроков предоставления выполненных заданий; способность отвечать на вопросы преподавателя и студентов в ходе устного опроса по заданной тематике.

✓ 100-86 баллов выставляется, если студент выразил своё мнение по сформулированной проблеме, аргументировал его, точно определив ее содержание и составляющие. Продемонстрировано знание и владение навыком самостоятельной исследовательской работы по теме вопроса. Фактических ошибок, связанных с пониманием проблемы, нет

✓ 85-76 баллов работа студента характеризуется смысловой цельностью, связностью и последовательностью изложения; допущено не более 1 ошибки при объяснении смысла или содержания проблемы. Для аргументации приводятся данные отечественных и зарубежных авторов. Продемонстрированы исследовательские умения и навыки. Фактических ошибок, связанных с пониманием проблемы, нет.

✓ 75-61 баллов проведен достаточно самостоятельный анализ основных этапов и смысловых составляющих проблемы; понимание базовых основ и теоретического обоснования выбранной темы. Привлечены основные источники по рассматриваемой теме. Допущено не более 2 ошибок в смысле или содержании проблемы

✓ 60-50 баллов если работа представляет собой пересказанный или полностью переписанный исходный текст без каких бы то ни было комментариев, анализа. Не раскрыта структура и теоретическая составляющая

темы. Допущено три или более трех ошибок смыслового содержания раскрываемой проблемы.

### **Критерии оценки конспекта:**

- ✓ 100-85 баллов - выставляется студенту, если студент выразил своё мнение по сформулированной проблеме, аргументировал его, точно определив ее содержание и составляющие. Приведены данные отечественной и зарубежной литературы, статистические сведения, информация нормативно-правового характера. Студент знает и владеет навыком самостоятельной исследовательской работы по теме исследования; методами и приемами анализа теоретических и/или практических аспектов изучаемой области. Фактических ошибок, связанных с пониманием проблемы, нет; графически работа оформлена правильно
- ✓ 85-76 баллов - работа характеризуется смысловой цельностью, связностью и последовательностью изложения; допущено не более 1 ошибки при объяснении смысла или содержания проблемы. Для аргументации приводятся данные отечественных и зарубежных авторов. Продемонстрированы исследовательские умения и навыки. Фактических ошибок, связанных с пониманием проблемы, нет. Допущены одна-две ошибки в оформлении работы
- ✓ 75-61 балл – студент проводит достаточно самостоятельный анализ основных этапов и смысловых составляющих проблемы; понимает базовые основы и теоретическое обоснование выбранной темы. Привлечены основные источники по рассматриваемой теме. Допущено не более 2 ошибок в смысле или содержании проблемы, оформлении работы
- ✓ 60-50 баллов – если работа представляет собой пересказанный или полностью переписанный исходный текст без каких бы то ни было комментариев, анализа. Не раскрыта структура и теоретическая составляющая темы. Допущено три или более трех ошибок в смысловом содержании раскрываемой проблемы, в оформлении работы.

### **Оценочные средства для промежуточной аттестации**

Промежуточная аттестация студентов по дисциплине «Машинное обучение (machine learning)» проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной. Согласно учебному плану видом промежуточной аттестации по дисциплине «Математические методы анализа данных» предусмотрен экзамен, который выставляется по результатам работы в семестре. Экзамен проводится в устной форме по билетам. По результатам изучения курса предусмотрен экзамен в конце 1

семестра и экзамен в конце 2 семестра. Экзамен может быть проведён в виде продуманной системы рейтинговой оценки. Не достигшие желаемого результата магистранты могут сдавать экзамен традиционным способом в устной форме ответа на вопросы билетов.

### **Вопросы к экзамену. 1 семестр**

1. Введение. Методы оптимизации. Градиентный спуск
  - Задача регрессии, классификации.
  - Функция потерь. Оптимизация. Перебор по сетке.
  - Производная, частные производные, градиент.
  - Градиентный спуск, проблема выбора шага.
  - Стохастический градиентный спуск. Использование момента.
  - Adagrad, Adadelata, Adam.
  - RMSProp\*.
2. Линейная регрессия
  - Постановка задачи линейной регрессии.
  - Метод наименьших квадратов.
  - Ковариация, корреляция.
  - Критерий R<sup>2</sup>.
  - Анализ остатков.
3. Глобальная оптимизация. Генетический алгоритм
  - Многопараметрическая оптимизация.
  - Доминанция и оптимальность по Парето.
  - Функция качества (fitness). Аппроксимация качества.
  - Общая идея генетического алгоритма.
  - Представление генома.
  - Методы селекции: пропорционально качеству, универсальная выборка (stochastic universal sampling), с наследием (reward-based), турнир. Стратегия элитизма.
    - Методы кроссовера. Двух и многоточечный, равномерный (по подмножествам), для перестановок.
    - Мутация. Влияние на скорость обучения.
    - Управление популяцией. Сегрегация, старение, распараллеливание.
    - Генетическое программирование.
4. Метод ближайших соседей (k-NN)
  - Понятие и свойства метрики. Ослабление требования к неравенству треугольника.
  - Базовый алгоритм классификации методом 1-NN и k-NN.

- Преимущества и недостатки.
  - Метрики L1, L2, Хемминга, Левенштейна, косинусное расстояние.
  - Потеря точности нормы в высоких размерностях.
  - Нормализация координат. Предварительная трансформация пространства признаков.
  - Метрика Махаланобиса.
  - Кросс-валидация методом "без одного" (leave one out).
  - Определение границ, показатель пограничности.
  - Сжатие по данным. Понятия выброса, прототипа, усвоенной точки.
5. Алгоритм Харта (Hart).
- Регрессия методом k-NN.
  - Взвешенные соседи.
  - Связь с градиентным спуском. Стохастическая формулировка, softmax.
  - Метод соседних компонент (neighbour component analysis)\*.
  - Связь с выпуклой оптимизацией. Метод большого запаса (Large margin NN).
  - Оптимизация классификатора, k-d дерева.
  - Хеши чувствительные к локальности, хеши сохраняющие локальность\*.

### **Вопросы к экзамену. 2 семестр**

1. Наивный байесов классификатор
  - Условная вероятность. Байесово решающее правило. Обновление вероятностей.
  - Наивный классификатор, предположение о независимости признаков.
  - Оценка плотности распределения для числовых признаков.
  - Алгоритмические оптимизации.
  - Алгоритм EM.
2. Логистическая регрессия
  - Сигмоид.
  - Метод наибольшего правдоподобия.
  - Логистическая регрессия для меток  $-1, 1$ .
3. Деревья решений
  - Понятие дерева решений.
  - Борьба с оверфиттингом: bagging, выборки признаков.
  - Ансамбли, случайный лес (Random Forest).
  - Понятие энтропии, определение информации по Шеннону.

- Метрики: примеси Джини (Gini impurity), добавленная информация (information gain).
  - Деревья регрессии. Метрика вариации.
  - Непрерывные признаки. Использование главных компонент вместо признаков.
  - Сокращение дерева (pruning).
4. Кластеризация
- Задача обучения без учителя, применения при эксплораторном анализе.
  - Неметрическая кластеризация: функция схожести, компоненты связности и остовные деревья, иерархическая кластеризация снизу вверх.
  - Метрики, понятие центроида и представителя класса.
  - Центроидные алгоритмы: k-means, k-medoid.
  - Алгоритмы, основанные на плотности: DBSCAN, OPTICS.
  - Алгоритмы, основанные на распределении: сумма гауссиан.
  - Нечёткая кластеризация, алгоритм c-means.
  - Метрики качества: leave-one-out, силуэт, индекс Дэвиса-Болдина (Davies-Bouldin), индекс Данна (Dunn).

### **Критерии выставления оценки студенту на экзамене**

- Оценка «отлично» выставляется студенту, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, свободно справляется с задачами, вопросами и другими видами применения знаний, причем не затрудняется с ответом при видоизменении заданий, использует в ответе материал монографической литературы, правильно обосновывает принятое решение, владеет разносторонними навыками и приемами выполнения практических задач.
- Оценка «хорошо» выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения.
- Оценка «удовлетворительно» выставляется студенту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала, испытывает затруднения при выполнении практических работ.