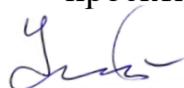


**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Утверждаю:  
Зав. кафедрой компьютерных  
интеллектуальных технологий  
проектирования

  
\_\_\_\_\_ М.И. ЧИЖОВ  
«21» декабря 2021 г.

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ  
ПО ДИСЦИПЛИНЕ**

**«Машинное обучение и анализ данных»**

**Направление подготовки:** 09.04.01 Информатика и вычислительная техника

**Направленность (профиль):** Искусственный интеллект

**Квалификация выпускника** магистр

**Нормативный период обучения** 2 года / 2 года и 5 м.

**Форма обучения** очная / заочная

**Год начала подготовки** 2022

Разработчик

  
\_\_\_\_\_

П.Ю. Гусев

Воронеж – 2021

Процесс изучения дисциплины «Машинное обучение и анализ данных» направлен на формирование следующих компетенций:

ПК-6 - Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач

ПК-8 - Способен руководить проектами по созданию, поддержке и использованию системы искусственного интеллекта на основе нейросетевых моделей и методов

**Перечень планируемых результатов обучения и показателей оценивания сформированности компетенций на этапе промежуточной аттестации**

№ п/п	Компетенция	Результаты обучения, характеризующие сформированность компетенции	Тип ОМ	Показатели оценивания
1	ПК-6	Знать актуальные методы обработки и анализа данных, методы алгоритмов машинного обучения в области	Вопросы (тест) к экзамену	Полнота знаний
		Уметь применять методы машинного обучения при решении задач в различных прикладных областях; уметь подготавливать данные для машинного обучения; использовать библиотеки языка Python для построения моделей машинного обучения	Стандартные задания	Наличие умений
		Владеть навыками использования библиотек языка Python для построения интеллектуальных систем.	Прикладные задания	Наличие навыков
2	ПК-8	Знать сквозные цифровые технологии искусственного интеллекта, включая «Компьютерное зрение», «Обработка естественного языка», «Рекомендательные системы и системы поддержки принятия решений», «Распознавание и синтез речи»; современный опыт применения систем искусственного интеллекта	Вопросы (тест) к экзамену	Полнота знаний
		Уметь проводить сравнительный анализ методов и инструментальных средств для решения задач искусственного интеллекта; разрабатывать и оценивать модели искусственного интеллекта; программировать на языках высокого уровня, ориентированных на применение методов и алгоритмов искусственного интеллекта.	Стандартные задания	Наличие умений
		Владеть навыками выявления требований заказчика к результатам решения задач искусственного интеллекта; определения возможностей применения методов искусственного интеллекта в предметной области решаемой задачи; использования имеющейся методологической и технологической инфраструктуры анализа и обработки данных.	Прикладные задания	Наличие навыков

## ОПИСАНИЕ ПОКАЗАТЕЛЕЙ, КРИТЕРИЕВ И ШКАЛ ОЦЕНИВАНИЯ КОМПЕТЕНЦИЙ НА ЭТАПЕ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Показатели оценивания компетенций	Шкала и критерии оценки уровня сформированности компетенции			
	Неудовлетворительный	Минимально допустимый (пороговый)	Средний	Высокий
<b>Полнота знаний</b>	Уровень знаний ниже минимальных требований. Имели место грубые ошибки	Минимально допустимый уровень знаний. Допущены не грубые ошибки.	Уровень знаний в объёме, соответствующем программе подготовки. Допущены некоторые погрешности.	Уровень знаний в объёме, соответствующем программе подготовки
<b>Наличие умений</b>	При выполнении стандартных заданий не продемонстрированы основные умения. Имели место грубые ошибки.	Продемонстрированы основные умения. Выполнены типовые задания с не грубыми ошибками. Выполнены все задания, но не в полном объеме (отсутствуют пояснения, неполные выводы)	Продемонстрированы все основные умения. Выполнены все основные задания с некоторыми погрешностями. Выполнены все задания в полном объёме, но некоторые с недочетами.	Продемонстрированы все основные умения. Выполнены все основные и дополнительные задания без ошибок и погрешностей. Задания выполнены в полном объёме без недочетов.
<b>Наличие навыков (владение опытом)</b>	При выполнении стандартных заданий не продемонстрированы базовые навыки. Имели место грубые ошибки	Имеется минимальный набор навыков для выполнения стандартных заданий с некоторыми недочетами.	Продемонстрированы базовые навыки при выполнении стандартных заданий с некоторыми недочетами.	Продемонстрированы все основные умения. Выполнены все основные и дополнительные задания без ошибок и погрешностей. Продемонстрирован творческий подход к решению нестандартных задач.
<b>Характеристика сформированности компетенции</b>	Компетенция в полной мере не сформирована. Имеющихся знаний, умений, навыков недостаточно для решения практических (профессиональных) задач. Требуется повторное обучение.	Сформированность компетенции соответствует минимальным требованиям. Имеющихся знаний, умений, навыков в целом достаточно для решения практических (профессиональных) задач, но требуется дополнительная практика по большинству профессиональных задач.	Сформированность компетенций в целом соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в целом достаточно для решения стандартных профессиональных задач.	Сформированность компетенции полностью соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в полной мере достаточно для решения сложных профессиональных задач.

## ПЕРЕЧЕНЬ ОЦЕНОЧНЫХ МАТЕРИАЛОВ

### Вопросы (тестовые задания) для оценки результатов обучения, характеризующих сформированность компетенций

<b><i>ПК-6 - Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач</i></b>	
1.	Преобработка. Масштабирование. Нормировка. Полиномиальные признаки. One-hot encoding.
2.	Кластеризация. kMeans, MeanShift, DBSCAN, Affinity Propagation
3.	Смещение и дисперсия (bias and variance). Понятие средней гипотезы
4.	Ансамблевые методы. Soft and Hard Voting. Bagging. Случайные леса. AdaBoost
5.	Типы обучения: с учителем, без учителя, с подкреплением, с частичным участием учителя, активное обучение
6.	Бустинг деревьев решений
7.	Ошибка внутри и вне выборки. Ошибка обобщения. Неравенство Хёфдинга. Валидация и кросс-валидация
<b><i>ПК-8 - Способен руководить проектами по созданию, поддержке и использованию системы искусственного интеллекта на основе нейросетевых моделей и методов</i></b>	
1.	Линейная регрессия. Полиномиальная регрессия. Гребневая регрессия
2.	Размерность Вапника-Червоненкиса. Размерность Вапника-Червоненкиса для перцептрона
3.	Логистическая регрессия. Градиентный спуск
4.	Пороговые условия. Эффективность по Парето. Precision-Recall и ROC кривые. AUC
5.	Ансамблевые методы регрессии. RANSAC. Theil-Sen. Huber
6.	Перцептрон. Перцептрон с карманом
7.	Метод опорных векторов. Постановка задачи. Формулировка и решение двойственной задачи. Типы опорных векторов. Ядра
8.	Гипотезы и дихотомии. Функция роста. Точка поломки. Доказательство полиномиальности функции роста в присутствии точки поломки
9.	Деревья решений. Информационный выигрыш, критерий Джини. Регуляризация деревьев. Небрежные решающие деревья
10.	Байесовский классификатор. Типы оценки распределений признаков (Gaussian, Bernoulli, Multinomial). EM алгоритм
11.	Нейронные сети. Перцептрон Розенблатта. Функции активации. Обратное распространение градиента. Softmax
12.	Стохастическая оптимизация. Hill Climb. Отжиг. Генетический алгоритм
13.	Метрические классификаторы. kNN. WkNN. Отбор эталонов. DROP5. Kdtree

### Практические задания для оценки результатов обучения, характеризующих сформированность компетенций

<b><i>ПК-6 - Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач</i></b>	
1.	<p>Какое из следующих утверждений является правильным? (Несколько вариантов)</p> <p>A. AdaGrad использует первую производную</p> <p>B. L-BFGS использует вторую производную</p> <p>C. AdaGrad использует вторую производную</p> <p>D. L-BFGS использует первую производную</p> <p>Ответ : AB</p>

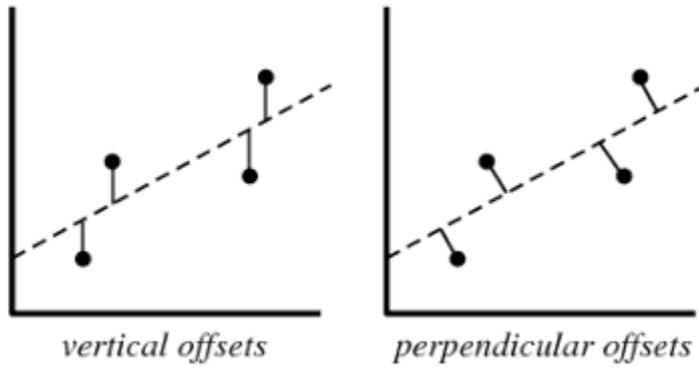
2.	<p>«Увеличение размера сверточного ядра, безусловно, улучшит производительность сверточной нейронной сети». Верно ли это утверждение?</p> <p>А. Правильно  Б. Ошибка  Ответ : В</p>
3.	<p>Предположим, у вас есть 5 ядер свертки в первом слое сверточной нейронной сети. Каждое ядро свертки имеет размер <math>7 \times 7</math>, имеет нулевое заполнение и имеет шаг 1. Размер входного изображения этого слоя составляет <math>224 \times 224 \times 3</math>. Каков размер выходного сигнала этого слоя?</p> <p>А. <math>217 \times 217 \times 3</math>  Б. <math>217 \times 217 \times 8</math>  С. <math>218 \times 218 \times 5</math>  Д. <math>220 \times 220 \times 7</math>  Ответ : С</p>
4.	<p>Если существует нейронная сеть и функция активации - ReLU, если вместо ReLU используется линейная функция активации, может ли нейронная сеть по-прежнему характеризовать функцию XNOR?</p> <p>А. Да  Б. Нет  Ответ : В</p>
5.	<p>Во время обучения машинному обучению размер мини-партии предпочтительно равен степени двойки, например 256 или 512. В чем причина этого?</p> <p>А. Когда мини-пакет является четным, алгоритм градиентного спуска обучается быстрее  Б. Для Mini-Batch установлено значение 2, чтобы удовлетворить требования к памяти процессора и графического процессора и облегчить параллельную обработку.  С. Функция потерь нестабильна, когда четные числа не используются  Д. Ничего из перечисленного  Ответ : В</p>
6.	<p>Какой из следующих методов можно использовать для уменьшения переоснащения? (Несколько вариантов)</p> <p>А. Больше данных об обучении  Б. L1 регуляризация  С. L2 регуляризация  Д. Уменьшить сложность модели  Ответ : ABCD</p>
7.	<p>Что не так со следующим утверждением?</p> <p>О. Когда целевая функция выпуклая, решение алгоритма градиентного спуска обычно является глобальным оптимальным решением.  В. Уменьшение размера PCA требует расчета ковариационной матрицы  С. Направление вдоль отрицательного градиента должно быть оптимальным направлением  Д. Лагранжа может решать задачи оптимизации с ограничениями  Ответ : С</p>
8.	<p>Какие из следующих форм алгоритм K-Means может не агрегировать?</p> <p>А. Круговое распределение  Б. Спиральное распределение  С. Ленточное распространение</p>

	<p>D. Выпуклое распределение полигонов</p> <p>Ответ : B</p>
9.	<p>Какова норма L1 вектора <math>X = [1,2,3,4, -9,0]</math>?</p> <p>A. 1</p> <p>B. 19</p> <p>C. 6</p> <p>D. <math>\sqrt{111}</math></p> <p>Ответ : B</p>
10.	<p>Какое из следующих утверждений о регуляризации L1 и L2 является правильным?</p> <p>A. Регуляризация L2 может предотвратить переоснащение и улучшить обобщающую способность модели, но L1 не может этого сделать</p> <p>B. Технология регуляризации L2 также известна как Регуляризация Лассо.</p> <p>C. Решение, полученное с помощью регуляризации L1, является более разреженным</p> <p>D. L2 регуляризационное решение более разреженное</p> <p>Ответ : C</p>
11.	<p>Есть N образцов, которые обычно используются для обучения и обычно используются для тестирования. Если значение N увеличивается, как изменяется разрыв между ошибкой обучения и ошибкой теста?</p> <p>A. Увеличение</p> <p>B. Уменьшение</p> <p>Ответ : B</p>
12.	<p>Предположим, вы используете функцию активации X в скрытом слое в нейронной сети. При любом входе определенного нейрона вы получите -0,01. Какие из следующих функций активации может иметь X?</p> <p>A. ReLU</p> <p>B. tanh</p> <p>C. Sigmoid</p> <p>D. Все вышеперечисленное возможно</p> <p>Ответ : B</p>
13.	<p>При каких условиях метод ближайшего соседа k-NN работает лучше?</p> <p>A. Больше образцов, но менее типично</p> <p>B. Небольшой образец, но хорошая типичность</p> <p>C. Образцы распределяются по кластерам.</p> <p>D. Образцы распределяются по цепочке</p>
14.	<p>Какой из следующих методов можно использовать для сокращения возможностей? (Несколько вариантов)</p> <p>A. PCA</p> <p>B. Линейный дискриминантный анализ LDA</p> <p>C. AutoEncoder</p> <p>D. Матрица сингулярного разложения SVD</p> <p>E. LeastSquares</p>
15.	<p>Какой из следующих методов не может напрямую классифицировать текст?</p> <p>A. K-Means</p> <p>B. Деревья решений</p> <p>C. опорная векторная машина</p> <p>D. kNN</p> <p>Ответ : A</p>

16.	<p>Что в модели регрессии оказывает наибольшее влияние на компромисс между недостаточной и чрезмерной подгонкой?</p> <p>А. Полиномиальный порядок          В. При обновлении весов <math>w</math> используется ли инверсия матрицы или градиентный спуск?          С. Использование постоянных терминов</p> <p>Ответ : А</p>								
17.	<p>Предположим, у вас есть следующие данные: есть только одна переменная для ввода и вывода. Модель линейной регрессии (<math>y = wx + b</math>) была использована для подгонки данных. Так что же является среднеквадратичной ошибкой, полученной с помощью перекрестной проверки Leave-One Out?</p> <table border="1" data-bbox="279 560 933 728"> <thead> <tr> <th>X(independent variable)</th> <th>Y(dependent variable)</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>2</td> </tr> <tr> <td>2</td> <td>2</td> </tr> <tr> <td>3</td> <td>1</td> </tr> </tbody> </table> <p>А. 10/27          В. 39/27          С. 49/27          D. 55/27</p> <p>Ответ : С</p>	X(independent variable)	Y(dependent variable)	0	2	2	2	3	1
X(independent variable)	Y(dependent variable)								
0	2								
2	2								
3	1								
18.	<p>Какое из следующих утверждений об оценке максимального правдоподобия (MLE) является правильным (множественный выбор)?</p> <p>А. MLE может не существовать          В. MLE всегда рядом          С. Если MLE присутствует, его решение не может быть уникальным          D. Если MLE существует, его решение должно быть уникальным</p> <p>Ответ : AC</p>								
19.	<p>Если мы говорим, что модель «линейной регрессии» идеально подходит для обучающей выборки (ошибка обучающей выборки равна нулю), какое из следующих утверждений верно?</p> <p>А. Ошибка тестового образца всегда равна нулю          В. Ошибка тестового образца не может быть нулевой          С. Ни один из приведенных выше ответов</p> <p>Ответ : С</p>								
20.	<p>В задаче линейной регрессии мы используем R-Squared для определения степени соответствия. В настоящее время, если функция добавлена, а модель не изменена, каков следующий оператор?</p> <p>А. Если R-Squared увеличивается, эта функция имеет смысл          В. Если R-Squared уменьшается, эта функция не имеет смысла          С. Просто взглянув на одну переменную R-Squared, невозможно определить, значима ли эта функция.          D. Ничего из перечисленного</p> <p>Ответ : С</p>								
<p><b>ПК-8 - Способен руководить проектами по созданию, поддержке и использованию системы искусственного интеллекта на основе нейросетевых моделей и методов</b></p>									
1.	<p>Какое из следующих утверждений об остатках в линейном регрессионном анализе является правильным?</p> <p>А. Среднее значение остатков всегда равно нулю</p>								

	<p>В. Средства от остатков всегда меньше нуля  С. Средство остатков всегда больше нуля  D. Ничего из перечисленного  Ответ : А</p>
2.	<p>Какое из следующих утверждений о гетероскедастичности является правильным?  A. Линейная регрессия имеет разные условия ошибки  B. Линейная регрессия имеет тот же термин ошибки  C. Член ошибки линейной регрессии равен нулю  D. Ничего из перечисленного  Ответ : А</p>
3.	<p>Что из следующего отражает сильную корреляцию между X и Y?  A. Коэффициент корреляции составляет 0,9  B. Значение p для нулевой гипотезы <math>\beta = 0</math> составляет 0,0001.  C. Для неверной гипотезы, <math>\beta = 0</math> с t-значением 30  D. Ничего из перечисленного  Ответ : А</p>
4.	<p>Каким из следующих предположений мы руководствуемся при получении параметров линейной регрессии (множественный выбор)?  A. X и Y имеют линейные отношения (полиномиальные отношения)  B. Модельные ошибки статистически независимы  C. Ошибки обычно следуют нормальному распределению 0 средних значений и фиксированному стандартному отклонению  D. X неслучайный и не имеет ошибки измерения  Ответ : ABCD</p>
5.	<p>Чтобы наблюдать линейную зависимость между тестом Y и X, X является непрерывной переменной, какая из следующих графиков больше подходит?  A. Разброс сюжета  B. Гистограмма  C. гистограмма  D. Ничего из перечисленного  Ответ : А</p>
6.	<p>В целом, какой из следующих методов обычно используется для прогнозирования непрерывных независимых переменных?  A. Линейная регрессия  B. Логический обзор  C. И линейная регрессия и логистическая регрессия  D. Ничего из перечисленного  Ответ : А</p>
7.	<p>Коэффициент корреляции между личным здоровьем и возрастом составляет -1,09. Какой вывод вы можете сказать доктору на основании этого?  A. Возраст - хороший показатель здоровья  B. Возраст - плохой показатель здоровья  C. Ничего из перечисленного  Ответ : С</p>

8. Какие из следующих смещений мы используем в случае подгонки линии наименьших квадратов? На рисунке абсцисса является входом  $X$ , а ордината - выходом  $Y$ .



- A. Вертикальные смещения
- B. перпендикулярные смещения
- C. Оба смещения в порядке
- D. Ничего из перечисленного

Ответ : A