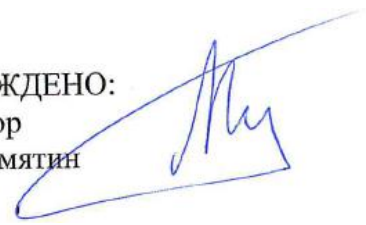


Министерство науки и высшего образования Российской Федерации  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт прикладной математики и компьютерных наук

УТВЕРЖДЕНО:  
Директор  
А. В. Замятин



Оценочные материалы по дисциплине

Методы машинного обучения в телекоммуникациях

по направлению подготовки

**02.04.02 Фундаментальная информатика и информационные технологии**

Направленность (профиль) подготовки:

**Математика беспроводных сетей связи и интернета вещей**

Форма обучения

**Очная**

Квалификация

**Магистр**

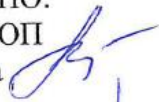
Год приема

**2024**

СОГЛАСОВАНО:

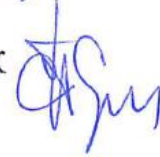
Руководитель ОП

С.П. Моисеева



Председатель УМК

С.П. Сущенко



Томск – 2024

## 1. Компетенции и индикаторы их достижения, проверяемые данными оценочными материалами

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-3 Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-3.1 Проводит анализ математических моделей и систем

ИОПК-3.2 Применяет математические модели, методы для решения прикладных задач профессиональной деятельности

## 2. Оценочные материалы текущего контроля и критерии оценивания

Элементы текущего контроля:

– лабораторная работа;

– коллоквиум.

Лабораторная работа 1. (ИОПК-3.2)

Задание 1. Для  $k=10,15,20,25,30$  сгенерируйте:

1. 200 реализаций нормально распределенной случайной величины с математическим ожиданием  $k$  и стандартным отклонением  $\sqrt{2k}$

2. 200 реализаций случайной величины, распределенной по закону  $\chi^2$  с  $k$  степенями свободы.

Используя тест Колмогорова-Смирнова, проверьте гипотезу о том, что данные выборки относятся к одному непрерывному распределению. Уровень значимости  $\alpha = 0.05$ .

Задание 2. Используя критерий  $\chi^2$  проверьте нулевую гипотезу, состоящую в том, что цвет глаз женщин не зависит от цвета волос.

Задание 3. В таблице содержится информация о скорости передачи данных в зависимости от применяемой технологии. Используя  $t$ -тест, проверьте гипотезу о том, что среднее время чтения для технологий совпадает (используйте разные альтернативные гипотезы). Объясните полученные результаты.

Лабораторная работа 2. (ИОПК-3.1)

*Задание:* Загрузите данные из файла. Постройте регрессию, выражающую зависимость количества корректно переданных текстовых сообщений пропускной способности канала, используя веса наблюдений. Оцените качество построенной модели.

Выполняется с помощью языка программирования R, Python.

В ходе выполнения лабораторных работ ожидается получение решения поставленных задач в полном объеме, демонстрация навыков применения методов и инструментов исследования, умение реализовывать численные расчеты, а также заключение практических выводов и интерпретаций результатов проделанной работы.

Критерии оценивания лабораторных работ (0-10 баллов за каждую работу)

Оценка	Критерии оценивания
10 баллов	Работа выполнена в полном объеме вовремя. Успешное решение поставленной задачи. Демонстрация высокого уровня знаний, умения применять методы исследований, владения аналитическим и цифровым инструментарием для анализа показателей системы и получения необходимых оценок и характеристик, умение реализовать численные расчеты и сделать верные практические выводы.

8 балла	Работа выполнена в полном объеме вовремя или с небольшой задержкой. Успешное решение поставленной задачи. Демонстрация наличия необходимых знаний, успешное применение методов исследования, применение аналитического и цифрового инструментария для решения поставленной задачи, реализация численных расчетов, верные практические выводы. Наличие небольших затруднений при реализации алгоритмов/незначительные ошибки в расчетах конечных показателей, не влияющие на практические выводы анализа системы.
6 балла	Работа выполнена в полном объеме с ошибками или выполнена не полностью вовремя или с небольшой задержкой. Частичное, фрагментарное владение знаниями, умение предложить метод, умение реализовать численные расчеты найденных характеристик. Трудности с анализом и практическими выводами, ошибки в расчетах.
4 балла	Работа выполнена не в полном объеме или не вовремя. Демонстрация низкого уровня знаний, незнание методов исследования. Трудности с анализом и практическими выводами. Трудности с численной реализацией.
2 балл	Попытка выполнения работы, самостоятельное получение отдельных результатов.
0 баллов	Работа не выполнена.

Примерные вопросы к коллоквиуму: (ОПК-3)

1. Предмет и задачи машинного обучения и анализа данных. Основные принципы, задачи и подходы, использование в различных областях науки и индустрии. Основные этапы эволюции алгоритмов машинного обучения.
  2. Обучение с учителем.
  3. Обучение без учителя.
  4. Рекомендательные системы.
  5. Обучение с подкреплением.
  6. Кодирование категориальных данных.
  7. Кросс-валидация (отбор переменных, переобучение на валидационное множество).
  8. Оценки ожидаемой ошибки для линейной регрессии.
  9. Метод главных компонент и *singular spectrum analysis*.
  10. Ядровые методы. Ядра и спрямляющие пространства, методы их построения.
- Ответ на вопрос дается в письменной форме в развернутом виде.

Критерии оценивания результатов текущего контроля

Коллоквиум (0-10 баллов):

Оценка	Критерии оценивания
10 баллов	Дан верный, развернутый и исчерпывающий ответ
7 баллов	Дан верный, но не исчерпывающий ответ или ответ в сжатой форме
3 балла	Дан в целом верный, но не полный ответ, требующий дополнительных пояснений.
0 баллов	Дан неверный ответ.

### 3. Оценочные материалы итогового контроля (промежуточной аттестации) и критерии оценивания

Примеры вопросов зачета: (ОПК-3)

1. Байесовский классификатор.
2. Линейные методы для классификации.
3. Логистическая регрессия.
4. Максимизация правдоподобия.
5. Кросс-энтропия.

Ответ на вопрос дается в письменной форме в развернутом виде. Ответ должен содержать необходимые основные определения, а также примеры, где это возможно.

Результаты зачета определяются оценками «зачтено», «незачтено».

Зачет (0-20 баллов):

Оценка	Критерии оценивания
20 баллов	Дан верный, развернутый и исчерпывающий ответ
15 баллов	Дан верный, но не исчерпывающий ответ или ответ в сжатой форме
10 баллов	Дан в целом верный, но не полный ответ, требующий дополнительных пояснений.
0 баллов	Дан неверный ответ.

Таблица распределения первичных баллов на основании результатов текущего и промежуточного контроля:

Вид работы	Удельный вес	Критерии оценки
Лабораторные работы	20	от 0-10 баллов за выполнение работы (максимум 20 баллов)
Коллоквиум	10	от 0-10 баллов
Зачет	20	от 0-20 баллов

Промежуточная аттестация определяется оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Итоговая оценка по предмету выставляется следующим образом:

- «отлично» – студент набрал не менее 40 первичных баллов и выполнил все лабораторные работы, нет неудовлетворительных оценок за коллоквиум/тест;
- «хорошо» – студент выполнил от 30 до 40 первичных баллов и выполнил все лабораторные работы, нет неудовлетворительных оценок за коллоквиум/тест;
- «удовлетворительно» – студент выполнил от 20 до 30 первичных баллов и выполнил все лабораторные работы, нет неудовлетворительных оценок за коллоквиум/тест;
- «неудовлетворительно» – студент не сдал лабораторные работы, набрал менее 20 первичных баллов или сдал коллоквиум/тест на «неудовлетворительно».

### 4. Оценочные материалы для проверки остаточных знаний (сформированности компетенций)

1. Что такое евклидово расстояние в методе k ближайших соседей?

- A. Расстояние между точками на карте.
- B. Расстояние между двумя линиями.
- C. Математическое ожидание расстояния.

D. Мера сходства между двумя примерами.

Евклидово расстояние является мерой сходства между двумя примерами в методе  $k$  ближайших соседей. Оно определяется как расстояние между двумя точками в  $n$ -мерном пространстве.

2. Как предпочтительно выбирать значение  $k$  в методе  $k$  ближайших соседей?

- A. Значение  $k$  выбирается на основе кросс-валидации.
- B. Значение  $k$  выбирается случайным образом.
- C. Значение  $k$  не имеет значения.
- D. Значение  $k$  выбирается на основе экспертного мнения.

Значение  $k$  предпочтительно выбирать на основе кросс-валидации. Для определения оптимального значения  $k$  часто используется метод кросс-валидации, который позволяет оценить качество модели на независимых данных.

3. Какой алгоритм используется для поиска  $k$  ближайших соседей?

- A. Алгоритм Беллмана-Форда.
- B. Алгоритм Дейкстры.
- C. Алгоритм поиска ближайших соседей.
- D. Жадный алгоритм.

Алгоритм поиска ближайших соседей используется для поиска  $k$  ближайших соседей в методе  $k$  ближайших соседей.

4. Какова сложность алгоритма поиска  $k$  ближайших соседей?

- A.  $O(n \log n)$ .
- B.  $O(n)$ .
- C.  $O(\log n)$ .
- D.  $O(n^2)$ .

. Сложность алгоритма поиска  $k$  ближайших соседей составляет  $O(n \log n)$ , где  $n$  — количество примеров в обучающем наборе.

5. Каким образом можно использовать метод  $k$  ближайших соседей для задачи классификации?

- A. Выбрать  $k$  на основе максимального расстояния до  $k$  ближайших соседей и отнести объект к классу, который наиболее часто встречается среди  $k$  ближайших соседей.
- B. Использовать метод  $k$  ближайших соседей только для задач регрессии.
- C. Выбрать  $k$  на основе минимального расстояния до  $k$  ближайших соседей и отнести объект к классу, который наиболее часто встречается среди  $k$  ближайших соседей.
- D. Нельзя использовать метод  $k$  ближайших соседей для задачи классификации.

Для задачи классификации метод  $k$  ближайших соседей можно использовать следующим образом: выбрать  $k$  на основе минимального расстояния до  $k$  ближайших соседей и отнести объект к классу, который наиболее часто встречается среди  $k$  ближайших соседей. Количество соседей  $k$  может быть выбрано на основе кросс-валидации или других методов выбора оптимального значения.

6. Что такое метод  $k$  ближайших соседей?

- A. Метод градиентного спуска.
- B. Метод машинного обучения для классификации и регрессии данных.
- C. Метод вычисления статистических показателей.
- D. Метод оптимизации функций.

Метод  $k$  ближайших соседей является методом машинного обучения для классификации и регрессии данных. Он используется для прогнозирования категории или значения нового примера на основе его сходства с ближайшими примерами из обучающего набора.

7. Каким образом можно уменьшить влияние выбросов непосредственно при обучении модели машинного обучения методом  $k$  ближайших соседей?

- A. Использовать алгоритмы кластеризации для определения выбросов и исключения их из обучающего набора.

- В. Уменьшить значение  $k$ .
- С. Использовать взвешенный метод  $k$  ближайших соседей, где вес каждого ближайшего соседа зависит от расстояния до нового примера.
- Д. Использовать метрики расстояния, устойчивые к выбросам, например, манхэттенское расстояние.

Использование метрик расстояния, устойчивых к выбросам, является одним из способов уменьшения влияния выбросов в методе  $k$  ближайших соседей. Например, манхэттенское расстояние учитывает только расстояние по каждой из осей, что делает его менее чувствительным к выбросам в одной из осей.

8. Каким образом выбираются веса для взвешенного метода  $k$  ближайших соседей?

- А. Используется расстояние до каждого ближайшего соседа.
- В. Используется расстояние до каждого ближайшего соседа и их классы.
- С. Задается вручную.
- Д. Используется расстояние до каждого ближайшего соседа и их порядковый номер.

Веса для взвешенного метода  $k$  ближайших соседей могут быть выбраны на основе расстояния до каждого ближайшего соседа. Чем ближе сосед к новому примеру, тем больший вес ему присваивается.

9. В чем заключается проблема проклятия размерности в методе  $k$  ближайших соседей?

- А. Увеличение размерности приводит к увеличению ошибки модели.
- В. Проблема проклятия размерности не относится к методу  $k$  ближайших соседей.
- С. Увеличение размерности приводит к увеличению числа примеров в обучающем наборе.
- Д. Увеличение размерности приводит к увеличению времени поиска ближайших соседей.

Проблема проклятия размерности в методе  $k$  ближайших соседей заключается в том, что с увеличением размерности пространства признаков увеличивается объем пространства, что приводит к разреженности обучающего набора и увеличению времени поиска ближайших соседей.

10. Какова максимальная точность, которую можно достичь с помощью метода  $k$  ближайших соседей?

- А. Точность не ограничена.
- В. 50%.
- С. 90%.
- Д. 70%.

Ключи: 1 D; 2 А; 3 С; 4 А; 5 С; 6 В; 7 D; 8 А; 9 D; 10 А.

#### **Информация о разработчиках**

Бронер Валентина Игоревна - кандидат физ.-мат. наук, доцент кафедры теории вероятностей и математической статистики ИПМКН ТГУ