Министерство науки и высшего образования РФ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Национальный исследовательский университет «МЭИ»

Направление подготовки/специальность: 27.03.04 Управление в технических системах

Наименование образовательной программы: Системы и технические средства автоматизации и управления

Уровень образования: высшее образование - бакалавриат

Форма обучения: Очная

Оценочные материалы
по дисциплине
Интеллектуальный анализ данных

Москва
2022

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ РАЗРАБОТАЛ:

Преподаватель	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ» Сведения о владельце ЦЭП МЭИ Владелец: Толчеев В.О. Идентификатор: Rfbd680da-TolcheevVO-692f9924	В.О. Толчеев
(должность)	(подпись)	(расшифровка подписи)

СОГЛАСОВАНО:

Руководитель образовательной программы	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ» Сведения о владельце ЦЭП МЭИ Владелец: Шилин Д.В. Идентификатор: R495daf18-ShilinDV-59db3f0e	Д.В. Шилин
(должность, ученая степень, ученое звание)	(подпись)	(расшифровка подписи)
Заведующий выпускающей кафедры	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ» Сведения о владельце ЦЭП МЭИ Владелец: Бобряков А.В. Идентификатор: R2c90f415-BobriakovAV-70dec1fa	А.В. Бобряков
(должность, ученая степень, ученое звание)	(подпись)	(расшифровка подписи)

ОБЩАЯ ЧАСТЬ

Оценочные материалы по дисциплине предназначены для оценки: достижения обучающимися запланированных результатов обучения по дисциплине, этапа формирования запланированных компетенций и уровня освоения дисциплины.

Оценочные материалы по дисциплине включают оценочные средства для проведения мероприятий текущего контроля успеваемости и промежуточной аттестации.

Формируемые у обучающегося компетенции:

1. ПК-2 Способен разрабатывать системы и технические средства автоматизации и управления на основе современных программных и аппаратных средств

ИД-1 Может формировать выборки и подготавливать данные для проведения анализа

ИД-2 Формулирует критерии качества, разработки, настройки и тестирования алгоритмов анализа данных

и включает:

для текущего контроля успеваемости:

Форма реализации: Защита задания

1. Защита лабораторной работы №1 (Лабораторная работа)
2. Защита лабораторной работы №2 (Лабораторная работа)
3. Защита лабораторной работы №3 (Лабораторная работа)
4. Защита лабораторной работы №4 (Лабораторная работа)
5. Защита расчетного задания «Разведочный анализ данных и построение моделей» (Расчетно-графическая работа)

Форма реализации: Письменная работа

1. Контрольная работа №1 (Контрольная работа)
2. Контрольная работа №2 (Контрольная работа)
3. Контрольная работа №3 (Контрольная работа)

БРС дисциплины

5 семестр

Раздел дисциплины	Веса контрольных мероприятий, %								
	Индекс КМ:	КМ-1	КМ-2	КМ-3	КМ-4	КМ-5	КМ-6	КМ-7	КМ-8
	Срок КМ:	4	6	8	10	12	12	15	16
Машинное обучение с учителем (регрессия)									
Постановка задачи машинного обучения. Формирование и исследование выборок		+	+	+		+		+	
Регрессионный анализ		+	+	+		+		+	
Машинное обучение без учителя (кластеризация)									
Методы кластеризации					+	+		+	+

Методы снижения размерности и визуализации				+	+		+	+
Машинное обучение с учителем (классификация)								
Методы классификации					+	+	+	+
Методы построения ансамблей классификаторов (коллективных решений)					+	+	+	+
Прикладные задачи машинного обучения								
Машинное обучение с подкреплением					+		+	
Методика анализа исследуемого датасета					+		+	
Вес КМ:	10	10	10	10	10	10	20	20

$Общая часть/Для промежуточной аттестации$

СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ТЕКУЩЕГО КОНТРОЛЯ

I. Оценочные средства для оценки запланированных результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Индекс компетенции	Индикатор	Запланированные результаты обучения по дисциплине	Контрольная точка
ПК-2	ИД-1$_{ПК-2}$ Может формировать выборки и подготавливать данные для проведения анализа	Знать: правила формирования выборок и методы предварительной обработки данных способы выявления неизвестных закономерностей из многомерных данных Уметь: формировать выборки и проводить предварительную обработку данных использовать стандартные пакеты прикладных программ и открытые библиотеки для решения практических задач	Контрольная работа №1 (Контрольная работа) Защита лабораторной работы №1 (Лабораторная работа) Контрольная работа №2 (Контрольная работа) Защита лабораторной работы №2 (Лабораторная работа)
ПК-2	ИД-2$_{ПК-2}$ Формулирует критерии качества, разработки, настройки и тестирования алгоритмов анализа данных	Знать: способы машинного обучения и методы интеллектуального анализа данных методику проведения	Контрольная работа №3 (Контрольная работа) Защита лабораторной работы №3 (Лабораторная работа) Защита лабораторной работы №4 (Лабораторная работа) Защита расчетного задания «Разведочный анализ данных и построение моделей» (Расчетно-графическая работа)

		обработки и анализа данных Уметь: обосновывать выбор метода анализа данных и оценивать качество полученных результатов проводить сбор, обработку и анализ данных с использованием современных информационных технологий и методик	

II. Содержание оценочных средств. Шкала и критерии оценивания

КМ-1. Контрольная работа №1

Формы реализации: Письменная работа
Тип контрольного мероприятия: Контрольная работа
Вес контрольного мероприятия в БРС: 10
Процедура проведения контрольного мероприятия: На группу выдается 4 варианта заданий, в каждом содержится трехмерный массив данных (т.е. содержится три переменные). Надо провести предварительный анализ данных и оценить силу связи между переменными.

Краткое содержание задания:
Для предварительного анализа выборки строится диаграмма Тьюки для каждой из переменных (признаков), оцениваются выборочные характеристики, выявляются выбросы. С помощью корреляционного анализа оценивается связи между переменными и обосновывается целесообразность отбросить один из признаков.

Контрольные вопросы/задания:

Знать: правила формирования выборок и методы предварительной обработки данных	1.Как вы понимаете «Информативность переменной»? 2.В каких случаях предпочтительно использовать выборочную медиану вместо среднеарифметического?

Описание шкалы оценивания:

Оценка: 5
Нижний порог выполнения задания в процентах: 85
Описание характеристики выполнения знания: Оценка "отлично" выставляется, если задание выполнено в полном объеме или выполнено преимущественно верно.

Оценка: 4
Нижний порог выполнения задания в процентах: 65
Описание характеристики выполнения знания: Оценка "хорошо" выставляется, если большинство вопросов раскрыто и выбрано верное направление для решения задачи.

Оценка: 3
Нижний порог выполнения задания в процентах: 50
Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется, если задание преимущественно выполнено.

КМ-2. Защита лабораторной работы №1

Формы реализации: Защита задания
Тип контрольного мероприятия: Лабораторная работа
Вес контрольного мероприятия в БРС: 10
Процедура проведения контрольного мероприятия: Изучение и анализ двух заданных датасетов. Проведение разведочного анализа данных и выявление информативных признаков. Построение парной и множественной регрессии. Оценка адекватности полученных моделей. Прогнозирование зависимой переменной. Проведение самостоятельного исследования по третьему датасету, заданному преподавателем.

Краткое содержание задания:

Построить диаграмму Тьюки. Проанализировать корреляционные зависимости между исследуемыми переменными. Определить входные и выходные переменные. Построить регрессионные модели и оценить значимость коэффициентов регрессии. Выполнить дисперсионный анализ остатков. Проанализировать избыточность и адекватность полученных моделей. Провести предсказание зависимой переменной по новым данным.

Контрольные вопросы/задания:

Уметь: формировать выборки и проводить предварительную обработку данных	1.Как построить диаграмму Тьюки? 2.Как рассчитать корреляционные зависимости?

Описание шкалы оценивания:

Оценка: 5
Нижний порог выполнения задания в процентах: 85
Описание характеристики выполнения знания: Оценка "отлично" выставляется, если задание выполнено в полном объеме или выполнено преимущественно верно.

Оценка: 4
Нижний порог выполнения задания в процентах: 65
Описание характеристики выполнения знания: Оценка "хорошо" выставляется, если большинство вопросов раскрыто и выбрано верное направление для решения задачи.

Оценка: 3
Нижний порог выполнения задания в процентах: 50
Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется, если задание преимущественно выполнено.

КМ-3. Контрольная работа №2

Формы реализации: Письменная работа
Тип контрольного мероприятия: Контрольная работа
Вес контрольного мероприятия в БРС: 10
Процедура проведения контрольного мероприятия: На группу выдается 4 варианта заданий, в каждом содержится трехмерный массив данных. Надо построить уравнение парной регрессии и оценить адекватность полученной модели.

Краткое содержание задания:

По имеющемуся трехмерному массиву обосновать выбор входных и выходной переменной. Построить регрессию. Оценить адекватность полученной модели с помощью расчета коэффициента детерминации, скорректированного коэффициента детерминации и критерия Дурбина-Ватсона.

Контрольные вопросы/задания:

Знать: способы выявления неизвестных закономерностей из многомерных данных	1.Как рассчитывается статистика в критерии Дурбина-Ватсона? 2.Чем отличаются коэффициент детерминации и скорректированный коэффициент детерминации?

Описание шкалы оценивания:

Оценка: 5
Нижний порог выполнения задания в процентах: 85

Описание характеристики выполнения знания: Оценка "отлично" выставляется, если задание выполнено в полном объеме или выполнено преимущественно верно.

Оценка: 4
Нижний порог выполнения задания в процентах: 65
Описание характеристики выполнения знания: Оценка "хорошо" выставляется, если большинство вопросов раскрыто и выбрано верное направление для решения задачи.

Оценка: 3
Нижний порог выполнения задания в процентах: 50
Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется, если задание преимущественно выполнено.

КМ-4. Защита лабораторной работы №2

Формы реализации: Защита задания
Тип контрольного мероприятия: Лабораторная работа
Вес контрольного мероприятия в БРС: 10
Процедура проведения контрольного мероприятия: Визуализация исходных данных в двумерном и трехмерном пространстве. Применение иерархических и неиерархических методов кластеризации с различными настройками параметров. Обоснование наличия (отсутствия) устойчивых кластеров на основе сравнительного анализа результатов, полученных с помощью различных методов. Анализ возможных «разногласий» методов по отдельным элементам выборки.

Краткое содержание задания:
Провести визуализацию исходных данных с помощью многомерного шкалирования и метода главных компонент, сделать предположение о степени однородности выборки. Применить метод иерархического кластерного анализа (ИКА) и построить дендрограммы для разных настроек ИКА. Применить методы к-средних и DSCAN. Сравнить результаты кластеризации трех методов. Обосновать число кластеров и их состав. Рассчитать метрики качества кластеризации.

Контрольные вопросы/задания:

Уметь: использовать стандартные пакеты прикладных программ и открытые библиотеки для решения практических задач	1.Как выбираются центроиды в методе к-средних? 2.Как построить дендрограмму?

Описание шкалы оценивания:

Оценка: 5
Нижний порог выполнения задания в процентах: 85
Описание характеристики выполнения знания: Оценка "отлично" выставляется, если задание выполнено в полном объеме или выполнено преимущественно верно.

Оценка: 4
Нижний порог выполнения задания в процентах: 65
Описание характеристики выполнения знания: Оценка "хорошо" выставляется, если большинство вопросов раскрыто и выбрано верное направление для решения задачи.

Оценка: 3
Нижний порог выполнения задания в процентах: 50

Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется, если задание преимущественно выполнено.

КМ-5. Контрольная работа №3

Формы реализации: Письменная работа
Тип контрольного мероприятия: Контрольная работа
Вес контрольного мероприятия в БРС: 10
Процедура проведения контрольного мероприятия: На группу выдается 4 варианта заданий, в каждом содержится трехмерный массив данных (два признака и метка класса). Надо провести кластеризацию и классификацию наблюдений.

Краткое содержание задания:
Проводится кластеризация исходных данных с помощью иерархического кластерного анализа (ИКА). Строятся центроиды классов и проводится классификация. Сравниваются результаты классификации и кластеризации.

Контрольные вопросы/задания:

Знать: методику проведения обработки и анализа данных	1.Должны ли совпадать результаты кластеризации и классификации? 2.Чем отличается правило ближайшего соседа в ИКА и метод ближайшего соседа, которые используется для классификации?

Описание шкалы оценивания:

Оценка: 5
Нижний порог выполнения задания в процентах: 85
Описание характеристики выполнения знания: Оценка "отлично" выставляется, если задание выполнено в полном объеме или выполнено преимущественно верно.

Оценка: 4
Нижний порог выполнения задания в процентах: 65
Описание характеристики выполнения знания: Оценка "хорошо" выставляется, если большинство вопросов раскрыто и выбрано верное направление для решения задачи.

Оценка: 3
Нижний порог выполнения задания в процентах: 50
Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется, если задание преимущественно выполнено.

КМ-6. Защита лабораторной работы №3

Формы реализации: Защита задания
Тип контрольного мероприятия: Лабораторная работа
Вес контрольного мероприятия в БРС: 10
Процедура проведения контрольного мероприятия: Выбор методов классификации для анализа датасетов. Выбор критериев качества классификации. Настройка параметров методов и оценка качества классификации. Сравнение результатов классификации, анализ элементов выборки, на которых все методы допустили ошибки. Формирование предложений по улучшению результатов классификации.

Краткое содержание задания:
Проверить выполнения предположений дискриминантного анализа (ДА) и применить ДА. Применить деревья решений, настроить параметры. Применить наивный

байесовский метод и метод к-ближайших соседей (к-БС). Для метода к-БС настроить параметры. Сопоставить результаты (качество классификации), которые достигаются каждым из методов, сделать вывод, объяснить полученные результаты.

Контрольные вопросы/задания:

Уметь: обосновывать выбор метода анализа данных и оценивать качество полученных результатов	1.Как выбирается число ближайших соседей в методе к-БС? 2.Как построить дерево решений?

Описание шкалы оценивания:

Оценка: 5
Нижний порог выполнения задания в процентах: 85
Описание характеристики выполнения знания: Оценка "отлично" выставляется, если задание выполнено в полном объеме или выполнено преимущественно верно.

Оценка: 4
Нижний порог выполнения задания в процентах: 65
Описание характеристики выполнения знания: Оценка "хорошо" выставляется, если большинство вопросов раскрыто и выбрано верное направление для решения задачи.

Оценка: 3
Нижний порог выполнения задания в процентах: 50
Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется, если задание преимущественно выполнено.

КМ-7. Защита лабораторной работы №4

Формы реализации: Защита задания
Тип контрольного мероприятия: Лабораторная работа
Вес контрольного мероприятия в БРС: 20
Процедура проведения контрольного мероприятия: Выдача индивидуального задания (датасета). Изучение датасета, проведение разведочного анализа. Формулирование вывода – какие модели для выборки можно построить. Реализация всех допустимых методов интеллектуального анализа данных. Анализ качества полученных моделей, обоснование их применимости.

Краткое содержание задания:
Проверить выполнения предположений дискриминантного анализа (ДА) и применить ДА. Применить деревья решений, настроить параметры. Применить наивный байесовский метод и метод к-ближайших соседей (к-БС). Для метода к-БС настроить параметры. Сопоставить результаты (качество классификации), которые достигаются каждым из методов, сделать вывод, объяснить полученные результаты.

Контрольные вопросы/задания:

Уметь: проводить сбор, обработку и анализ данных с использованием современных информационных технологий и методик	1.Как выявить информативные признаки? 2.Как оценить качество моделей?

Описание шкалы оценивания:

Оценка: 5

Нижний порог выполнения задания в процентах: 85

Описание характеристики выполнения знания: Оценка "отлично" выставляется, если задание выполнено в полном объеме или выполнено преимущественно верно.

Оценка: 4

Нижний порог выполнения задания в процентах: 65

Описание характеристики выполнения знания: Оценка "хорошо" выставляется, если большинство вопросов раскрыто и выбрано верное направление для решения задачи.

Оценка: 3

Нижний порог выполнения задания в процентах: 50

Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется, если задание преимущественно выполнено.

КМ-8. Защита расчетного задания «Разведочный анализ данных и построение моделей»

Формы реализации: Защита задания
Тип контрольного мероприятия: Расчетно-графическая работа
Вес контрольного мероприятия в БРС: 20
Процедура проведения контрольного мероприятия: Каждый студент получает индивидуальное задание в соответствии с его номером в журнале и номером группы. Задание выполняется с помощью ППП Matlab и Statistica или открытых программных библиотек Python.

Краткое содержание задания:

Рассчитать ковариационную и корреляционную матрицы для анализируемого трехмерного признака. Вычислить оценки частных коэффициентов корреляции. Сравнить значения парного и частного коэффициентов корреляции. Проверить гипотезу (при уровне значимости α=0,05) об их статистической значимости. Найти оценку множественного коэффициента корреляции. Рассчитать оценки коэффициентов парной регрессии и проверить значимость полученного уравнения (при уровне значимости α=0,05).

Для исходных данных составить матрицу евклидовых расстояний и построить дендрограмму. С помощью ППП MATLAB рассчитать главные компоненты. Определить относительные доли суммарной дисперсии, обусловленные одной и двумя главными компонентами.

Контрольные вопросы/задания:

Знать: способы машинного обучения и методы интеллектуального анализа данных	1.В чем заключаются различия между задачами прогнозирования, кластеризации и классификации? 2.Какие методы разведочного анализа данных вам известны?

Описание шкалы оценивания:

Оценка: 5

Нижний порог выполнения задания в процентах: 85

Описание характеристики выполнения знания: Оценка "отлично" выставляется, если задание выполнено в полном объеме или выполнено преимущественно верно.

Оценка: 4

Нижний порог выполнения задания в процентах: 65

Описание характеристики выполнения знания: Оценка "хорошо" выставляется, если большинство вопросов раскрыто и выбрано верное направление для решения задачи.

Оценка: 3
Нижний порог выполнения задания в процентах: 50
Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется, если задание преимущественно выполнено.

СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

5 семестр

Форма промежуточной аттестации: Экзамен

Пример билета

Предварительный анализ выборки, выявление выбросов. Метод ближайшего соседа и его модификации. Задача.

Процедура проведения

Экзамен с оценкой проводится в устной форме по билетам. На подготовку ответа студенту отводится 45 минут.

I. Перечень компетенций/индикаторов и контрольных вопросов проверки результатов освоения дисциплины

1. Компетенция/Индикатор: ИД-1$_{ПК-2}$ Может формировать выборки и подготавливать данные для проведения анализа

Вопросы, задания

1.Многомерное нормальное распределение.
2.Метрики расстояния и меры близости.
3.Иерархический кластерный анализ, способы объединения кластеров.
4.Основные этапы анализа датасета.

Материалы для проверки остаточных знаний

1.Укажите последовательность проведения анализа данных (датасетов):

Ответы:

Построение регрессионной зависимости или классификатора.
Анализ качества построенной модели.
Разведочный анализ (Предварительный анализ выборки, выявление выбросов, анализ корреляций, определение информативных переменных, визуализация).
Выбор критерия качества.

Верный ответ: #1 Разведочный анализ (Предварительный анализ выборки, выявление выбросов, анализ корреляций, определение информативных переменных, визуализация). #2 Выбор критерия качества. #3 Построение регрессионной зависимости или классификатора. #4 Анализ качества построенной модели.

2.Как называется зависимость, описываемая следующей формулой:

$$f(X) = \frac{1}{\sqrt{(2\pi)^M |K|}} \exp\{\frac{1}{2}(X - m_x)^T K^{-1}(X - m_x)\}$$

Здесь M – размерность вектора X, $mx = M[X]$ – вектор математического ожидания, К – ковариационная матрица размера [MxM], $|K| = \det(K)$ – определитель ковариационной матрицы, Т – знак транспонирования.

Верный ответ: многомерный нормальный закон распределения

3.Понятия генеральной совокупности и выборки. Основные способы формирования выборок

Верный ответ: Генеральной совокупностью называют множество всех мыслимых (возможных) результатов наблюдений над случайной величиной, полученных при данном комплексе условий. Выборка из генеральной совокупности – это конечный

набор значений случайной величины, полученный в результате наблюдений. Число элементов выборки N называется ее объемом (или размером)

4.Понятие репрезентативной выборки и основные способы ее формирования

Верный ответ: Выборка называется репрезентативной (представительной), если она достаточно полно характеризует генеральную совокупность и, значит, по ней возможно определить интересующие исследователя закономерности. Для обеспечения репрезентативности выборки чаще всего используется случайный (рандомизированный) отбор элементов. При случайном отборе (например, с помощью генератора случайных чисел) вероятность попадания в выборку для каждого элемента генеральной совокупности одинаковы. При этом отбор может быть “с возвращением”, когда извлеченный элемент возвращается обратно в генеральную совокупность и может быть извлечен повторно, и “без возвращения”. Для достаточно больших генеральных совокупностей оба подхода практически идентичны.

5.Как называется и что характеризует показатель, который рассчитывается по формуле $S_{xx}^2 = \frac{1}{N}\sum_{j=1}^{N}(X_j - \bar{X})(X_j - \bar{X})$

Верный ответ: по указанной формуле рассчитывается оценка дисперсии. Дисперсия является мерой разброса значений случайной величины относительно её математического ожидания

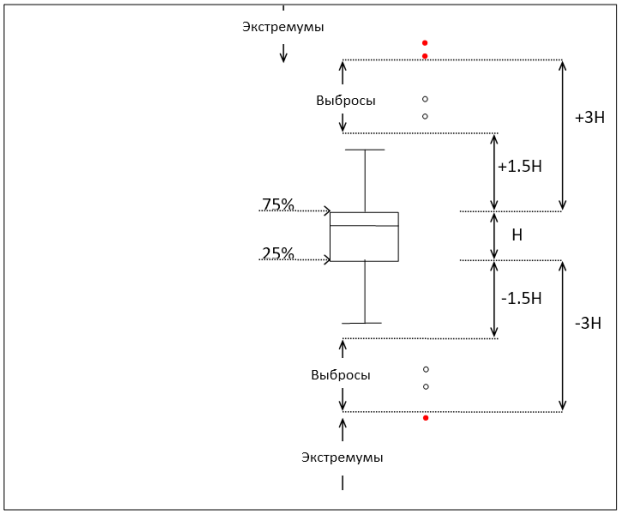
6.Расшифруйте смысл величин в формуле выборочного коэффициента корреляции. Что характеризует

и в каких пределах изменяется

$$\hat{r}_{xy} = \frac{S_{xy}}{S_{xx}S_{yy}}$$

Верный ответ: В числителе находится оценка ковариации, в знаменателе – произведение оценок среднеквадратических отклонений. Выборочный коэффициент корреляции показывает силу линейной связи между двумя случайными величинами, принимает значения в диапазоне [-1;1].

7.Как строится диаграмма Тьюки, какие характеристики она содержит?



Верный ответ: Для построения ДТ вариационные ряд делится тремя квартилями на 4 части. ДТ представляет собой прямоугольник («ящик») высотой H, верхняя и нижняя граница которого соответствуют 75-му и 25-му процентилю (или первой и третьей квартилям), с отходящими от него «усами», размер которых определяется максимальным и минимальным значениями случайной величиной (размахом). Внутри «ящика» заключены 50% выборочных значений, горизонтальной линией, обозначается медиана – второй квартиль. Для выявления выбросов с помощью диаграммы размаха Тьюки необходимо найти высоту прямоугольника H и отсечь

наблюдения, которые лежат на расстоянии, превышающем 1.5H от верхней и нижней границы «ящика». Иногда вместо размаха и медианы используют среднее арифметическое значение и среднеквадратическое отклонение. Примечание: Квартили — это процентили, которые делят набор данных на четверти. Первый квартиль - Q1 равен 25-ому процентилю, третий квартиль - Q3 равен 75-ому процентилю.

2. Компетенция/Индикатор: ИД-2$_{ПК-2}$ Формулирует критерии качества, разработки, настройки и тестирования алгоритмов анализа данных

Вопросы, задания

1.Выбор функции регрессии.
2.Определение коэффициентов парной и множественной регрессии.
3.Свойства оценок метода наименьших квадратов.
4.Деревья решений (ДР). Способы выявления информативных признаков.

Материалы для проверки остаточных знаний

1.Какой коэффициент корреляции применяется для проверки адекватности регрессионной модели?

Ответы:

парный коэффициент корреляции,
частный коэффициент корреляции,
коэффициент детерминации.

Верный ответ: парный коэффициент корреляции, частный коэффициент корреляции, *коэффициент детерминации.

2.Как называются характеристики, рассчитанные по следующим формулам? Какая из них используется в регрессионном анализе для оценки дисперсии шума?

$$S^2_{\text{общ.}} = \frac{1}{N-1}\sum_{j=1}^{N}(y_j - \bar{y})^2;$$

$$S^2_{\text{регр.}} = \frac{1}{\kappa-1}\sum_{j=1}^{N}(\hat{y}_j - \bar{y})^2;$$

$$S^2_{\text{остат.}} = \frac{1}{N-k}\sum_{j=1}^{N}(y_j - \hat{y}_j)^2.$$

Здесь N – размер выборки, k – число оцениваемых параметров.

Верный ответ: приведены формулы для расчета оценок общей, регрессионной и остаточных дисперсий. Для оценки дисперсии шума в регрессионном анализе используется оценка остаточной дисперсии

3.Какие основные показатели качества используются при классификации данных?

Верный ответ: Полнота (Recall), точность (Precision), доля правильных ответов (Accuracy=1-ошибка).

4.Для чего используются обучающая и тестовая выборки? Могут ли они содержать общие элементы (т.е. может ли одно и то же наблюдение входить в обе выборки).

Верный ответ: Обучающая выборка используется для настройки параметров модели, а тестовая для независимой проверки качества модели. Обучающая и тестовая выборки не могут содержать общих элементов.

5.Какие параметры надо задавать при проведении иерархического кластерного анализа?

Верный ответ: Мера близости между наблюдениями и правило объединения кластеров.

6.Какие параметры настраиваются в методе к-ближайших соседей?

Верный ответ: Мера близости между наблюдениями и параметр к - число соседей.

7.Какие методы классификации можно использовать для построения регрессионных зависимостей?

Ответы:

наивный байесовский метод;
метод к-ближайших соседей;
деревья решений;
логистическая регрессия;
дискриминантный анализ;

Верный ответ: наивный байесовский метод; *метод к-ближайших соседей; *деревья решений; логистическая регрессия; дискриминантный анализ;

8.Что такое сбалансированная и несбалансированная выборки?

Верный ответ: в сбалансированной выборке все классы имеют приблизительно одинаковый размер, в несбалансированной выборке размеры классов могут существенно различаться.

II. Описание шкалы оценивания

Оценка: 5
Нижний порог выполнения задания в процентах: 85
Описание характеристики выполнения знания: Работа выполнена в рамках "продвинутого" уровня. Ответы даны верно, четко сформулированные особенности практических решений.

Оценка: 4
Нижний порог выполнения задания в процентах: 65
Описание характеристики выполнения знания: Работа выполнена в рамках "базового" уровня. Большинство ответов даны верно. В части материала есть незначительные недостатки.

Оценка: 3
Нижний порог выполнения задания в процентах: 50
Описание характеристики выполнения знания: Работа выполнена в рамках "порогового" уровня. Основная часть задания выполнена верно.

III. Правила выставления итоговой оценки по курсу

Оценка определяется в соответствии с Положением о балльно-рейтинговой системе для студентов НИУ «МЭИ» на основании семестровой и аттестационной составляющих.