


МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«Тульский государственный университет»

Институт прикладной математики и компьютерных наук  
Кафедра прикладной математики и информатики

Утверждено на заседании кафедры  
«Прикладная математика и информатика»  
24 января 2023 г., протокол № 5

И.о. заведующего кафедрой

 Н.В. Ларин

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ (ОЦЕНОЧНЫЕ МАТЕРИАЛЫ) ДЛЯ  
ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И  
ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО  
ДИСЦИПЛИНЕ (МОДУЛЮ)**

**«Методы машинного обучения»**

**основной профессиональной образовательной программы  
высшего образования – программы магистратуры**

по направлению подготовки

**01.04.02 Прикладная математика и информатика**

с направленностью (профилем)

**Перспективные методы искусственного интеллекта  
в сетях передачи и обработки данных**

Форма обучения: очная

Идентификационный номер образовательной программы: 010402-03-23

Тула 2023 год

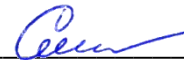
**ЛИСТ СОГЛАСОВАНИЯ**  
**фонда оценочных средств (оценочных материалов)**

**Разработчик:**

Смирнов О.И., доцент каф. ПМий, к.ф.-м.н., доцент

---

*(ФИО, должность, ученая степень, ученое звание)*



---

*(подпись)*

## 1. Описание фонда оценочных средств (оценочных материалов)

Фонд оценочных средств (оценочные материалы) включает в себя контрольные задания и (или) вопросы, которые могут быть предложены обучающемуся в рамках текущего контроля успеваемости и промежуточной аттестации по дисциплине (модулю). Указанные контрольные задания и (или) вопросы позволяют оценить достижение обучающимся планируемых результатов обучения по дисциплине (модулю), установленных в соответствующей рабочей программе дисциплины (модуля), а также сформированность компетенций, установленных в соответствующей общей характеристике основной профессиональной образовательной программы.

Полные наименования компетенций и индикаторов их достижения представлены в общей характеристике основной профессиональной образовательной программы.

## 2. Оценочные средства (оценочные материалы) для проведения текущего контроля успеваемости обучающихся по дисциплине (модулю)

### Перечень контрольных заданий и (или) вопросов для оценки сформированности компетенции ПК-3 (контролируемый индикатор достижения компетенции ПК-3.1)

1. Место и роль ИАД в современной структуре человеческой деятельности. Место ИАД в передаче научного знания. Уровни технологий анализа данных, их назначение, место ИАД в технологиях АД. Понятие о моделировании реального мира в науке. Физическая модель.
2. Модель “решателя”. Информационная модель. Эвристическая модель. Основная особенность ИАД (обучение и эксплуатация эвристической информационной модели). Понятие о машинном обучении.
3. Основные модели данных (dataframe, multidimensional, similaritytensor, transactional). Их назначение научное, технологическое. Гомогенные и гетерогенные модели.
4. Фундаментальные задачи ИАД и основные инструменты статистики. Прикладная жизнь ИАД: декомпозиция содержательных задач предметной области. Научная жизнь ИАД: сведение к задачам фундаментальной математики. Обучение и эксплуатация в фундаментальных задачах.
5. Модель данных «признаковое описание объектов». Понятие о шкалах значений атрибутов. Представление реляционными технологиями. Схемы «звезда» и «снежинка».
6. Многомерная модель данных. Группирование объектов как переход к многомерной модели данных. Аналитические пространства. Измерения и категории. Показатели. Детализация. Функции агрегирования, типы показателей по агрегированию.
7. Транзакционная модель данных. Связанные с ней задачи.
8. Общая задача классификации. Понятие об обучении и использовании. Объект, модель, алгоритм-классификатор. Универсальные ограничения.
9. Локальные ограничения. Оптимизационный подход. Функционалы качества на размеченной выборке. Частотные функционалы качества. Случай бинарной классификации. Стоимостные функционалы качества. Несоответствие частотных и стоимостных функционалов качества человеческому поведению.
10. Подходы к многокритериальной оптимизации.

### Перечень контрольных заданий и (или) вопросов для оценки сформированности компетенции ПК-3 (контролируемый индикатор достижения компетенции ПК-3.2)

1. Понятие байесовского классификатора как оптимального алгоритма распознавания.

Классификаторы, основанные на использовании формулы Байеса. Линейный дискриминант Фишера. Логистическая регрессия.

2. Метрические методы распознавания образов. Метод k ближайших соседей. Взвешенное голосование. Метод парзеновского окна.
3. Метрические методы распознавания образов. Ядровая оценка плотности (kernel density estimation), связь с методами k ближайших соседей и парзеновского окна.
4. Метрические методы распознавания образов. Проклятие размерности (curse of dimensionality).
5. Форматы представления информации. Текстовые файлы, их атрибуты, проблема определения атрибутов.
6. Текстовые форматы представления таблиц: separate dvalues, delimitedtext. Экранирование символов.
7. Форматы представления транзакционных данных.
8. Диаграммы для наборов точек из конечномерных евклидовых пространств.
9. Диаграммы для многомерной модели данных. Системы отчётности.
10. Задача восстановления регрессии: аппроксимационный подход, статистический подход. Простая регрессия. Множественная регрессия. Поиск коэффициентов по МНК. Недостатки МНК. Трёхкомпонентное разложение ошибки регрессионных моделей. Регуляризация по Тихонову. Гребневая регрессия, лассо, эластичные сети.

### **3. Оценочные средства (оценочные материалы) для проведения промежуточной аттестации обучающихся по дисциплине (модулю)**

#### **Перечень контрольных заданий и (или) вопросов для оценки сформированности компетенции ПК-3 (контролируемый индикатор достижения компетенции ПК-3.1)**

1. Рассматривается задача классификации на два класса: положительный и отрицательный. В ходе тестирования классификатора получены следующие результаты: полнота составляет 75%, общая точность составляет 80%. Какие значения может принимать точность?
2. Магазин собрал сведения о покупках (транзакции в файле). Были построены ассоциативные правила. Какое правило, содержащее в условии 2 элемента, имеет наибольшую поддержку?
3. Государственная избирательная комиссия зафиксировала результаты выборов по партиям и по регионам (таблица в файле). Требуется кластеризовать регионы по правилу k-средних для числа кластеров K от 1 до 12. Для каждого числа кластеров K найти максимальный радиус кластера. Построить график этой величины от K. На основании графика предположить, сколько групп регионов разумно выделить по итогам выборов.
4. В алгоритме вычисления оценок написать формулу для числа голосов, если система опорных множеств состоит из всех непустых подмножеств, а функция близости определяется только порогами  $e_i, \dots, e_n$ .
5. Обоснуйте способ построения всех тупиковых тестов через приведение системы тестовых уравнений к не упрощаемой ДНФ.
6. Модель данных «метрические тензоры», гомогенные и гетерогенные многомерные матрицы сходства. Группирование объектов как кластеризация по метрическим описаниям. Гомогенная кластеризация, бикластеризация, мультикластеризация. Основные типы результатов кластеризации (плоская, последовательная плоская, иерархическая, нечёткая, стохастическая, ранговая).
7. Плоская кластеризация. Задача и метод k-means.

8. Последовательная плоская кластеризация. Метод ФОРЕЛЬ.
9. Иерархическая кластеризация. Дивизивная. Агломеративная, функционалы связи (linkage).
10. ROC анализ.

**Перечень контрольных заданий и (или) вопросов для оценки сформированности компетенции ПК-3 (контролируемый индикатор достижения компетенции ПК-3.2)**

1. Линейная модель. Линейная машина как метод обучения линейной модели.
2. Линейная модель. Метод опорных векторов.
3. Модель перцептрона Розенблатта. Метод его обучения. Теорема Новикова. Переход от сдвига к фиктивному признаку.
4. Многослойные перцептроны. Метод обратного распространения ошибки. Функции активации, удобные для распространения ошибки.
5. Многослойные перцептроны. Возможность разделения множеств для перцептронов разной глубины.
6. Решающие деревья.
7. Алгоритмы, основанные на голосовании по наборам закономерностей.
8. Bagging. Boosting. Решающие леса (random forest).
9. Линейные методы распознавания образов. Линейная регрессия методом наименьших квадратов. Логистическая регрессия. Вероятностная интерпретация методов.
10. Переобучение. Регуляризация. Основные методы регуляризации линейных моделей (L2, L1, elasticnet), их свойства. Механизм гребневой регрессии подробно. Вероятностная интерпретация регуляризации.

**4. Оценочные средства (оценочные материалы) для проведения промежуточной аттестации обучающихся (защиты курсовой работы (проекта)) по дисциплине (модулю)**

**Перечень контрольных заданий и (или) вопросов для оценки сформированности компетенции ПК-3 (контролируемый индикатор достижения компетенции ПК-3.1)**

1. Распознавание рукописных цифр, написанных разными людьми.
2. Выделение на изображении сегментов кожи.
3. Периоды покоя и извержения гейзеров.
4. Анализ зависимости времени вылета рейса от пункта назначения.
5. Анализ результатов ЕГЭ по регионам.
6. Связь стоимости просмотра фильма с его характеристиками.
7. Связь стоимости монитора с его характеристиками.
8. SVM. Постановка задачи, вид решения, понятие опорного объекта. Нелинейное обобщение SVM (kernel trick), определение ядра, примеры.
9. Решающие деревья. Жадный алгоритм построения решающего дерева. Критерии ветвления (энтропийный, Джини), их обоснование.
10. Композиции решающих правил. Бустинг: постановка задачи, идея жадного алгоритма. AdaBoost, градиентный бустинг, AnyBoost.

**Перечень контрольных заданий и (или) вопросов для оценки сформированности компетенции ПК-3 (контролируемый индикатор достижения компетенции ПК-3.2)**

1. Композиции решающих правил. Стохастические методы: bagging (+ bootstrap), RSM. Случайные леса. Обоснование стохастических методов: bias-variance decomposition.
2. Нейронные сети. Число слоёв и аппроксимационные возможности сети, причины глубокого обучения. Метод обратного распространения ошибки. Dropout, batch normalization.
3. Анализ изображений. Свёрточные нейронные сети, решение задачи классификации. Задача сегментации изображений: постановка, идея решения. Upsampling (подходы). U-net
4. Анализ изображений. Свёрточные нейронные сети, решение задачи классификации. Задача локализации объекта на изображении: постановка, идея решения.
5. Анализ изображений. Свёрточные нейронные сети, решение задачи классификации. Задача обнаружения объектов: постановка, подходы к решению (R-CNN, FastR-CNN, FasterR-CNN).
6. Анализ текстов. Представление текстов в виде мешка слов. tf-idf. Модель языка. N-граммы.
7. Анализ текстов. Скрытые марковские модели, алгоритм Витерби. Применение скрытых марковских моделей для выделения частей речи.
8. Анализ последовательностей. Рекуррентные нейросети, разные конфигурации входа и выхода. Обучение генератора текста. Задача описания изображений предложениями на естественном языке.
9. Рекуррентные нейросети. Проблема взрыва и затухания градиентов. LSTM.
10. Основания таксономии (способы группирования) фундаментальных задач. Таксономия по наличию в исходных данных целевого признака. Таксономия по моделям данных: в разрезе исходных данных, в разрезе результатов.