

МИНОБРНАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Институт прикладной математики и компьютерных наук

УТВЕРЖДАЮ
Директор института прикладной
математики и компьютерных наук
А.В. Замятин

« 16 » _____ 2022 г.

Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине
(Оценочные средства по дисциплине)

Статистические методы машинного обучения - II

по направлению подготовки

01.04.02 Прикладная математика и информатика

Направленность (профиль) подготовки:

Big Data and Data Science

ОС составил(и):

канд. физ.-мат. наук, доцент
доцент кафедры теории вероятностей
и математической статистики

Т.В. Кабанова

Рецензент:

канд. физ.-мат. наук, доцент
доцент кафедры системного анализа
и математического моделирования

Ж.Н. Зенкова

Оценочные средства одобрены на заседании учебно-методической комиссии
института прикладной математики и компьютерных наук (УМК ИПМКН).

Протокол от 12.05.2022 г. № 4

Председатель УМК ИПМКН,
д-р техн. наук, профессор

С.П. Сущенко

Оценочные средства (ОС) являются элементом системы оценивания сформированности компетенций у обучающихся в целом или на определенном этапе ее формирования.

ОС разрабатывается в соответствии с рабочей программой (РП) дисциплины.

1. Компетенции и результаты обучения, формируемые в результате освоения дисциплины

Компетенция	Индикатор компетенции	Код и наименование результатов обучения (планируемые результаты обучения, характеризующие этапы формирования компетенций)	Критерии оценивания результатов обучения			
			Отлично	Хорошо	Удовлетворительно	Неудовлетворительно
ОПК-1 – способен решать актуальные задачи фундаментальной и прикладной математики	ИОПК-1.3 Демонстрирует навыки использования основных понятий, фактов, концепций, принципов математики, информатики и естественных наук для решения практических задач, связанных с прикладной математикой и информатикой.	ОР-1.1.1. обучающийся сможет: - выбирать адекватные метод для решения поставленной задачи; - реализовывать выбранный метод в программе анализа данных; - делать выводы и интерпретировать полученные результаты.	Демонстрация высокого уровня знаний математических основ и базовых понятий, которые необходимы для понимания статистических методов анализа данных.	В целом успешные, но содержащие отдельные пробелы знания о математических основах и базовых понятиях, которые необходимы для понимания статистических методов анализа данных.	Фрагментарное, неполное знание без грубых ошибок математически базовых понятий, которые необходимы для понимания статистических методов анализа данных.	Не знает математические основы и базовые понятия, которые необходимы для понимания статистических методов анализа данных.

ОПК-2 – способен совершенствовать и реализовывать новые математические методы решения прикладных задач.	ИОПК-2.1 Использует результаты прикладной математики для освоения, адаптации новых методов решения задач в области своих профессиональных интересов.	ОР-2.1.1. обучающийся сможет: адаптировать модели для описания процессов реальной предметной области	Демонстрация высокого уровня знаний математических основ и базовых понятий, которые необходимы для понимания статистических методов анализа данных.	В целом успешные, но содержащие отдельные пробелы знания о математических основах и базовых понятиях, которые необходимы для понимания статистических методов анализа данных.	Фрагментарное, неполное знание без грубых ошибок математически базовых понятий, которые необходимы для понимания статистических методов анализа данных.	Не знает математические основы и базовые понятия, которые необходимы для понимания статистических методов анализа данных.
	ИОПК-2.2 Реализует и совершенствует новые методы, решения прикладных задач в области профессиональной деятельности.	ОР-2.2.1. обучающийся сможет: реализует и интерпретирует построенные модели для описания процессов реальной предметной области				
	ИОПК-2.3. Проводит качественный и количественный анализ полученного решения с целью построения оптимального варианта.	ОР-2.3.1. Обучающийся сможет проводить качественный и количественный анализ построенных моделей и полученных на их основании прогнозов и выбирать наиболее оптимальный в соответствии с выбранной метрикой				

2. Этапы формирования компетенций и виды оценочных средств

№	Этапы формирования компетенций (разделы дисциплины)	Код и наименование результатов обучения	Вид оценочного средства (тесты, задания, кейсы, вопросы и др.)
1.	Множественная регрессия	ОР-1.1.1, ОР-2.1.1, ОР-2.2.1, ОР-2.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
2.	Дополнительные вопросы регрессионного анализа.	ОР-1.1.1, ОР-2.1.1, ОР-2.2.1, ОР-2.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
3.	Задачи классификации	ОР-1.1.1, ОР-2.1.1, ОР-2.2.1, ОР-2.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста

3. Типовые контрольные задания или иные материалы, необходимые для оценки образовательных результатов обучения

3.1. Типовые задания для проведения текущего контроля успеваемости по дисциплине.

Текущий контроль осуществляется на основании посещаемости и проверки лабораторных работ.

Примеры заданий для лабораторных работ

Лабораторная работа. Множественная регрессия. Фиктивные переменные

Выполняется в R.

Задание.

Импортировать таблицу с данными в R.

1. Построить графики для визуализации данных и их взаимосвязей.
2. Проверить связи факторов друг с другом и их влияние на зависимую целевую переменную.
3. Построить и провести анализ множественной модели регрессии целевой переменной от всех представленных количественных и порядковых факторов.
4. Провести обработку и кодирование категориальных факторов.
5. Построить и провести анализ множественной модели регрессии с учетом всех предложенных факторов.
6. Удалить незначимые факторы. Построить окончательную модель.
7. Проверить остатки модели на нормальность.
8. Задать новое наблюдение со своими значениями признаков и построить прогноз целевого показателя для него.

Лабораторная работа. Логистическая регрессия.

Задание.

Сформировать наблюдения, связанные однофакторной логистической регрессией.

1. Задать объем выборки $n = 20 : 50$.
2. Значения фактора x сформировать как реализацию целочисленной равномерно распределенной случайной величины в интервале $[a, b]$.
3. Задать нормально распределенный шум $\varepsilon \sim N(0, \sigma)$.
4. Определить регрессионную модель

$$P(x) = \frac{e^{\theta_0 + \theta_1 x + \varepsilon}}{1 + e^{\theta_0 + \theta_1 x + \varepsilon}}.$$

5. Значение бинарной зависимой переменной определить как

$$y_i = \begin{cases} 0, & P(x_i) < \frac{1}{2}; \\ 1, & P(x_i) \geq \frac{1}{2}. \end{cases}$$

Все параметры задать самостоятельно, ориентируясь на диаграмму рассеяния.

6. Оценить параметры модели.
Проверить общее качество модели.

3.2. Типовые задания для проведения промежуточной аттестации по дисциплине
Примерный перечень теоретических вопросов и тем для подготовки к зачету:

1. Нелинейные модели и линейаризация.
2. Случай смещенного шума.
3. Случай коррелированных гомоскедастических наблюдений.
4. Случай некоррелированных гетероскедастических наблюдений.
5. Мультиколлинеарность.
6. Фиктивные переменные.
7. Постановка задачи классификации.
8. Логистическая регрессия.
9. Метрики качества бинарного классификатора.
10. ROC-анализ.

4. Методические материалы, определяющие процедуры оценивания образовательных результатов обучения

4.1. Методические материалы для оценки текущего контроля успеваемости по дисциплине.

Для текущей аттестации необходимо иметь посещаемость не менее 75% от всех проведенных на момент аттестации занятий и сдать все данный на момент аттестации лабораторные работы. Лабораторные работы оцениваются на зачтено/незачтено.

4.2. Методические материалы для проведения промежуточной аттестации по дисциплине.

Результаты экзамена определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Для теста из 10 вопросов. За каждый вопрос в зависимости от его сложности можно получить от 1 до 3 баллов. Максимально 20.

отлично	От 18 до 20 баллов
хорошо	От 14 до 17 баллов
удовлетворительно	От 11 до 13 баллов
неудовлетворительно	От 0 до 10 баллов