

Министерство образования и науки Российской Федерации
федеральное государственное бюджетное образовательное
учреждение высшего образования
«Казанский национальный исследовательский технический университет
им. А.Н. Туполева-КАИ»
(КНИТУ-КАИ)

Институт Компьютерных технологий и защиты информации

Кафедра Компьютерных систем

УТВЕРЖДАЮ

Ответственный за ОП

Верш И.С.Вершинин

«31» 08 2017 г.

Регистрационный № 4010-17/4-
137

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

для проведения промежуточной аттестации обучающихся по дисциплине

«Технология интеллектуального анализа данных»

Индекс по учебному плану: Б1.В.ДВ.04.02

Направление: 09.04.01 «Информатика и вычислительная техника»

Квалификация: магистр

Программа: «Высокопроизводимые вычислительные системы»

Виды профессиональной деятельности: научно-исследовательская

Заведующий кафедрой С.С.Зайдуллин

Разработчик Л.Ю.Емалетдинова

Казань 2017 г.

Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине Технология интеллектуального анализа данных
(наименование дисциплины, практики)

Содержание фонда оценочных средств (ФОС) соответствует требованиям федерального государственного стандарта высшего образования (ФГОС ВО) по магистратуре «Информатика и вычислительная техника» учебному плану направления подготовки : 09.04.01 «Информатика и вычислительная техника»,

Дается оценка:

- полноты и актуальности ФОС;
- соответствия ФОС задачам будущей профессиональной деятельности обучающихся, установленных ФГОС ВО;
- наличие оценочных средств для проведения различных форм контроля;
- разнообразию форм заданий, наличие контекстных заданий, заданий различного уровня трудности (сложности), вариантности.

Оценивается уровень приближенности ФОС к задачам будущей профессиональной деятельности обучающихся.

Закключение. Учебно-методическая комиссия делает вывод о том, что представленные материалы соответствуют требованиям ФГОС ВО по направлению подготовки : 09.04.01 «Информатика и вычислительная техника» и рекомендуются для использования в учебном процессе.

Рассмотрено на заседании учебно-методической комиссии института КТЗИ от 31 августа 2017 г., протокол №.8

Председатель УМК института КТЗИ  В.В. Родионов

Содержание

ВВЕДЕНИЕ	4
1. ФОРМЫ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ	5
2. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ	5
3. ПЕРЕЧЕНЬ КОМПЕТЕНЦИЙ С УКАЗАНИЕМ ЭТАПОВ ИХ ФОРМИРОВАНИЯ В ПРОЦЕССЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ	5
4. ОПИСАНИЕ ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ ОЦЕНИВАНИЯ КОМПЕТЕНЦИЙ НА РАЗЛИЧНЫХ ЭТАПАХ ИХ ФОРМИРОВАНИЯ, ОПИСАНИЯ ШКАЛЫ ОЦЕНИВАНИЯ	6
5. МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ, ОПРЕДЕЛЯЮЩИЕ ПРОЦЕДУРУ ОЦЕНИВАНИЯ ЗНАНИЙ, УМЕНИЙ, НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ, ХАРАКТЕРИЗУЮЩИХ ЭТАПЫ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ	8
6. КОНТРОЛЬНЫЕ ЗАДАНИЯ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ, НАВЫКОВ, ХАРАКТЕРИЗУЮЩИХ ЭТАПЫ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ В ПРОЦЕССЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ	9
ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ И ДОПОЛНЕНИЙ	36

Введение

Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине (ФОС ПА) **Технология интеллектуального анализа данных** – это комплект методических и контрольно-измерительных материалов, предназначенных для определения уровня сформированности компетенций, оценивания знаний, умений, владений на разных этапах освоения дисциплины для проведения промежуточной аттестации обучающихся по дисциплине.

ФОС ПА является составной частью учебного и методического обеспечения программы магистратуры по направлению подготовки **Информатика и вычислительная техника**

Задачи ФОС по дисциплине **Технология интеллектуального анализа данных**

- оценка запланированных результатов освоения дисциплины обучающимися в процессе изучения, в соответствии с разработанными и принятыми критериями по каждому виду контроля;
- контроль и управление процессом приобретения обучающимися необходимых знаний, умений, навыков и формирования компетенций, определенных в ФГОС ВО по направлению подготовки

ФОС ПА по дисциплине **Технология интеллектуального анализа данных** сформирован на основе следующих основных принципов оценивания:

- пригодности (валидности) (объекты оценки соответствуют поставленным целям обучения);
- надежности (использования единообразных стандартов и критериев для оценивания запланированных результатов);
- эффективности (соответствия результатов деятельности поставленным задачам).

ФОС ПА по дисциплине **Технология интеллектуального анализа данных** разработан в соответствии с требованиями ФГОС ВО по направлению подготовки : **09.04.01 «Информатика и вычислительная техника»** для аттестации обучающихся на соответствие их персональных достижений требованиям поэтапного формирования соответствующих составляющих компетенций и включает контрольные вопросы (или тесты) и типовые задания, необходимые для оценки знаний, умений и навыков, характеризующих этапы формирования компетенций.

1. Формы промежуточной аттестации по дисциплине

Дисциплина Технологии интеллектуального анализа данных изучается во 2 семестре при очной форме обучения и завершается промежуточной аттестацией в форме экзамена.

2. Оценочные средства для промежуточной аттестации

Оценочные средства для промежуточной аттестации по дисциплине Технологии интеллектуального анализа данных при очной форме обучения.

Таблица 1

Оценочные средств для промежуточной аттестации
(очная форма обучения)

№ п/п	Семестр	Форма промежуточной аттестации	Оценочные средства
1.	2 / 4	экзамен	ФОС ПА

3. Перечень компетенций с указанием этапов их формирования в процессе освоения дисциплины

Перечень компетенций и их составляющих, которые должны быть сформированы при изучении темы соответствующего раздела дисциплины Технологии интеллектуального анализа данных, представлен в таблице 2.

Таблица 2

Перечень компетенций и этапы их формирования
в процессе освоения дисциплины

№ п/п	Этап формирования (семестр)	Наименование раздела	Код формируемой компетенции (составляющей компетенции)		Форма промежуточной аттестации
1.	2/4	<i>Раздел 1. Технологии и методы интеллектуального анализа данных</i> <i>Раздел 2. Вспомогательные методы интеллектуального анализа данных</i>	ПК-4	ПК-4.3, ПК-4.У, ПК-4.В	экзамен

4. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описания шкалы оценивания

Показатели и критерии оценивания сформированности компетенций на зачете / экзамене, приведены в таблице 3.

Таблица 3

Показатели и критерии оценивания сформированности компетенций на экзамене

№ п/п	Этап формирования (семестр)	Код формируемой компетенции (составляющей компетенции)		Критерии оценивания	Показатели оценивания (планируемые результаты обучения)		
					Пороговый уровень	Продвину-тый уровень	Превосход-ный уро-вень
1.	2/4	ПК-4	ПК-4.3, ПК-4.У	Теоретиче-ские навыки	Знать мате-матические модели за-дач, приме-няемых в технологиях интеллекту-ального анализа данных	Знать мате-матические модели за-дач, приме-няемых в технологиях интеллекту-ального анализа данных и основные методы их решения	Знать мате-матические модели за-дач, приме-няемых в технологиях интеллекту-ального анализа данных, различные подходы к их решению и соответ-ствующие методы их решения

2.	2/4	ПК-4	ПК-4.У, ПК-4.В	Практические навыки (опыт практической деятельности)	Применять современные методы и средства интеллектуального анализа данных для решения практических	Разрабатывать и применять современные методы и средства технологии интеллектуального анализа данных для решения практических	Разрабатывать и применять современные методы и средств атехнологии интеллектуального анализа данных для решения практических и научных задач
----	-----	------	-------------------	--	---	--	--

Формирование оценки при промежуточной аттестации по итогам освоения дисциплины зависит от уровня освоения компетенций, которые обучающийся должен освоить по данной дисциплине. Связь между итоговой оценкой и уровнем освоения компетенций (шкала оценивания) представлена в таблице 5.

Таблица 5

Описание шкалы оценивания

Шкала оценивания		Описание оценки в требованиях к уровню и объему компетенций
Словесное выражение	Выражение в баллах	
Отлично (Зачтено)	от 86 до 100	Освоен превосходный уровень всех компетенций (составляющих компетенций)
Хорошо (Зачтено)	от 71 до 85	Освоен продвинутый уровень всех компетенций (составляющих компетенций)
Удовлетворительно (Зачтено)	от 51 до 70	Освоен пороговый уровень всех компетенций (составляющих компетенций)
Неудовлетворительно (Не зачтено)	до 51	Не освоен пороговый уровень всех компетенций (составляющих компетенций)

5. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Формирование оценки по результатам текущего контроля успеваемости и промежуточной аттестации по итогам освоения дисциплины **Технологии интеллектуального анализа данных** приведено в таблице 6.

Таблица 6

Формирование оценки по итогам освоения дисциплины

Наименование контрольного мероприятия	Рейтинговые показатели				
	I аттестация	II аттестация	III аттестация	по результатам текущего кон- троля	по итогам промежуточной аттестации (зачета /экзамена)
<i>Раздел 1. Технологии и методы интеллектуального анализа данных</i>	16			16	
Тест текущего контроля по разделу	10			10	
Выполнение лабораторных работ	6			6	
<i>Раздел 2. Вспомогательные методы интеллектуального анализа данных</i>		16		16	
Тест текущего контроля по разделу		10		10	
Выполнение лабораторных работ		6		6	
Промежуточная аттестация (экзамен):					52
– тест промежуточной аттестации по дисциплине					22
– в письменной форме по билетам					30

6 Контрольные задания, необходимые для оценки знаний, умений, навыков, характеризующих этапы формирования компетенций в процессе освоения дисциплины

ФОСТК-1

Вопросы по этапам и средствам технологии данных(тема 1.1.)

1. Технология Knowledge Discovery in Databases (KDD). Схема этапов и суть каждого этапа. Пример.
2. Определение Data Mining. Четыре базовых класса задач, их обобщенные формулировки, примеры.
3. Критерии выбора алгоритма.
4. Программное обеспечение в области анализа данных.

Вопросы по классификации данных (тема 1.2.)

5. Этапы процесса классификации данных.
6. Построение логистической регрессии для случая бинарной выходной переменной.
7. Значимость входной переменной.
8. Шанс и отношение шансов (для случая, когда входная переменная принимает только два значения).
9. Простой байесовский критерий. Пример.
10. Структура дерева решений. Пример.
11. Алгоритм построения дерева решений ID3. Недостатки.
12. Построение правил классификации на основе дерева решений.
13. Сложность и точность деревьев решений.
14. Определение нейронной сети. Основные свойства. Недостатки.
15. Модель искусственного нейрона. Примеры активационных функций нейрона

16. Архитектура нейронной сети. Пример архитектуры.
17. Рекомендации по выбору числа нейронов в многослойном персептроне.
Пример расчета.
18. Процесс обучения сети
19. Алгоритм обратного распространения ошибки.
20. Схема алгоритма обратного распространения ошибки для нейронной сети с одним скрытым слоем

Вопросы по кластеризации данных (тема 1.3)

21. Формальная постановка задачи кластеризации.
22. Идея алгоритма k-средних.
23. Алгоритм k-средних.
24. Достоинства и недостатки алгоритма
25. Пример применения алгоритма k-средних
26. Виды иерархических алгоритмов.
27. Агломеративный алгоритм. Недостатки алгоритма.
28. Пример применения агломеративного алгоритма.
29. Критерии принятия решения о составе кластеров на основании использования дендрограммы.
30. Сеть Кохонена: архитектура сети (количество нейронов входного и выходного слоев). Пример.
31. Общие процедуры процесса обучения сети Кохонена.
32. Алгоритм обучения сети Кохонена.
33. Пример работы сети Кохонена
34. Проблемы алгоритмов кластеризации

Вопросы по линейной регрессии (тема 1.4)

35. Определение простой линейной регрессии. Уравнение. Метод вычисления коэффициентов регрессии. Стандартная ошибка оценивания. Коэффициент детерминации. Общая сумма Q , регрессионная Q_R и остаточная Q_E :
36. Линейная регрессионная модель. Проверка гипотез об отсутствии или присутствии линейной связи между переменными с помощью t – критерия, построенного на основе t – распределения Стьюдента.
37. Линейная регрессионная модель. Проверка гипотез об отсутствии или присутствии линейной связи между переменными с помощью F – критерия, использующего F – распределение Фишера.
38. Уравнение множественной регрессии. Регрессионная модель. Оценка значимости переменных, входящих в регрессионную модель на основе t – критерия Стьюдента.
39. Уравнение множественной регрессии. Регрессионная модель. Оценка наличия линейной зависимости между набором входных переменных $x_1, x_2 \dots x_m$ и переменной y на основе F критерия Фишера.
40. Метод прямого отбора переменных в регрессионной модели. Пример.
41. Проверка качества регрессионной модели на основе теста Чоу.

Вопросы по ассоциативным правилам (тема 1.5)

42. Определение и вид ассоциативных правил. Области приложения ассоциативных правил. Понятия транзакции и предметного набора. Понятие анализа рыночной корзины.
43. Определение показателей ассоциативных правил: Поддержка и Достоверность, Лифт, Левередж, Улучшение. Формулы для их вычисления. Назначение этих показателей. Примеры.

44. Этап 1 алгоритма поиска ассоциативных правил Apriori: алгоритм построения частых наборов на множестве транзакций. Пример.
45. Этап 2 алгоритма поиска ассоциативных правил Apriori: генерация ассоциативных правил. Пример.
46. Понятие иерархических ассоциативных правил. Иерархия товаров. Способ использования иерархии при построении правил.
47. Методы поиска иерархических правил.

ФОС ПА

Экзаменационные вопросы

1. Определение Data Mining. Четыре базовых класса задач, их обобщенные формулировки, примеры.
2. Определение простой линейной регрессии. Уравнение. Метод вычисления коэффициентов регрессии. Стандартная ошибка оценивания. Коэффициент детерминации. Общая сумма Q , регрессионная Q_R и остаточная Q_E
3. Построение логистической регрессии для случая бинарной выходной переменной. Значимость входной переменной.
4. Линейная регрессионная модель. Проверка гипотез об отсутствии или присутствии линейной связи между переменными с помощью t – критерия, построенного на основе t – распределения Стьюдента
5. Построение правил классификации на основе дерева решений. Сложность и точность деревьев решений.
6. Линейная регрессионная модель. Проверка гипотез об отсутствии или присутствии линейной связи между переменными с помощью F – критерия, использующего F – распределение Фишера
7. Простой байесовский критерий. Пример.
8. Уравнение множественной регрессии. Регрессионная модель. Оценка значимости переменных, входящих в регрессионную модель на основе t – критерия Стьюдента.
9. Алгоритм построения дерева решений ID3. Недостатки.

10. Уравнение множественной регрессии. Регрессионная модель. Оценка значимости переменных, входящих в регрессионную модель на основе t – критерия Стьюдента.
11. Архитектура нейронной сети. Пример архитектуры. Рекомендации по выбору числа нейронов в многослойном персептроне.
12. Уравнение множественной регрессии. Регрессионная модель. Оценка наличия линейной зависимости между набором входных переменных $x_1, x_2, \dots, x_m$ и переменной y на основе F критерия Фишера.
13. Алгоритм обратного распространения ошибки.
14. Метод прямого отбора переменных в регрессионной модели. Пример
15. Формальная постановка задачи кластеризации. Алгоритм k -средних.
16. Проверка качества регрессионной модели на основе теста Чоу
17. Агломеративный алгоритм. Критерии принятия решения о составе кластеров на основании использования дендрограммы.
18. Определение и вид ассоциативных правил. Области приложения ассоциативных правил. Понятия транзакции и предметного набора. Понятие анализа рыночной корзины.
19. Сеть Кохонена: архитектура сети (количество нейронов входного и выходного слоев). Общие процедуры процесса обучения сети Кохонена. Алгоритм обучения сети Кохонена.
20. Определение показателей ассоциативных правил: Поддержка и Достоверность, Лифт, Левередж, Улучшение. Формулы для их вычисления. Назначение этих показателей. Примеры.

Тесты

Вариант 1

№	Вопрос	Ответ
1	Укажите последовательность этапов технологии извлечения знаний из баз данных	1. Выборка данных, трансформация, очистка, добыча данных, интерпретация. 2. Выборка данных, очистка, добыча данных, трансформация, интерпретация.

	(Knowledge Discovery in Databases)	3. Выборка данных, чистка, трансформация, добыча данных, интерпретация.
2	Логит-преобразование имеет следующий вид:	1. $\lg\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \beta_n x_n$ 2. $\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \beta_n x_n$ 3. $p = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \beta_n x_n$
3	Если вероятность того, что событие произойдет, равно p , то шанс события (для случая, когда входная переменная принимает только два значения) вычисляется как	1. $\frac{p}{1-p}$ 2. $\frac{1-p}{p}$ 3. $\frac{p^2}{1-p^2}$
4	Иерархические древовидные структуры позволяют решать задачу	1. кластеризации 2. построения регрессии 3. классификации
5	В алгоритма ID3 энтропия $Info(T)$ определенного узла вычисляется как:	1. $-\bigcup_{j=1}^n P_j \log_2 P_j$, 2. $-\prod_{j=1}^n P_j \log_2 P_j$, 3. $-\sum_{j=1}^n P_j \log_2 P_j$, где P_j – вероятность появления примеров, относящихся к j-му классу
6	Архитектура нейронной сети определяется.	1. количеством нейронов и способом их соединения внутри сети 2. количеством нейронов, их активационной функцией, количеством слоев и способом их соединения внутри сети 3. количеством слоев и способом их соединения внутри сети
7	Формулы вычисления весов скрытых нейронов n -го слоя, где $x_i^{(n)}$ – вход нейрона n -го слоя (выход $y_i^{(n-1)}$)	1. $\delta_j^{hid(n)} = y_j^{hid(n)}(1 - y_j^{hid(n)}) \sum_k \delta_k^{hid(n+1)} w_{jk}^{(n+1)}$, $w_{ij}^{hid(n)}(t+1) = w_{ij}^{hid(n)}(t) - \eta \delta_j^{hid(n)} x_i^{(n)}$ 2. $\delta_j^{hid(n)} = y_j^{hid(n)}(1 + y_j^{hid(n)}) \sum_k \delta_k^{hid(n+1)} w_{jk}^{(n+1)}$, $w_{ij}^{hid(n)}(t+1) = w_{ij}^{hid(n)}(t) - \eta \delta_j^{hid(n)} x_i^{(n)}$ 3. $\delta_j^{hid(n)} = y_j^{hid(n)}(1 - y_j^{hid(n)}) \sum_k \delta_k^{hid(n+1)} w_{jk}^{(n+1)}$,

		$w_{ij}^{hid(n)}(t+1) = w_{ij}^{hid(n)}(t) + \eta \delta_j^{hid(n)} x_i^{(n)}$
8	Задача кластеризации состоит в построении множества $C = \{c_1, c_2 \dots c_k \dots c_g\}$, где c_k - кластер, содержащий объекты из множества I , $d(j, p)$ - мера близости между объектами, называемая расстоянием.	1. $c_k = \{j, p j \in I, p \in I \text{ и } d(j, p) = \sigma\}$ 2. $c_k = \{j, p j \in I, p \in I \text{ и } d(j, p) > \sigma\}$ 3. $c_k = \{j, p j \in I, p \in I \text{ и } d(j, p) < \sigma\}$
9	Линия регрессии – это прямая наилучшего приближения для множества пар значений входной и выходной переменной (x, y) , выбираемая таким образом, чтобы	1. $\sum_{i=1}^n (\hat{y}_i - y_i)^2 \rightarrow \max$ 2. $\sum_{i=1}^n (\hat{y}_i - y_i)^2 \rightarrow \min$ 3. $\sum_{i=1}^n (\hat{y}_i - y_i)^2 \rightarrow \sup$
10	Простая линейная регрессионная модель задается следующим образом: $y = \beta_0 + \beta_1 x + \varepsilon$ Где β_0, β_1 – коэффициенты модели. Предполагается, что ошибка ε	1. является нормально распределенной случайной величиной с математическим ожиданием $m(\varepsilon) = 0$ и постоянной дисперсией σ^2. 2. является равномерно распределенной случайной величиной 3. является нормально распределенной случайной величиной с ненулевым математическим ожиданием $m(\varepsilon)$ и постоянной дисперсией σ^2.
11	Для проверки наличия связи между выходной переменной y и некоторой входной переменной x_i проверяется следующая гипотеза:	1. $H_0: \beta_i = 0$ (остальные $\beta_k \neq 0, k \neq i$) 2. $H_0: \beta_i \neq 0$ (остальные $\beta_k \neq 0, k \neq i$)
12	Анализ рыночной корзины - это	1. Экономический метод 2. Метод исследования взаимной связи между событиями 3. Определение состава продуктов корзины

Вариант 2

№	Вопрос	Ответ
1	Дайте определение понятию Data Mining (добыча данных)	1. Обнаружение в данных ранее неизвестных, тривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. 2. Обнаружение в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. 3. Обнаружение в данных ранее известных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.
2	Логистическая функция имеет вид:	1. $\rho(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$ 2. $\frac{1 - \rho(x)}{\rho(x)} = e^{\beta_0 + \beta_1 x} / (1 + e^{\beta_0 + \beta_1 x})$ 3. $\ln\left(\frac{1 - \rho(x)}{\rho(x)}\right) = e^{\beta_0 + \beta_1 x} / (1 + e^{\beta_0 + \beta_1 x})$
3	Отношение шансов события для случая, когда входная переменная логистической регрессии принимает только два значения (0 и 1) определяется как $OR = \frac{p(1)/(1-p(1))}{p(0)/(1-p(0))}$. Тогда коэффициент β_1 можно рассматривать как	1. десятичный логарифм отношения шансов при увеличении на единицу значения входной переменной 2. логарифм по основанию 2 отношения шансов при увеличении на единицу значения входной переменной 3. натуральный логарифм отношения шансов при увеличении на единицу значения входной переменной
4	Узлы дерева решений содержат правила, с помощью которых	1. производится проверка атрибутов и множества объектов в данном узле разбивается на подмножества 2. производится проверка атрибутов 3. множества объектов в данном узле разбивается на подмножества
5	В алгоритме ID3 общая энтропия узлов $Info_S(T)$ после разбиения S вычисляется	1. $\prod_{i=1}^k \frac{N_i}{N} Info(T_i)$ 2. $\sum_{i=1}^k \frac{N_i}{N} Info(T_i)$

	как:	3. $\bigcup_{i=1}^k \frac{N_i}{N} \text{Info}(T_i)$
6	Число нейронов входного слоя персептрона определяется	1. числом выходных переменных моделей 2. числом скрытых слоев 3. числом входных переменных модели
7	Какому из условий не удовлетворяет функция $d(j, p)$, являющаяся расстоянием между элементами j и p	1. $d(j, p) \geq 0$ для всех j и p 2. $d(j, p) = 0$ тогда и только тогда, когда объекты с номерами j, p одинаковы 3. $d(j, p) = d(p, j)$ 4. $d(j, p) > d(j, r) + d(r, p)$
8	В иерархических алгоритмах число кластеров	1. определяется в результате работы алгоритма 2. задано определенным числом 3. задается случайно
9	Уравнение регрессии имеет вид:	1. $\hat{y} = b_0 * b_1 x$ 2. $\hat{y} = b_0 + b_1 x$ 3. $\hat{y} = b_0 / b_1 x$
10	Для утверждения того, что построенная регрессионная модель действительно описывает линейную связь между переменными для всей генеральной совокупности необходимо принять в результате статистической проверки следующую гипотезу:	1. $H_0: \beta_1 = 0$ 2. $H_1: \beta_1 \neq 0$
11	Для оценки наличия линейной зависимости между набором входных переменных $x_1, x_2 \dots x_m$ и переменной y проверяется следующая гипотеза:	1. $H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_m = 0$ 2. H_0 : хотя бы один коэффициент $\beta_i \neq 0$
12	Ассоциативные правила – это	1. правила для объединения двух или более событий 2. правила для формального описания

		взаимной связи между двумя или более событиями 3. правила для разъединения множества событий на подмножества
--	--	--

Вариант 3

№	Вопрос	Ответ
1	Укажите четыре базовых класса задач	1. Классификация, линейная регрессия, кластеризация, построение нечетких моделей 2. Построение нейронных сетей, регрессия, построение нечетких моделей, ассоциация 3. Классификация, линейная регрессия, кластеризация, ассоциация
2	Для выборки независимых наблюдений функция правдоподобия $l(\beta/x)$ является функцией вектора параметров β и вычислется:	1. $l(\beta/x) = \prod_{i=1}^n [\rho(x_i)]^{y_i} [1 - \rho(x_i)]^{1-y_i}$ 2. $l(\beta/x) = \sum_{i=1}^n [\rho(x_i)]^{y_i} [1 - \rho(x_i)]^{1-y_i}$ 3. $l(\beta/x) = \bigcup_{i=1}^n [\rho(x_i)]^{y_i} [1 - \rho(x_i)]^{1-y_i}$
3	Байесовский подход представляет собой группу алгоритмов	1. кластеризации, основанных на принципе максимума апостериорной вероятности. 2. классификации, основанных на принципе максимума апостериорной вероятности. 3. ассоциации, основанных на принципе максимума апостериорной вероятности.
4	Листья дерева решений – это конечные узлы дерева, в которых содержатся	1. подмножества объектов, ассоциированные с соответствующими классами 2. наиболее характерные атрибуты объектов 3. атрибуты объектов, встречающиеся с наибольшей частотой
5	В алгоритме ID3 этапы для определения оптимального порога следующие: упорядочим значения атрибутов $v_1, v_2 \dots v_m$. $i = 1$ В качестве порога выбирается нижняя граница v_i интервала (v_i, v_{i+1}) или его	1. $i^* = \underset{i}{\operatorname{argmin}} \operatorname{Gain} S^i$ 2. $i^* = \underset{i}{\operatorname{argmax}} \operatorname{Gain} S^i$

	среднее значение Используя выбранный порог, осуществляется разбиение S_i и вычисляется $Gain(S_i)$. $i = i + 1$. Если $i \leq m - 1$, то переходят к пункту 3. Выбираем разбиение S^{i*}, где	
6	Число нейронов в скрытых слоях и число скрытых слоев выбирают таким образом, чтобы	1. количество образованных ими связей было равно числу обучающих примеров 2. количество образованных ими связей было больше числа обучающих примеров как минимум в два раза 3. количество образованных ими связей было меньше числа обучающих примеров как минимум в два раза
7	Евклидово расстояние между элементами j и p определяется по следующей формуле:	1. $d(j, p) = \sqrt{\sum_{k=1}^m (x_{jk} - x_{pk})^2}$ 2. $d(j, p) = \sum_{k=1}^m x_{jk} - x_{pk}$ 3. $d(j, p) = \sqrt{\sum_{k=1}^m x_{jk} - x_{pk} }$
8	Агломеративные алгоритмы кластеризации характеризуются	1. последовательным разбиением исходного множества элементов и соответствующим увеличением числа кластеров 2. последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров
9	Ограничение линейной регрессии: линия регрессии является подходящей аппроксимацией некоторой функции	1. в том диапазоне изменений входной переменной x, в котором распределены исходные наблюдения. 2. за пределами распределения исходных наблюдений. 3. не только в котором распределены исходные наблюдения, но и за пределами
10	Для проверки истинности гипотезы H_0 об отсутствии линейной зависимости между входной и выходной переменной используют t – критерий, построенный на основе t – распределения	1. $t = \frac{S_R^2}{S_E^2}$, где $S_R^2 = \frac{Q_R}{m} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{1}$; $S_E^2 = \frac{Q_E}{n-m-1} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}$ 2. $t = \frac{b_1}{S_{b_1}}$, $S_{b_1} = \frac{E_{ст}}{\sqrt{\sum_{i=1}^n x_i^2 - \frac{(\sum x_i)^2}{n}}} = \frac{E_{ст}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$, $E_{ст} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n-m-1}}$, $\hat{y}_i = b_0 + b_1 x_i$

	Стьюдента с $(n - 2)$ степенями свободы:	
11	Какой из методов не является методом отбора входных переменных в регрессионные модели?	1. Метод прямого отбора. 2. Метод обратного исключения. 3. Метод обратного отбора. 4. Метод лучших подмножеств
12	Транзакция – это	1. Множество событий, происходящих совместно. 2. Множество финансовых операций. 3. Регрессионная зависимость

Вариант 4

№	Вопрос	Ответ
1	Какие критерии используются при выборе алгоритмов, если данные зашумлены	1. Производительность, точность 2. Производительность, устойчивость, точность 3. Точность, устойчивость
2	Для выборки независимых наблюдений логарифмическая функция правдоподобия $L(\beta/x)$ является функцией вектора параметров β и вычислится:	1. $L(\beta/x) = \prod_{i=1}^n (y_i \ln[\rho(x_i)] + (1 - y_i) \ln[1 - \rho(x_i)])$ 2. $L(\beta/x) = \sum_{i=1}^n (y_i \ln[\rho(x_i)] + (1 - y_i) \ln[1 - \rho(x_i)])$ 3. $L(\beta/x) = \cup_{i=1}^n (y_i \ln[\rho(x_i)] + (1 - y_i) \ln[1 - \rho(x_i)])$
3	Простой байесовский критерий вычисляет	1. эмпирическую вероятность принадлежности объекта к каждому из классов, затем выбирается тот класс, для которого она максимальна. 2. априорную вероятность принадлежности объекта к каждому из классов, затем выбирается тот класс, для которого она максимальна. 3. апостериорную вероятность принадлежности объекта к каждому из классов, затем выбирается тот класс, для которого она максимальна.
4	Мерой оценки возможного разбиения исходного множества объектов при постро-	1. чистота, под которой понимается присутствие в получаемых подмножествах объектов только одного класса 2. максимальная частота появления объектов раз-

	ении дерева решений является	ных классов 3. минимальная частота появления объектов разных классов
5	Что не является условием метода ранней остановки при построении дерева решений	1. минимальное количество примеров в узле 2. глубина дерева 3. ошибка классификации дерева 4. количество узлов
6	Какой из условий не является условием останова обучения нейронной сети	1. ошибка E будет минимальной 2. ошибка E больше некоторого порога ε 3. число итераций достигнет некоторого заданного числа
7	Расстояние Манхэттена между элементами j и p определяется по следующей формуле:	4. $d(j, p) = \sqrt{\sum_{k=1}^m (x_{jk} - x_{pk})^2}$ 5. $d(j, p) = \sum_{k=1}^m x_{jk} - x_{pk} $ 6. $d(j, p) = \sqrt{\sum_{k=1}^m x_{jk} - x_{pk} }$
8	Дивизимный алгоритмы кластеризации характеризуются	3. последовательным разбиением исходного множества элементов и соответствующим увеличением числа кластеров 4. последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров
9	Стандартная ошибка оценивания соответствия простой линейной регрессии реальным данным определяется по следующей формуле: где n – число наблюдений в выборке исходных данных; m – количество независимых переменных, для линейной регрессии $m = 1$.	1. $E_{ст} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$ 2. $E_{ст} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n-m}}$ 3. $E_{ст} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n-m-1}}$
10	Для проверки истинности гипотезы H_0 об отсутствии линейной зависимости между входной и выходной переменной используют F –	3. $F = \frac{S_R^2}{S_E^2}$, где $S_R^2 = \frac{Q_R}{m} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{1}$; $S_E^2 = \frac{Q_E}{n-m-1} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}$ 4. $F = \frac{b_1}{S_{b_1}}$, $S_{b_1} = \frac{E_{ст}}{\sqrt{\sum_{i=1}^n x_i^2 - \frac{(\sum x_i)^2}{n}}} = \frac{E_{ст}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$

	критерий Фишера.	$E_{ст} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n-m-1}}, \hat{y}_i = b_0 + b_1 x_i$
11	Для введения переменной в регрессионную модель должна быть отвергнута гипотеза	1. Гипотеза H_1: квадратичная сумма $Q_R^{доп}$, связанная с введенной на этом этапе дополнительной переменной, вносит значительный вклад в квадратичную сумму регрессии для модели, которая уже содержит введенные ранее переменные. 2. Гипотеза H_0: квадратичная сумма $Q_R^{доп}$, связанная с введенной на этом этапе дополнительной переменной, не вносит значительного вклада в квадратичную сумму регрессии для модели, которая уже содержит введенные ранее переменные.
12	Предметный набор - это	1. Непустое множество предметов, появившихся в одной транзакции. 2. Непустое множество предметов 3. Непустое множество событий

Вариант 5

№	Вопрос	Ответ
1	Какие критерии используются при выборе алгоритмов, если объем исходных данных велик и данные не зашумлены	1. Производительность, точность 2. Производительность, устойчивость, точность 3. Точность, устойчивость
2	Оценки вектора коэффициентов β логистической регрессии вычисляются из:	1. условия экстремума их математического ожидания 2. условия экстремума их среднеквадратического отклонения 3. условия экстремума логарифмической функции правдоподобия
3	Для вычисления вероятности $P(H/X)$ используется формула Байеса:	1. $P(H/X) = \frac{P(H)P(X/H)}{P(X)}$ 2. $P(H/X) = \frac{P(X)P(X/H)}{P(H)}$ 3. $P(H/X) = P(H)P(X)P(X/H)$
4	Алгоритм ID3 начинает работу со всеми записями исходных данных (примерами) в корне-	1. Более двух атрибутов 2. Два атрибута 3. Один атрибут

	вом узле дерева. Для разделения множества примеров корневого узла выбирается	
5	Искусственная нейронная сеть представляет собой	1. параллельно-распределенную систему бинарных элементов 2. параллельно-распределенную систему процессорных элементов 3. последовательно-распределенную систему процессорных элементов
6	Для числовых атрибутов обучающей выборки нейронной сети осуществляется минимаксная нормализация, причем	1. входные переменные нормируются в диапазоне $[-1,1]$, а выходные – в $[0,1]$. 2. входные переменные нормируются в диапазоне $[0,1]$, а выходные – в $[-1,1]$. 3. входные переменные нормируются в диапазоне $[0,1]$, а выходные – в $[0,1]$.
7	В алгоритме кластеризации k-средних число кластеров k	1. вычисляется в результате работы алгоритма 2. меняется с каждым шагом 3. задается на первом шаге фиксированным числом
8	Основная цель нейронной сети Кохонена	1. преобразование сложных многомерных данных в более простую структуру малой размерности 2. аппроксимация функций специальными зависимостями 3. выявление числовой закономерности данных
9	Какая из сумм не используется в оценке степени соответствия регрессии реальным данным?	1. $Q = \sum_{i=1}^n (y_i - \bar{y})^2$, 2. $Q_R = \sum_{i=1}^n (\hat{y}_i - y_i)^2$, 3. $Q_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2$.
10	Ошибка первого рода	1. $\alpha = P(H_0 - \text{принимается} / H_0 - \text{не справедлива})$ 2. $\alpha = P(H_0 - \text{отклоняется} / H_0 - \text{не справедлива})$ 3. $\alpha = P(H_0 - \text{отклоняется} / H_0 - \text{справедлива})$
11	Коэффициент детерминации r^2 показывает	1. степень линейной зависимости между соответствующей входной и выходной переменной 2. силу линейной зависимости между соответствующей входной и выходной переменной
12	Поддержка ассоциативного правила – это	1. отношение числа транзакций, которые содержат условие к общему числу транзакций 2. отношение числа транзакций, которые содержат следствие, к общему числу транзакций 3. отношение числа транзакций, которые содержат как условие, так и следствие, к общему числу транзакций

Вариант 6

№	Вопрос	Ответ
1	Суть задачи классификации заключается в:	- установлении зависимости непрерывной выходной переменной от входных переменных и определение на ее основе значения выходной переменной для конкретных значений входных переменных. - установлении зависимости дискретной выходной переменной от входных переменных и определение на ее основе значения выходной переменной из известного конечного множества для конкретных значений входных переменных. - определении значения выходной переменной из известного конечного множества для конкретных значений входных переменных.
2	Значимость входной переменной логистической регрессии определяется на основе значения статистики, если n_1 – число значений $y = 1$; n_0 – число значений $y = 0$; n – общее число значений $n_1 + n_0$	$G = 2\{ \sum_{i=1}^n (y_i \ln[\rho(x_i)] + (1 - y_i) \ln[1 - \rho(x_i)]) \} - (n_1 \ln n_1 + n_0 \ln n_0)$ $G = 2\{ \prod_{i=1}^n (y_i \ln[\rho(x_i)] + (1 - y_i) \ln[1 - \rho(x_i)]) \} - (n_1 \ln n_1 + n_0 \ln n_0)$ $G = 2\{ \sum_{i=1}^n (y_i \ln[\rho(x_i)] + (1 - y_i) \ln[1 - \rho(x_i)]) \} - (n_1 \ln n_1 + n_0 \ln n_0)$
3	В формуле Байеса $P(C_k/X^*) = \frac{P(C_k)P(X^*/C_k)}{P(X^*)}$ для $\forall k = 1, m$	C_k - это класс, к которому нужно отнести объект, X^* - реализация атрибутов для которых класс известен C_k - это класс, к которому нужно отнести объект, X^* - общая реализация атрибутов C_k - это класс, к которому нужно отнести объект, X^* - реализация атрибутов для которых класс неизвестен
4	Алгоритм ID3 работает рекурсивно до тех пор, пока	- в узлах не останутся примеры только одного класса, после чего узлы будут объявлены листьями и разбиение прекратится. - в узлах не останутся примеры разных классов, после чего узлы будут объявлены листьями и разбиение прекратится. - в узлах не останется ни одного примера, после чего узлы будут объявлены листьями и разбиение прекратится.

5	Метод упрощения полного дерева путем отсечения ветвей заключается в том, что	1. после его построения удаляются верхние узлы, правила в которых имеют низкую ценность 2. после его построения удаляются в направлении «снизу вверх» узлы и листья, правила в которых имеют низкую ценность 3. после его построения удаляются только нижние листья, правила в которых имеют низкую ценность
6	Какое из предложений не является требованием к обучающей выборке:	1. выборка должна использовать возможные значения всех переменных 2. значения переменных должны быть распределены равномерно 3. обучающее множество должно содержать только один пример для каждого значения категориального параметра, а также для всех диапазонов значений упорядоченных переменных
7	В алгоритме кластеризации k-средних роль начальных центров k кластеров играют	1. параметры случайно выбранных k объектов 2. усредненные параметры k подмножеств исходного множества объектов 3. параметры первых k объектов
8	Сеть Кохонена состоит из узлов, которые	1. содержат активационные функции 2. осуществляют распределительные действия 3. объединяются в кластеры
9	Если при построении линейной регрессии остаточная сумма $Q_E \ll Q$ (общей суммы), то можно сделать вывод, что	1. качестве оценки целесообразно использовать среднее значение выходной переменной. 2. качестве оценки целесообразно использовать регрессионное значение выходной переменной 3. данных для принятия решений недостаточно
1 0	Ошибка второго рода	4. $\beta = P(H_0 - \text{принимается} / H_0 - \text{не справедлива})$ 5. $\beta = P(H_0 - \text{отклоняется} / H_0 - \text{не справедлива})$ 6. $\beta = P(H_0 - \text{отклоняется} / H_0 - \text{справедлива})$
1 1	Коэффициент b_i линейной регрессии показывает	3. степень линейной зависимости между соответствующей входной и выходной переменной 4. силу линейной зависимости между соответствующей входной и выходной переменной
1 2	Достоверность ассоциативного правила $A \rightarrow B$ определяется как	1. отношение количества транзакций, содержащих следствие B, к количеству транзакций, содержащих только условие A 2. отношение количества транзакций, содержащих условие A, к количеству транзакций, содержащих только условие A 3. отношение количества транзакций, содержащих и условие A, и следствие B, к количеству транзакций, содержащих только условие A

Вариант 7

№	Вопрос	Ответ
1	Задача линейной регрессии заключается в:	1. установлении зависимости дискретной выходной переменной от входных переменных и определение на ее основе значения выходной переменной из известного конечного множества для конкретных значений входных переменных. 2. определении значения выходной переменной из известного конечного множества для конкретных значений входных переменных. 3. установлении зависимости непрерывной выходной переменной от входных переменных и определение на ее основе значения выходной переменной для конкретных значений входных переменных.
2	В случае, когда логистическая регрессия имеет одну входную переменную и гипотеза H_0 о том, что $\beta_1 = 0$ справедлива, то статистика G имеет	1. распределение χ^2 с одной степенью свободы. 2. распределение t-распределение Стьюдента с одной степенью свободы. 3. нормальное распределение с нулевым математическим ожиданием и единичной дисперсией
3	В каком случае имеет место формула $P(X^*/C_k) = P(X_1^*/C_k) P(X_2^*/C_k) \dots P(X_n^*/C_k)$.	1. Если компоненты вектора коррелированы 2. Если компоненты вектора являются векторами 3. Если компоненты вектора независимы
4	Для выбора атрибута алгоритм ID3 использует критерий, называемый	1. увеличением энтропии 2. уменьшением энтропии 3. постоянством энтропии
5	В процессе обучения накопленные знания нейронной сети сосредоточены	1. в весах межэлементных связей. 2. в вычислительных элементах 3. в весах и вычислительных элементах
6	Алгоритм обучения нейронной сети	1. начинает работу с заданного набора весов, затем веса корректируются из условия наибольшего сокращения ошибки значений

		выходных переменных и процедура повторяется. 2. начинает работу со случайного набора весов, затем веса корректируются из условия наибольшего сокращения ошибки значений выходных переменных и процедура повторяется. 3. начинает работу со случайного набора весов, затем веса корректируются из условия наибольшего сокращения весов связей и процедура повторяется.
7	В алгоритме кластеризации k-средних для каждого объекта исходного множества I определяется	1. самый удаленный центр кластера 2. средний центр кластера 3. ближайший к нему центр кластера
8	В процессе обучения и функционирования сеть Кохонена выполняет три процедуры. Укажите их порядок	1. Подстройка весов 2. Объединение 3. Конкуренция
9	Коэффициент детерминации $r^2 = \frac{Q_R}{Q} = \frac{Q_R}{Q_R + Q_E}$ показывает	1. степень приближения линейной зависимости между входной и выходной переменными с реальными данными 2. степень приближения нелинейной зависимости между входной и выходной переменными с реальными данными 3. степень отсутствия зависимости между входной и выходной переменными
10	Какой пункт алгоритма проверки гипотезы с помощью : t – критерия Стьюдента содержит ошибку:	1. Вычислим значение t^* статистики t. 2. Задав α и степень свободы $(n - 2)$, по таблице t – распределения Стьюдента найдем граничное значение t_α. 3. Применим решающее правило: Если $t^* < t_\alpha$, то H_0 – отклоняется. Если $t^* \geq t_\alpha$, то H_0 – принимается.
11	Уравнение регрессии строится	1. на основе генеральной совокупности данных 2. на основе выборки, формируемой

		из имеющегося множества наблюдений и затем зависимость распространяется на все наблюдаемые данные. 3. на основе генеральной совокупности данных и проверяется на тестовой выборке
12	Лифт – это	1. отношение частоты появления следствия B в транзакциях $A \rightarrow B$, которые содержат также и условие A, к частоте появления следствия B в целом 2. отношение частоты появления условия A в транзакциях $A \rightarrow B$, которые содержат также и следствие B, к частоте появления следствия B в целом 3. отношение частоты появления условия A в транзакциях $A \rightarrow B$, которые содержат также и следствие B, к частоте появления условия A в целом

Вариант 8

№	Вопрос	Ответ
1	Задача кластеризации заключается в:	1. определении класса объекта на основе данных, описывающих свойства объектов. 2. группировке объектов (наблюдений, событий) на основе данных, описывающих свойства объектов. Объекты внутри кластера должны быть похожи друг на друга, объекты разных кластеров должны отличаться друг от друга. 3. Отнесении объекта к кластеру содержащих похожие объекты.
2	При проверке значимости входной переменной логистической регрессии по таблицам χ^2 вычисляется вероятность $P(\chi^2 > G^*) = p$, где	1. гипотеза H_0 – отклоняется и рассматриваемая переменная x является значимой при определении вероятности значения y. 2. гипотеза H_0 – отклоняется и рассмат-

	G^* - значение статистики. Если p – мало, то	риваемая переменная x является незначимой при определении вероятности значения y. 3. гипотеза H_0 – принимается и рассматриваемая переменная x является незначимой при определении вероятности значения y.
3	Если атрибут X_i со значением X_i^* является категориальным, то в формуле Байеса	1. $P(X_i^*/C_k) = \frac{S_{ik}}{S}$, где S_{ik} – число наблюдений класса C_k в исходных данных, в которых X_i принимает значение X_i^*, а S – число классов. 2. $P(X_i^*/C_k) = \frac{S_{ik}}{S_k}$, где S_{ik} – число наблюдений класса C_k в исходных данных, в которых X_i принимает значение X_i^*, а S_k – число наблюдений класса C_k. 3. $P(X_i^*/C_k) = \frac{S_k}{S}$, где S_k – число наблюдений класса C_k, а S – число классов.
4	Введенное Шенноном понятие энтропия характеризует	1. степень определенности системы, имеющей n возможных состояний с вероятностями $P(A_1), P(A_2) \dots P(A_n)$. 2. степень неадекватности системы, имеющей n возможных состояний с вероятностями $P(A_1), P(A_2) \dots P(A_n)$. 3. степень неопределенности системы, имеющей n возможных состояний с вероятностями $P(A_1), P(A_2) \dots P(A_n)$.
5	Какое свойство не является свойством нейронных сетей:	1. Обучение на примерах. 2. Линейность. 3. Параллельная обработка данных. 4. Адаптивность. 5. Отказоустойчивость.
6	Алгоритм обратного распространения ошибки –	1. градиентный алгоритм обучения, основанный на минимизации средне квадратичного отклонения выходов сети. 2. алгоритм обучения, основанный на случайном поиске и минимизации средне квадратичного отклонения выходов сети. 3. генетический алгоритм обучения, основанный на минимизации средне квадратичного отклонения выходов сети.

7	В алгоритме кластеризации к средних центр тяжести каждого кластера вычисляется как:	1. усреднение разности параметров по всем объектам, попавшим в кластер 2. усреднение произведения параметров по всем объектам, попавшим в кластер 3. усреднение параметров по всем объектам, попавшим в кластер
8	Правило обучения нейронов, которые являются соседями нейрона-победителя и подстраивают свои веса, заключается в следующем: Где $0 < \beta \leq 1$ – коэффициент скорости обучения, j – номер нейрона, i – номер компоненты объекта, l – номер объекта в исходном множестве объектов.	1. $w_{ij}^{\text{новое}} = w_{ij}^{\text{текущее}} + \beta(x_{li} - w_{ij}^{\text{текущее}})$, 2. $w_{ij}^{\text{новое}} = w_{ij}^{\text{текущее}} - \beta(x_{li} - w_{ij}^{\text{текущее}})$, 3. $w_{ij}^{\text{новое}} = w_{ij}^{\text{текущее}} + \beta(x_{li} + w_{ij}^{\text{текущее}})$,
9	$r^2 \in [0,1]$. Чем ближе r^2 к 1, тем	1. увеличивается степень неопределенности 2. лучше регрессионная модель описывает исходные данные. 3. хуже регрессионная модель описывает исходные данные
10	Какой пункт алгоритма проверки гипотезы с помощью F – критерия Фишера содержит ошибку:	1. Вычислим значение F^* статистики F. 2. Задав α и число степеней свободы $(1, n - 2)$, по таблице F – распределения Фишера найдем граничное значение F_α. 3. Применим решающее правило: Если $F^* < F_\alpha$, то H_0 – отклоняется. Если $F^* \geq F_\alpha$, то H_0 – принимается.
11	Если при построении регрессионной модели множество исходных данных содержит неоднородные данные, то в этом случае необходимо:	1. разделить все множество данных на однородные подмножества и для каждого из них строить свою регрессионную модель 2. построить регрессионную модель на выборке этого исходного множества и оценить ее адекватность с помощью статистического критерия Стьюдента

		3. построить регрессионную модель на выборке этого исходного множества и оценить ее адекватность с помощью статистического критерия Фишера
12	Леввередж –это характеристика ассоциативного правила, которая вычисляется как	1. $T(A \rightarrow B) = P(A)P(B) - P(A \cap B)$ 2. $T(A \rightarrow B) = P(A \cap B) - P(A)P(B)$. 3. $T(A \rightarrow B) = P(A \cap B)/P(A)P(B)$.

Вариант 9

№	Вопрос	Ответ
1	Задача ассоциации заключается в:	1. выявлении общих свойств между связанными событиями. 2. выявлении закономерностей между связанными событиями. 3. описании общих характеристик связанных событий.
2	Согласно методу Вальда для проверки значимости входной переменной логистической регрессии, если справедлива гипотеза H_0: $\beta_1 = 0$, то статистика $Z_W = \frac{b_1}{E_{ст}(b_1)}$ (Здесь b_1 – оценка коэффициента регрессии; $E_{ст}(b_1)$ – стандартная ошибка оценивания коэффициента регрессии на основе наблюдаемых значений: $E_{ст}(b_1) = \sqrt{E_{СКО}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-m-1}},$ где m – число независимых переменных $m = 1$)	4. имеет распределение χ^2 с одной степенью свободы. 5. распределение t -распределение Стьюдента с одной степенью свободы. 6. распределена по нормальному закону с нулевым математическим ожиданием и единичным среднеквадратическим отклонением.
3	Если атрибут X_i со значением X_i^* является непрерывным, тогда предполагается, что его значения подчиняются закону распределе-	1. Стьюдента 2. Фишера 3. Гаусса

	ния	
4	Энтропия $Info(A)$ вычисляется по следующей формуле:	1. $-\sum_{i=1}^n P(A_i) \log P(A_i)$ 2. $-\prod_{i=1}^n P(A_i) \log P(A_i)$ 3. $-\cup_{i=1}^n P(A_i) \log P(A_i)$
5	Модель искусственного нейрона имеет вид	1. $S = \prod_{i=1}^n w_i x_i + b_0$, 2. $S = \sum_{i=1}^n w_i x_i + b_0$, 3. $S = \cup_{i=1}^n w_i x_i + b_0$, где b_0 – смещение, w_i – вес связи, $f(S)$ – активационная функция.
6	Какой этап вычисления с помощью нейронной сети является лишним?	1. Каждый нейрон входного слоя получает возбуждение от вектора его входных значений. 2. Каждый нейрон скрытого слоя производит их взвешенное суммирование и преобразует сумму в его выход с помощью активационной функции. 3. Каждый скрытый нейрон вычисляет свою часть ошибки сети. 4. Нейрон выходного слоя вычисляет значение выходной переменной.
7	Алгоритм кластеризации k средних завершает свою работу, если:	1. расположение центров не изменяется 2. расположение центров не изменяется, а границы кластеров изменяются 3. расположение центров меняется, а границы кластеров остаются неизменными
8	В сети Кохонена нейрон j^* является победителем, для которого выполняется следующее действие:	1. $D(W_{j^*}, X_l) = \max_{j=\overline{1,N}} D(W_j, X_l)$ 2. $D(W_{j^*}, X_l) = \sup_{j=\overline{1,N}} D(W_j, X_l)$ 3. $D(W_{j^*}, X_l) = \min_{j=\overline{1,N}} D(W_j, X_l)$
9	Значение коэффициента корреляции $r \in [-1,1]$. Если r близок к -1 , то	1. с ростом x значение y будет уменьшаться. 2. с ростом x значение y будет увеличиваться. 3. зависимость между переменными носит неопределенный характер
10	В уравнении множественной регрессии: $y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_m x_m$, коэффициенты $b_0, b_1, b_2 \dots b_m$	1. метода средних 2. метода наименьших квадратов 3. метода квадратичного отклонения

	определяются на основе	
11	Статистический метод, называемый критерием Чоу, используется для разрешения проблемы:	1. включения переменной в регрессионную модель 2. является ли значимым улучшение модели после разделения исходной совокупности наблюдений на однородные группы и построения отдельных моделей для каждой группы. 3. исключения переменной из регрессионной модели
12	Улучшением называется характеристика ассоциативного правила, которая вычисляется как	1. $I(A \rightarrow B) = \frac{P(A \cap B)}{P(A/B)P(B)}$. 2. $I(A \rightarrow B) = \frac{P(A \cap B)}{(A/B)}$. 3. $I(A \rightarrow B) = \frac{P(A \cap B)}{P(A)P(B)}$.

Вариант 10

№	Вопрос	Ответ
1	Для чего используется логистическая регрессия?	1. Служит для оценки вероятности того, что зависимая переменная примет заданное значение. 2. Служит для описания зависимости выходной переменной от множества входных переменных 3. Служит для описания зависимости вероятности того, что зависимая переменная примет заданное значение от множества входных переменных.
2	При проверке значимости входной переменной логистической регрессии методом Вальда по таблицам нормального закона распределения вычисляется вероятность $P(Z_W > Z_W^*) = p$, где Z_W^* - значение статистики. Если p – мало, то	4. гипотеза H_0 – отклоняется и рассматриваемая переменная x является незначимой при определении вероятности значения y. 5. гипотеза H_0 – отклоняется и рассматриваемая переменная x является значимой при определении вероятности значения y. 6. гипотеза H_0 – принимается и рассматриваемая переменная x является незначимой при определении вероятности значения y.

		сти значения y .
3	В методе Байеса вероятность $P(X^*)$ вычисляется следующим образом:	1. $P(X^*) = \prod_{k=1}^m (P(C_k)P(X^*/C_k))$ 2. $P(X^*) = \sum_{k=1}^m (P(C_k)P(X^*/C_k))$ 3. $P(X^*) = \bigcup_{k=1}^m (P(C_k)P(X^*/C_k))$
4	В Алгоритме ID3 мера прироста информации вычисляется: $Gain(S) = Info(T) - Info_S(T)$ где	1. $Info(T)$ – энтропия множества исходных примеров после разбиения, 2. $Info(T)$ – суммарная энтропия множества исходных примеров до и после разбиения 3. $Info(T)$ – энтропия множества исходных примеров до разбиения
5	Логистическая активационная функция нейрона имеет вид:	1. $y = \frac{1}{1+e^{-as}}$ 2. $y = \frac{e^{as}-e^{-as}}{e^{as}+e^{-as}}$ 3. $= \begin{cases} 1, S \geq 0 \\ 0, S < 0 \end{cases}$
6	Формулы вычисления весов для выходных нейронов методом обратного распространения ошибок	1. $\delta_j^{out} = y_j^{out}(1 + y_j^{out})(d_j - y_j^{out})$, $w_{ij}^{out}(t+1) = w_{ij}^{out}(t) - \eta \delta_j^{out} x_i$ 2. $\delta_j^{out} = y_j^{out}(1 - y_j^{out})(d_j + y_j^{out})$, $w_{ij}^{out}(t+1) = w_{ij}^{out}(t) - \eta \delta_j^{out} x_i$ 3. $\delta_j^{out} = y_j^{out}(1 - y_j^{out})(d_j - y_j^{out})$, $w_{ij}^{out}(t+1) = w_{ij}^{out}(t) - \eta \delta_j^{out} x_i$
7	Что является недостатком алгоритма кластеризации k средних?	1. умеренные вычислительные затраты; 2. результат не зависит от порядка следования записей, а определяется только набором исходных точек; 3. отсутствие критерия выбора числа кластеров; 4. отсутствие необходимости задания величины близости объектов σ
8	В сети Кохонена производится подстройка весов нейронов, которые расположены относительно нейрона-победителя в пределах радиуса обучения R. Как определяется это множество нейронов? Замечание. N – число нейронов. Нейрон j^* является победителем. W_j-	1. $J = \{j \in \{1, N\} : D(W_j, X_i) - D(W_{j^*}, X_i) \leq R\}$ 2. $J = \{j \in \{1, N\} : D(W_j, X_i) - D(W_{j^*}, X_i) > R\}$ 3. $J = \{j \in \{1, N\} : D(W_j, X_i) - D(W_{j^*}, X_i) = R\}$

	вектор весов j нейрона выходного слоя. X_i - вектором входного воздействия	
9	Значения коэффициента корреляции близких к нулю указывают на	1. отсутствие нелинейной зависимости 2. отсутствие линейной зависимости 3. отсутствие какой-либо зависимости
10	В регрессионной модели $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \varepsilon$, где $\varepsilon \sim N(0, \sigma^2)$ в качестве оценок коэффициентов регрессионной модели рассматриваются	1. средние значения соответствующих переменных 2. математические ожидания соответствующих переменных 3. коэффициенты уравнения регрессии, полученные методом наименьших квадратов
11	В статистическом методе, называемым критерием Чоу, принятие гипотезы H_0 означает что после разделения исходной совокупности наблюдений на однородные группы:	1. улучшение модели не является значимым, то есть качество модели не повышается. 2. улучшение модели является значимым, то есть качество модели повышается.
12	В случае огромного числа исходных данных показатели Лифт, Левередж, Улучшение используются при	1. формировании всего множества рассматриваемых ассоциаций 2. ограничении множества рассматриваемых ассоциаций путем установки порога значимости, ниже которого ассоциации отбрасываются 3. ограничении множества рассматриваемых ассоциаций путем установки порога значимости, выше которого ассоциации отбрасываются

Лист регистрации изменений и дополнений

№ п/п	№ страницы внесения изменений	Дата внесения изменения	Краткое содержание изменений (основание)	Ф.И.О., подпись	«Согласовано» заве- дующий кафедрой, ведущей дисциплину