

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования

«Московский физико-технический институт
(национальный исследовательский университет)»
(МФТИ)

УТВЕРЖДАЮ

Ректор МФТИ

д-р физ.-мат. наук, профессор,
член-корреспондент РАН



И.Н. Кудрявцев

« 08 » сентября 2020 г.

**Дополнительная профессиональная
программа повышения квалификации**

**«Методика преподавания курса «Машинное обучение» для учащихся 10
класса»**

УГСН 01.00.00 Математика и механика

Направление подготовки 01.03.02 Прикладная математика и информатика

ОКВЭД – Образование

85.41 Образование дополнительное детей и взрослых

Москва 2020

1. Характеристика программы

1.1. Цель реализации программы

Целью реализации программы является совершенствование профессиональных компетенций обучающихся в области методики преподавания курса «Машинное обучение» для учащихся 10 (ИТ) классов.

1.2. Совершенствуемые и/или приобретаемые компетенции

Компетенции, формируемые и совершенствуемые в результате обучения, представлены в таблице 1.

таблица 1

№ п/п	Компетенция	Направление подготовки: «Прикладная математика и информатика» код: 01.03.02
		Квалификация: бакалавриат
		Код компетенции
1.	Способен применять фундаментальные знания, полученные в области математических и (или) естественных наук, и использовать их в профессиональной деятельности	ОПК 1 из федерального государственного образовательного стандарта высшего образования – бакалавриат по направлению подготовки 01.03.02 Прикладная математика и информатика, утв. приказом Министерства образования и науки Российской Федерации от 10 января 2018 №9

1.3. Планируемые результаты обучения

Планируемые результаты обучения представлены в таблице 2.

таблица 2

№ п/п	Уметь - знать	«Прикладная математика и информатика» код: 01.03.02
		Квалификация: бакалавриат
		Код компетенции
1.	Знать: 1. Классификацию задач машинного обучения. 2. Особенности языка программирования Python и библиотеки NumPy. 3. Методы машинного обучения.	ОПК-1

	4. Алгоритм решения задач бинарной классификации методом линейной регрессии / методом k-ближайших соседей Уметь: Решать задачи бинарной классификации методом линейной регрессии / методом k-ближайших соседей	
2.	Знать: 1. Особенности методики преподавания курса «Машинное обучение» для ИТ-класса 2. Алгоритм составления заданий для индивидуальной работы обучающихся ИТ-классов и проведения текущего контроля по курсу «Машинное обучение» Уметь: Составлять задания для индивидуальной работы обучающихся ИТ-классов и проведения текущего контроля по курсу «Машинное обучение»	ОПК-1

1.4. Категории обучающихся

Уровень образования – ВО, область профессиональной деятельности – обучение информатике в ИТ-классах на уровне среднего общего образования.

1.5. Форма обучения

Очная (с применением дистанционных образовательных технологий).

1.6. Трудоемкость программы

18 академических часов.

1.7. Режим обучения

1 неделя (18 часов в неделю).

2. Содержание программы

2.1. Учебный (тематический) план

Учебный (тематический) план программы представлен в таблице 3.

таблица 3

№ п/п	Наименование разделов (модулей) и тем	Аудиторные учебные занятий, учебные работы			Внеаудит. работа	Формы контроля	Трудоемкость
		Всего ауд., час	Лекции	Практ. занятия	с/р		
1.	<i>Машинное обучение</i>	7	5	2	3		10
1.1	Введение в машинное обучение	1	1				1
1.2	Язык программирования Python	2	2				2
1.3	Библиотека NumPy	1	1				1
1.4	Методы машинного обучения	3	1	2	3	Практическое задание №1, Практическое задание №2	6
2.	<i>Методика преподавания курса «Машинное обучение»</i>	4	4		2		6
2.1	Структура и особенности курса «Машинное обучение» для 10 класса	1	1				1
2.2	Особенности организации групповой и индивидуальной работы обучающихся во время учебных занятий	1	1		1	Практическое задание №3	2
2.3	Особенности организации текущего и промежуточного контроля	1	1		1	Практическое задание №4	2
2.4	Программное обеспечение и оборудование компьютерного класса	1	1				1

	Итоговая аттестация	2		2		Зачет на основании совокупности выполненных работ и результатов итогового тестирования	2
	Итого:	13	9	4	5		18

2.2. Учебная программа

Содержание учебной программы приведено в таблице 4.

таблица 4

№ п/п	Виды учебных занятий, учебных работ	Содержание
Раздел 1. Машинное обучение		
Тема 1.1 Введение в машинное обучение	<i>Лекция (1 час)</i>	Определение машинного обучения. Примеры задач, в которых используется машинное обучение. Обучение с учителем и без учителя. Классификация задач машинного обучения: регрессия, классификация, кластеризация, уменьшение размерности, выявление аномалий.
Тема 1.2 Язык программирования Python	<i>Лекция (2 час)</i>	Установка Python. Введение в Python: типы данных, ввод и вывод данных, арифметические операции, условия и циклы, списки.
Тема 1.3 Библиотека NumPy	<i>Лекция (1 час)</i>	Введение в линейную алгебру (вектора и операции над ними, скалярное и векторное произведение, матрицы и операции над матрицами). Одномерные массивы. Многомерные массивы. Методы и операции над массивами. Манипуляции с элементами массива.
Тема 1.4 Методы машинного обучения	<i>Лекция (1 час)</i>	Линейная регрессия. Метод наименьших квадратов. Решение задач бинарной классификации методом линейной регрессии. Решение задач регрессии методом линейной регрессии. Метод k-ближайших соседей (KNN). Алгоритм решения задач бинарной классификации методом линейной регрессии / методом k-ближайших соседей.
	<i>Практическое занятие (2 часа)</i>	Практическое задание 1. Решение задач бинарной классификации методом линейной регрессии для предсказания успеваемости учащихся
	<i>Самостоятельная работа (3 час.)</i>	Практическое задание 2. Решение задач бинарной классификации методом «k-ближайших

		соседей» для разделения учащихся на учебные группы по признакам
Раздел 2. Методика преподавания курса «Машинное обучение»		
Тема 2.1 Структура и особенности курса «Машинное обучение» для 10 класса	<i>Лекция (1 час)</i>	Цели и задачи курса «Машинное обучение». Особенности методики преподавания курса «Машинное обучение» для ИТ-класса. Основные понятия, закономерности, технологии, рассматриваемые в курсе. Формируемые компетенции школьников. Декомпозиция результатов обучения. Входные требования к обучению по программе «Машинное обучение». Проведение диагностики готовности обучающихся к освоению курса, организация входного тестирования. Последовательность тем изучения курса. Взаимосвязь тем курса между собой. Типы учебных занятий, предусмотренные программой курса «Машинное обучение». Характеристика типов уроков. Определение типа урока в зависимости от его целевой установки. Структура уроков различного типа. Последовательность этапов урока и тайминг.
Тема 2.2 Особенности организации групповой и индивидуальной работы обучающихся во время учебных занятий	<i>Лекция (1 час)</i>	Роль групповой и индивидуальной работы в формировании компетенций учащихся. Виды заданий для организации групповой работы в классе. Виды индивидуальных заданий для учащихся. Уровневые задания различной сложности. Проектная деятельность учащихся: особенности организации. Виды заданий для организации групповой и индивидуальной самостоятельной работы учащихся. Оценка эффективности самостоятельной работы учащихся. Алгоритм составления заданий для индивидуальной работы обучающихся ИТ-класса и проведения текущего контроля по курсу «Машинное обучение».
	<i>Самостоятельная работа (1 час.)</i>	Практическое задание 3. Составление задания для индивидуальной работы обучающихся по теме «Методы машинного обучения»
Тема 2.3 Особенности организации текущего и промежуточного контроля Особенности организации текущего и	<i>Лекция (1 час)</i>	Виды контроля в учебном процессе. Приемы организации текущего контроля на уроке. Средства контроля. Типы тестов. Тестовые задания закрытого и открытого типов. Тестовые задания со множественным выбором и тестовые задания с выбором одного правильного ответа. Тестовые задания с альтернативным выбором, тестовые задания с установлением соответствия, тестовые задания с установлением

промежуточного контроля		последовательности. Решение задач. Задания в виде ноутбука. Определение критериев оценивания.
	<i>Самостоятельная работа (1 час.)</i>	Практическое задание 4. Разработка заданий для проведения текущего контроля по одной из тем курса «Машинное обучение»
Тема 2.4 Программное обеспечение и оборудование компьютерного класса	<i>Лекция (1 час)</i>	Характеристика материально-технической базы, необходимой для реализации образовательного процесса по курсу «Машинное обучение». Программное обеспечение курса.
Итоговая аттестация	<i>(2 час.)</i>	Итоговое тестирование Зачет на основании совокупности выполненных работ и результатов тестирования

3. Формы аттестации и оценочные материалы

3.1. Форма аттестации

Итоговая аттестация по программе является обязательной. Итоговая аттестация осуществляется на основании совокупности работ, выполненных на положительную оценку.

Форма итоговой аттестации: зачет на основании совокупности выполненных работ и результатов тестирования.

Критерии оценивания:

выполнение / не выполнение тестирования не менее, чем на 75%;

выполнение / не выполнение практических заданий, предусмотренных программой.

таблица 5

Наименование модуля, разделов и тем	Основные показатели оценки	Формы и методы контроля и оценки
<i>Машинное обучение</i>	ОПК-1	Тест Практическое задание 1 Практическое задание 2
<i>Методика преподавания курса «Машинное обучение»</i>	ОПК-1	Практическое задание 3 Практическое задание 4

Критерий	Соответствует	Не соответствует
выполнение / не выполнение тестирования не менее, чем на 75%	1 балл	0 баллов
выполнение / не выполнение практических заданий, предусмотренных программой	1 балл	0 баллов

Оценивание проводится преподавателем на основе представленных критериев и шкалы оценки.

Зачет выставляется слушателю, если по результатам оценивания слушатель получает 2 балла.

3.2. Оценочные материалы

Примеры заданий

Примерный тест:

Вопрос 1.

У машинного обучения есть ряд задач. Как называется та, что направлена на предсказание значения той или иной непрерывной числовой величины для входных данных?

1. Переобучение
2. Регрессия
3. Кластеризация
4. Классификация

Вопрос 2.

Какой из видов машинного обучения основывается на взаимодействии обучаемой системы со средой?

1. Обучение с подкреплением
2. Обучение без учителя
3. Обучение с учителем
4. Глубинное обучение

Вопрос 3.

В таблице приведены сведения о выставленных на продажу квартирах

Квартира	Площадь	Материал стен	Наличие лифта
А	50	кирпич	да
В	70	панель	да
С	100	Дерево	нет

Каким является признак «Площадь»?

1. Числовым
2. Бинарным
3. Номинальным
4. Порядковым

Вопрос 4.

Даны значения признака Р для шести объектов выборки $P=(-1,0,4,2,1,2)$.

Среднее значение признака Р равно: 4/3

Медиана признака Р равна: 1.5

Вопрос 5.

Дана таблица с пропущенным значением признака Р3 для последнего объекта.

Объект	Р1	Р2	Р3
А	1	1	0
В	2	0	1
С	0	1	5
Д	4	1	?

Восстановим пропущенное значение с помощью различных подходов. Если пропущенное значение заменить на среднее значение признака Р3, то оно будет равно 2

Если пропущенное значение заменить на медиану признака Р3, то оно будет равно 1

Вопрос 6.

Даны два вектора значений признаков $P1=(0,1,2)$ $P2=(2,1,0)$. Чему равно расстояние между ними в евклидовой метрике?

$$d = 2\sqrt{2}$$

Вопрос 7.

Следующая таблица содержит информацию об оценках, выставленных фильмам.

Человек\Фильм	Молчание ягнят	Титаник	Матрица	Гарри Поттер
Вася	5	5	5	3
Петя	5	3	4	4
Маша	2	5	3	5
Саша	3	4	4	?

Требуется оценить, какую оценку поставит Саша фильму «Гарри Поттер». Сделаем это с помощью метрики Манхэттен (для простоты вычислений данные в таблице нормировать не нужно). Для этого подсчитаем расстояния от Саши до других людей, используя информацию из первых 3 столбцов. Чему равна ожидаемая оценка для «Гарри Поттера» (округлить до одного знака после запятой)? Метрика Манхэттен имеет вид:

$$d_1(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|_1 = \sum_{i=1}^n |p_i - q_i|$$

6	5	4	3	2	3	4	5
5	4	3	2	1	2	3	4
4	3	2	1		1	2	3
5	4	3	2	1	2	3	4
6	5	4	3	2	3	4	5
7	6	5	4	3	4	5	6
8	7	6	5	4	5	6	7
9	8	7	6	5	6	7	8

N = 4.3

Вопрос 8.

Таблица содержит сведения о заказах анонимных пользователей в интернет-магазине.

Товар\Заказ	1	2	3	4	5	6
A	1	0	1	0	1	0
B	0	1	1	1	0	0
C	1	1	0	1	1	0
D	1	1	0	1	1	1

По строкам отложены наименования товаров, а столбцы — номера заказов (в ячейке стоит 1, если товар входит в соответствующий заказ; 0 — в противном случае). При построении рекомендательной системы необходимо оценить степень похожести товаров с помощью вычисления евклидовой метрики. Наиболее похожим (близким) на товар А будет товар С и расстояние между этими товарами равно $\sqrt{3}$.

Вопрос 9.

Пусть объекты из следующей таблицы образуют один кластер.

Объект	P1	P2	P3
A	1	1	0
B	0	2	-1
C	2	3	1
D	1	0	4

Нужно указать три числа — координаты центра кластера.

1 2 1

Вопрос 10.

Дана таблица с объектами

Объект	P1	P2	P3
A	1	1	0
B	0	2	-1
C	2	3	1
D	1	0	4

В данной задаче выбросы будем искать по следующему правилу: «Выбросом будет считаться объект, у которого суммарное расстояние от него до остальных объектов выборки наибольшее». Таким образом, в указанной таблице выбросом будет (написать имя объекта в виде заглавной латинской буквы):

D , сумма расстояний от него до остальных объектов будет равна: **20**

При вычислении использовать метрику Манхэттен, нормализацию не проводить.

Вопрос 11.

Дана таблица с объектами

	P1	P2
A	4	2
B	3	2
C	1	-1
D	-1	1
E	0	4

Какие объекты попадают в окружность с центром (1,2) и радиусом 2,5 (используется tax-метрика)?

B, D, E

Вопрос 12.

Дана таблица с объектами

	P1	P2
A	4	2
B	3	2
C	1	-1
D	-1	1
E	0	4

Планируется разбить объекты на 2 кластера. На первой итерации работы алгоритма k-means были выбраны точки (2,3) и (1,1). После первой итерации алгоритма к кластеру, определяемому первой точкой, будут отнесены объекты (метрика Манхэттен)... **A, B, E**

Используется следующий подвид алгоритма:

Шаг 1. Генерируем k случайных точек – центры кластеров.

Шаг 2. Объект будет отнесен к тому кластеру, чей центр расположен ближе всех к этому объекту.

Шаг 3. Пересчитываем центры кластеров, возврат на Шаг 2.

Цикл 2-3 крутится, пока изменяются центры кластеров.

Вопрос 13.

Объекты в следующей таблице имеют один нецелевой признак X.

Объект	X	Y
A	0	0
B	1	1
C	2	0
D	3	3

Постройте модель линейной регрессии для предсказания признака Y.

$$Y = 0.8 * X - 0.2$$

Вопрос 14.

Объекты в следующей таблице имеют два нецелевых признака X_1 , X_2 .

Объект	X_1	X_2	Y
A	0	3	0
B	1	2	1
C	2	1	0
D	3	0	3

Постройте модель линейной регрессии с регуляризацией для предсказания признака Y (значение константы регуляризации C положить равным 1).

$$Y = 0,172 + 0.62 * X_1 - 0.104 * X_2$$

Вопрос 15.

Даны объекты тренировочной выборки. На вход алгоритму kNN подается объект F с признаками $X_1=0$, $X_2=0$.

Объект	X_1	X_2	Y
A	-1	0	1
B	1	2	0
C	2	-2	0
D	-3	-1	1
E	3	2	1

При $k=3$ объект F будет отнесен к классу **0**

При $k=5$ объект F будет отнесен к классу **1**

При решении задачи использовать евклидову метрику, нормализацию данных не проводить.

Вопрос 16.

Задача классификации - это:

- 1. множество объектов, разделенных на классы**
2. исследование влияние одного или нескольких признаков на объект
3. определение порядка признака согласно рангу

Вопрос 17.

Что, из ниже перечисленного, относится к обучающей выборке?

1. классификация данных
2. **объекты с известными ответами**
3. алгоритм, решающий функцию

Вопрос 18.

Решаются ли задачи классификации и регрессии при помощи метода " k-ближайших соседей"?

1. **Да**
2. Только задачи классификации (регрессии не решаются)
3. Только задачи регрессии (классификации не решаются)

Вопрос 19.

Назовите метод, недостаток которого приведен ниже:

"Существует сложность выбора меры "близости", от этой меры главным образом зависит объем множества записей, которые нужно хранить в памяти для достижения удовлетворительной классификации или прогноза"

1. метод байесовской классификации
2. **метод "k-ближайших соседей"**
3. метод опорных векторов

Вопрос 20.

Вы хотите предсказать суммы, которые клиенты потратят на оплату трафика в разные месяцы, исходя из истории их предыдущего потребления. Это задача:

1. **регрессии**
2. классификации
3. регрессии и классификации.

Практические задания

Практическое задание 1. Решение задач бинарной классификации методом линейной регрессии для предсказания успеваемости учащихся

Метод линейной множественной регрессии с регуляризацией может быть использован для предсказания успеваемости учащихся.

Для этого зададим m учеников, и n оценок $x_1, x_2, \dots, x_n$ для каждого из них по тем или иным признакам. Вектор Y будет состоять из разницы результатов выходного теста и входного тестов для каждого ученика. Все признаки, даже бинарные, должны быть переведены в 10-балльную шкалу. Целью решения задачи является получение функции:

$$f(x_1, x_2, \dots, x_n) = w_0 + w_1 \cdot x_1 + \dots + w_n \cdot x_n$$

Данная функция является моделью линейной множественной регрессии с регуляризацией. Регуляризация нужна для того, чтобы модель была более корректной, менее подверженной к ошибкам переобучения, влиянию случайных отклонений и случайному получению завышенных весовых коэффициентов.

Без регуляризации функция ошибок в методе наименьших квадратов равна:

$$\mathcal{L}(X, \vec{y}, \vec{w}) = \frac{1}{2}(\vec{y} - X\vec{w})^T (\vec{y} - X\vec{w})$$

Здесь X – матрица коэффициентов, полученная из оценок каждого ученика, к которой слева приписали единичный столбец. Например, пусть признаки учеников имеют следующие значения:

Ученик	Занятия спортом	Кружки	Посещаемость	Активность на уроках
Иванов	5	3	9	5
Петров	7	4	8	5
Сидоров	1	8	9	3

Тогда матрица X будет иметь вид:

$$\begin{pmatrix} 1 & 5 & 3 & 9 & 5 \\ 1 & 7 & 4 & 8 & 5 \\ 1 & 1 & 8 & 9 & 3 \end{pmatrix}$$

Для регуляризации добавим слагаемое.

$$R(\vec{w}) = \frac{1}{2} \|\vec{w}\|_2^2 = \frac{1}{2} \sum_{j=1}^m w_j^2 = \frac{1}{2} \vec{w}^T \vec{w}$$

Тогда новая функция оценки ошибки будет иметь вид:

$$\mathcal{L}(X, \vec{y}, \vec{w}) = \frac{1}{2} (\vec{y} - X\vec{w})^T (\vec{y} - X\vec{w}) + \frac{\lambda}{2} \vec{w}^T \vec{w}$$

Вычислим производную по параметрам.

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \vec{w}} &= \frac{\partial}{\partial \vec{w}} \left(\frac{1}{2} (\vec{y} - X\vec{w})^T (\vec{y} - X\vec{w}) + \frac{\lambda}{2} \vec{w}^T \vec{w} \right) \\ &= \frac{\partial}{\partial \vec{w}} \left(\frac{1}{2} (\vec{y}^T \vec{y} - 2\vec{y}^T X\vec{w} + \vec{w}^T X^T X\vec{w}) + \frac{\lambda}{2} \vec{w}^T \vec{w} \right) \\ &= -X^T \vec{y} + X^T X\vec{w} + \lambda \vec{w} \end{aligned}$$

И найдём решение в явном виде.

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \vec{w}} = 0 &\Leftrightarrow -X^T \vec{y} + X^T X\vec{w} + \lambda \vec{w} = 0 \\ &\Leftrightarrow X^T X\vec{w} + \lambda \vec{w} = X^T \vec{y} \\ &\Leftrightarrow (X^T X + \lambda E) \vec{w} = X^T \vec{y} \\ &\Leftrightarrow \vec{w} = (X^T X + \lambda E)^{-1} X^T \vec{y} \end{aligned}$$

Придумайте набор признаков, критерии их оценки по 10-балльной шкале, реализуйте модель линейной множественной регрессии с регуляризацией с помощью python, используя numpy.

Написанная программа должна брать параметры для учеников из текстового файла и строить модель регрессии.

В качестве признаков могут быть также текущие оценки, сдача рефератов, своевременность сдачи домашних заданий и баллы по ним, и тому подобное.

Сделайте визуализацию результатов: гистограмму величин полученных коэффициентов w , а также сравнение значения среднеквадратичной ошибки и модуля значений компонент вектора Y .

Сделайте модуль оценки применимости модели и реализуйте программу на основе него, которая определяет все возможные наборы подходящих признаков для модели линейной регрессии. Для того, чтобы можно было бы применять модель множественной линейной регрессии, при анализе множественной корреляции должны быть установлены следующие факты:

- зависимая переменная тесно зависит от независимых переменных (тесноту связи, как и в случае парной корреляции, показывают значения коэффициента корреляции)

- нет тесной зависимости между независимыми переменными.

Коэффициенты множественной корреляции между зависимой переменной Y и независимыми переменными $X_1, X_2, \dots, X_m$ записываются в корреляционную матрицу:

$$\begin{pmatrix} r_{yy} & r_{yx_1} & r_{yx_2} & \dots & r_{yx_m} \\ r_{x_1y} & r_{x_1x_1} & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{x_2y} & r_{x_2x_1} & r_{x_2x_2} & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{x_my} & r_{x_mx_1} & r_{x_mx_2} & \dots & r_{x_mx_m} \end{pmatrix}$$

На диагонали стоят единицы, а недиагональные элементы являются обычными парными коэффициентами корреляции. Для любой пары векторов значений X и Y этот коэффициент считается следующим образом.

$$\begin{aligned} r &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \\ &= \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \cdot \sqrt{n \sum y_i^2 - (\sum y_i)^2}}. \end{aligned}$$

После вычисления корреляционной матрицы можно вычислить парные коэффициенты частной корреляции. С помощью коэффициента частной корреляции определяется теснота связи между двумя факторами при фиксировании или исключении влияния остальных. Коэффициент частной корреляции рассчитывается по следующей формуле:

$$r_{y12} = \frac{r_{y1} - r_{y2}r_{12}}{\sqrt{(1-r_{y2}^2)(1-r_{12}^2)}}$$

Здесь 1 и 2 – номера признаков, между которыми оценивают тесноту связи.

Значение коэффициента корреляции	Линейная зависимость
-1	функциональная отрицательная
0	не существует
1	функциональная положительная
$ r < 0,5$	слабая
$0,5 \leq r \leq 0,8$	средней тесноты
$ r > 0,8$	тесная

Критерии оценивания:

- программа берет / не берет параметры для учеников из текстового файла;
- проверяет / не проверяет корректность модели регрессии;
- программа строит / не строит модель регрессии;
- программа имеет / не имеет визуализацию.

Оценивание

таблица 7

Критерий	Соответствует	Не соответствует
программа берет / не берет параметры для учеников из текстового файла	1 балл	0 баллов
проверяет / не проверяет корректность модели регрессии	1 балл	0 баллов
программа строит / не строит модель регрессии	1 балл	0 баллов
программа имеет / не имеет визуализацию	1 балл	0 баллов

Оценивание проводится преподавателем на основе представленных критериев и шкалы оценки.

Практическое задание считается выполненным, если по результатам оценивания слушатель получает не менее 3 баллов.

Практическое задание 2. Решение задач бинарной классификации методом «k-ближайших соседей» для разделения учащихся на учебные группы по признакам

Зачастую удобно разделять обучающихся на группы не просто по принципу общего уровня подготовки, но и по принципу сходства знаний и умений. Для этого можно провести серию тестов, проверяющих разные знания и умения, и ранжируя каждое из них по 100-балльной шкале.

Пусть проведено N тестов для M учащихся, и мы хотим разбить их на L групп. Эта задача может быть решена с помощью метода «k-ближайших соседей». Используется следующий подвид алгоритма:

Шаг 1. Генерируем $k-2$ случайных точек – центры кластеров.

Шаг 2. В качестве оставшихся 2 точек берём в качестве одного центра кластера точку, все координаты которой равны нулю, в качестве второго центра кластера – точку, все координаты которой равны 100.

Шаг 3. Объект будет отнесен к тому кластеру, чей центр расположен ближе всех к этому объекту.

Шаг 4. Пересчитываем центры кластеров, возврат на Шаг 3.

Цикл 3-4 крутится, пока изменяются центры кластеров.

Напишите на Python программу, которая по результатам сдачи этих тестов разделяет учеников на группы и выводит результаты разделения на экран:

- 1) размер полученных групп и список принадлежащих к ним учеников;
- 2) центры и диаметры каждого кластера (диаметром называется максимальное расстояние между двумя учениками внутри одного кластера)

3) матрицу расстояний между парами кластеров (расстоянием между кластерами называется наименьшее расстояние между двумя учениками, один из которых принадлежит первому кластеру, а второй – второму)

При решении задачи использовать евклидову метрику.

Критерии оценивания:

- программа выводит / не выводит данные о размере полученных групп учеников и список принадлежащих к ним учеников;
- программа выводит / не выводит данные о центрах и диаметрах каждого кластера;
- программа выводит / не выводит матрицу расстояний между парами кластеров.

Оценивание

таблица 7

Критерий	Соответствует	Не соответствует
программа выводит / не выводит данные о размере полученных групп учеников и список принадлежащих к ним учеников	1 балл	0 баллов
программа выводит / не выводит данные о центрах и диаметрах каждого кластера	1 балл	0 баллов
программа выводит / не выводит матрицу расстояний между парами кластеров	1 балл	0 баллов

Оценивание проводится преподавателем на основе представленных критериев и шкалы оценки.

Практическое задание считается выполненным, если по результатам оценивания слушатель получает не менее 2 баллов.

Практическое задание 3. Составление задания для индивидуальной работы обучающихся по теме «Методы машинного обучения»

Цель – сформировать умение составлять задания для индивидуальной работы обучающихся по теме «Методы машинного обучения».

Необходимо разработать задание для индивидуальной работы обучающихся по совершенствованию умения решать задачи бинарной классификации методом линейной регрессии или методом k-ближайших соседей.

Требования к оформлению работы:

- титульный лист – указывается название дополнительной профессиональной программы, название работы, ФИО слушателя;
- текст задания для индивидуальной работы – должен содержать следующие разделы:
 - Исходные данные
 - Постановка задачи
 - Требования к оформлению
 - Критерии и способы оценивания

Критерии оценивания:

- содержание разработанного задания соответствует / не соответствует теме «Методы машинного обучения»;
- разработанное задание позволяет / не позволяет оценить умение школьников решать задачи бинарной классификации методом линейной регрессии или методом k-ближайших соседей;
- работа соответствует / не соответствует установленным требованиям к оформлению.

Оценивание

таблица 7

Критерий	Соответствует	Не соответствует
содержание разработанного задания соответствует / не соответствует теме «Методы машинного обучения»	1 балл	0 баллов
разработанное задание позволяет / не позволяет оценить умение школьников решать задачи бинарной классификации методом линейной регрессии или методом k-ближайших соседей	1 балл	0 баллов
работа соответствует / не соответствует установленным требованиям к оформлению	1 балл	0 баллов

Оценивание проводится преподавателем на основе представленных критериев и шкалы оценки.

Практическое задание считается выполненным, если по результатам оценивания слушатель получает 3 балла.

Практическое задание 4. Разработка заданий для проведения текущего контроля по одной из тем курса «Машинное обучение»

Цель – сформировать умение разрабатывать задания для проведения текущего контроля по курсу «Машинное обучение».

Требования к оформлению работы:

– титульный лист – указывается название дополнительной профессиональной программы, название работы, ФИО слушателя;

– текст задания для проведения текущего контроля – должен содержать следующие разделы:

– Название темы, по которой проводится текущий контроль (по выбору слушателя: *(Введение в машинное обучение; Язык программирования Python; Библиотека NumPy; Методы машинного обучения)*)

– Перечень знаний / умений, на проверку сформированности которых направлено задание

– Метод контроля (опрос, тест, практическая проверка и др.)

– Текст задания

– Критерии и способы оценивания

Критерии оценивания:

– содержание разработанного задания соответствует / не соответствует выбранной теме;

– разработанное задание позволяет / не позволяет оценить знания / умения школьников, указанные в перечне;

– работа соответствует / не соответствует установленным требованиям к оформлению.

Оценивание

таблица 8

Критерий	Соответствует	Не соответствует
содержание разработанного задания соответствует / не соответствует выбранной теме	1 балл	0 баллов
разработанное задание позволяет / не позволяет оценить знания / умения школьников, указанные в перечне	1 балл	0 баллов

работа соответствует / не соответствует установленным требованиям к оформлению	1 балл	0 баллов
--	--------	----------

Оценивание проводится преподавателем на основе представленных критериев и шкалы оценки.

Практическое задание считается выполненным, если по результатам оценивания слушатель получает 3 балла.

4. Организационно-педагогические условия реализации программы

4.1. Учебно-методическое обеспечение и информационное обеспечение программы

4.1.1. Список литературы:

1. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных /пер. с английского А.А. Слинкина - М. : ДМК Пресс 2015. -400 с.

2. Дэви С., Мейсман А., Мохамед А. Основы DataScienceи BigData. Python и наука о данных - СПб. : Питер, 2017. - 336 с. ил. – серия «Библиотека программиста».

3. Хенрик Б., Джозеф Р., Феверолф М. Машинное обучение.– СПб. : Питер, 2017, 336 с.

4. Бастиан Ш., Массаран Л., Боскети А. Крупномасштабное машинное обучение вместе с Python. - М. : ДМК Пресс 2017. - 358 с.

5. Рашка С. Python и машинное обучение. - М. : ДМК Пресс 2017. -418 с.

6. Коэльо Л.П., Ричарт В. Построение систем машинного обучения на языке Python. 2016. 302 с.

4.1.2. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

1. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2014. — 739 p.
2. Bishop C. M. Pattern Recognition and Machine Learning. — Springer, 2006. — 738 p.
3. William M. Bolstad. Introduction to Bayesian Statistics, 2nd Edition // Wiley-Interscience; 2nd edition.
4. Dr. Anasse Bari, Mohamed Chaouchi, Tommy Jung. Predictive Analytics For Dummies // For Dummies; 2 edition, 2016.
5. Домингос П. Верховный алгоритм. – М.: Издательство «МИФ». – 336 с.
6. Мерков А. Б. Распознавание образов. Введение в методы статистического обучения. 2011. 256 с.
7. Мерков А. Б. Распознавание образов. Построение и обучение вероятностных моделей. 2014. 238 с.
8. Нейт Сильвер. Сигнал и Шум. Почему одни прогнозы сбываются, а другие — нет // Азбука-Аттикус, КоЛибри, 2015.

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

1. Популярная статья «машинное обучение для чайников» <https://newtonew.com/tech/machine-learning-novice> - Дата обращения 03.11.2020
2. Перевод статьи разработчика алгоритмов машинного обучения, бизнес-консультанта и популярного автора Ганта Лаборде «*Machine Learning: from Zero to Hero*». <https://neurohive.io/ru/osnovy-data-science/vvedenie-v-mashinnoe-obuchenie-kto-ego-primenjaet-i-kak-stat-razrabotchikom/> - Дата обращения 03.11.2020
3. Обзор «машинное обучение для людей» https://vas3k.ru/blog/machine_learning/ - Дата обращения 03.11.2020

4. Метод ближайших k-соседей <https://basegroup.ru/community/articles/knn>
- Дата обращения 03.11.2020
5. Форекс-Python <https://github.com/MicroPyramid/forex-python>
<https://forex-python.readthedocs.io/en/latest/usage.html>, описание функций и
установка - Дата обращения 03.11.2020
6. Как создать собственную нейросеть с нуля
<https://neurohive.io/ru/tutorial/kak-sozdat-nejronnuju-set-s-nulja-na-jazyke-python/> на
python - Дата обращения 03.11.2020
7. Обучи свою первую нейросеть – руководство на сайте tensorflow
<https://www.tensorflow.org/tutorials/keras/classification?hl=ru> и учебный курс по
tensorflow <https://neurohive.io/ru/tutorial/tensorflow-tutorial-tenzory-i-vektory/> - Дата
обращения 03.11.2020
8. Метод опорных векторов для задач классификации , объяснение SVM с
нуля <https://habr.com/ru/company/ods/blog/484148/> - Дата обращения 03.11.2020

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

таблица 9

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
Аудитория	Лекции	Персональные компьютеры/ноутбуки, доступ в Интернет, проектор, камера
	Сам. работа	Персональные компьютеры/ноутбуки, доступ в Интернет
Аудитория	Итоговая аттестация	Персональные компьютеры/ноутбуки, доступ в Интернет, проектор, камера

5. Организация образовательного процесса

В таблице 10 описаны образовательные технологии.

таблица 10

№ п/п	Вид занятия	Форма проведения занятий	Цель
1	Лекция	изложение теоретического материала	получение теоретических знаний по дисциплине

6. Составители программы

Воронцов Игорь Андреевич, ассистент кафедры математического анализа института естественных наук и математики уральского федерального университета (УрФУ);

Сырцова Елена Леонидовна, PhD, кандидат педагогических наук, доцент