Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный технический университет»
Кафедра экономической информатики

# Паспорт экзамена

по дисциплине «Машинное обучение и нейронные сети», 6 семестр

## 1. Методика оценки

Экзамен проводится в форме письменного тестирования с использованием электронной информационно-образовательной среда НГТУ (http:// www.nstu.ru/sveden/eos).

Тестовые задания охватывают все содержание «Машинное обучение и нейронные сети».

Тест состоит из 40 вопросов различного вида и позволяет проверить результаты обучения по дисциплине, соотнесенные с установленными в программе индикаторами достижения компетенций.

## 2. Критерии оценки результатов обучения, соотнесенных с уровнями освоения индикаторов достижения компетенций

Выполнение теста засчитывается на **продвинутом** уровне, если студент правильно ответил на все вопросы теста, знает определения всех понятий, продемонстрировал способность безошибочно устанавливать последовательность и соответствие между процессами и явлениями, выявлять проблемы, предлагать механизмы их решения, представляет количественные и качественные характеристики определенных процессов и не допускает ошибок. Совокупность результатов обучения по дисциплине и соотнесенных с ними индикаторов достижения компетенций не содержит пробелов. Установленные в программе компетенции сформированы в полном объеме. Оценка составляет *от 35 до 40 баллов.*

Выполнение теста засчитывается на **базовом** уровне, если студент правильно ответил на ⅔ вопросов теста, знает определения основных понятий, продемонстрировал способность устанавливать последовательность и соответствие между процессами и явлениями, правильно характеризует процессы, явления, не допускает существенных ошибок. Совокупность результатов обучения по дисциплине и соотнесенных с ними индикаторов достижения компетенций содержит несущественные пробелы. Установленные в программе компетенции сформированы на базовом уровне. Оценка составляет *от 34 до 27 баллов.*

Выполнение теста засчитывается на **пороговом** уровне, если студент правильно ответил от ½ до ⅔ вопросов теста, знает определения основных понятий, продемонстрировал способность устанавливать последовательность и соответствие между процессами и явлениями, допускает непринципиальные ошибки. Совокупность результатов обучения по дисциплине и соотнесенных с ними индикаторов достижения компетенций содержит пробелы. Установленные в программе компетенции сформированы на пороговом уровне. Оценка составляет *от 26 до 20 баллов.*

Выполнение теста считается **неудовлетворительным**, если студент правильно ответил менее чем на половину вопросов теста, не знает определений понятий, не продемонстрировал способность устанавливать последовательность и соответствие между процессами и явлениями, допускает принципиальные ошибки. Совокупность результатов обучения по дисциплине и соотнесенных с ними индикаторов достижения компетенций содержит существенные пробелы. Установленные в программе компетенции не

сформированы. Оценка составляет *менее 20 баллов.*

### 3. Шкала оценки

Экзамен считается сданным, если сумма баллов по всем заданиям билета составляет от 20 до 40 баллов включительно. Сумма менее 20 баллов признается неудовлетворительным результатом промежуточной аттестации по дисциплине.

В общей оценке по дисциплине экзаменационные баллы учитываются в соответствии с правилами балльно-рейтинговой системы, установленными в НГТУ.

### 4. Примерный тест для экзамена

Утверждаю:
Зав. кафедрой ____ _____ ФИО
«____» _____ 20_ г.

## Тест

по дисциплине «Машинное обучение и нейронные сети»[1]

Вопрос №1 Для сего служит виртуальное окружение Python?

- Данный вопрос не разбирался в рамках курса
- Данное требование является обязательным для всех проектов по аналитике данных.
- **Для управления версиями библиотек и возможности работы с разными версиями для разных проектов.**

Вопрос №2 В чем основная проблема Google Colaboratory?

- Программа исполняется долго.
- **Имеется риск потерять доступ.**
- Никаких проблем нет, эта технология просто работает.

Вопрос №3 Всегда ли требуется производить анализ предметной области перед началом разработки модели?

- Нет, данный шаг является опциональным.
- Да, всегда, так как позволяет больше заработать, если оплата почасовая.
- **Да, всегда, в противном случае, имеется риск предложить неверное решение.**
- Данный вопрос не разбирался в рамках курса.

Вопрос №4 Выберите универсальный (который всегда дает идеальный результат) алгоритм кластеризации из предложенных?

- К-средних (k-means).
- Агломеративная кластеризация.
- **Все зависит от задачи, универсального алгоритма нет.**

Вопрос №5 Что отличает проект с промышленным внедрением алгоритма с использованием машинного обучения от соревнований?

- **На соревнованиях, как правило, нулевая цена ошибки, в то время как при промышленном внедрении ошибка имеет свою цену.**
- Ничем не отличаются, самое главное — это точность модели.

[1] Правильные ответы выделены жирным шрифтом

Вопрос №6 Почему решающее дерево, обученное на результатах кластеризации, не дает объяснение поведения алгоритма кластеризации?

- Вопрос некорректный, решающее дерево позволяет объяснить поведение алгоритма кластеризации.
- **Решающее дерево является отдельной моделью, результаты построения которой на кластеризованных данных могут иметь не 100% точность.**

Вопрос №7 Выберите алгоритм, всегда дающий лучший прогноз на любых данных.

- **Такого не существует**
- Нейронная сеть, обученная на большом объеме данных
- Данную задачу может решить только человек
- ARIMA
- Линейная регрессия

Вопрос №8 Выберите методы, позволяющие извлечь числовые признаки из текста. [MB]

- **TF-IDF**
- **Использование векторов встраивания (embeddings)**
- Из текста нельзя извлечь числовые признаки

Вопрос №9 В чем разница между интерполяцией и аппроксимацией?

- Никакой, термины взаимозаменяемы.
- **Интерполяция - нахождение неизвестных промежуточных значений некоторой функции, по имеющемуся дискретному набору её известных значений, определенным способом, а аппроксимация приближение набора значений некой выбранной функцией.**
- Аппроксимация - нахождение неизвестных промежуточных значений некоторой функции, по имеющемуся дискретному набору её известных значений, определенным способом, а интерполяция приближение набора значений некой выбранной функцией.

Вопрос №10 Какие проблемы возникают при анализе длинных текстов, по сравнению с анализом коротких?

- **В длинных текстах возможно наличие кореференций, требующих разрешения.**
- Никаких проблем нет, методы и подходы одинаковые.
- Все дело в модели, ChatGPT работает одинаково хорошо с любыми текстами.

Вопрос №11 Какой из предложенных алгоритмов машинного обучения не чувствителен к масштабу признаков?

- **Дерево решений (Decision Tree)**
- Нейронная сеть
- Метод к-средних
- Все методы ведут себя одинаково

Вопрос №12 Какие из указанных метрик позволяют сравнивать качество решения задачи регрессии на данных, имеющих разный масштаб?

- Среднеквадратическое отклонение
- Абсолютная ошибка
- **Относительная ошибка**
- Средняя абсолютная ошибка

Вопрос №13 Для чего необходимо разбивать выборку на обучающую и тестовую?

- **Дабы оценить поведение модели на данных, которые модель «не видела».**
- Необходимости нет, модель можно просто обучить на обучающей выборке.

- Без этого никак нельзя, так как это best practice, и иначе проделанную работу не примут.

Вопрос №14 Что такое модель?

- **Это система, исследование которой служит средством для получения информации о другой системе.**
- Это сущность, созданная для решения таких задач как прогнозирование или классификация.
- Обученная нейронная сеть или иной алгоритм машинного обучения.

Вопрос №15 Сколько метрик качества допустимо использовать для оценки качества работы модели?

- Метрика должна быть одна, так как требуется однозначный ответ на вопрос о работоспособности модели.
- Необходимо стремиться к разнообразию метрик качества, так как это позволяет максимально точно определить характеристики модели.
- **Число метрик качества ограничивается целесообразностью их применения.**

Вопрос №16 Что такое признак?

- **Признак – это индивидуальное измеримое свойство или характеристика наблюдаемого явления.**
- Это коэффициент, который необходимо оптимизировать в процессе обучения алгоритма машинного обучения.
- Это список наблюдений за системой или явлением в разные моменты времени.

Вопрос №17 С чего должен начинаться любой проект, связанный с обработкой и/или анализом данных?

- С поиска и первичного анализа данных.
- **С бизнес-анализа или изучения предметной области.**
- С построения базовой (baseline)

модели. Вопрос №18 Что такое «обучение» модели?

- Это вызов метода .fit() в языке Python.
- Ответ на данный вопрос зависит от задачи.
- **Процесс подбора значений параметров модели, обеспечивающих оптимум целевой функции на некоторых данных.**
- Данный вопрос не разбирался в рамках

курса. Вопрос №19 Что такое метод градиентного спуска?

- Это метод классификации, основанный на принципе спуска в потенциальную яму разделяющей гиперплоскости в пространстве признаков.
- **Это метод оптимизации (поиска минимума/максимума) некоторой функции.**
- Это метод кластеризации, основанный на принципе спуска в потенциальную яму разделяющей гиперплоскости в пространстве признаков.

Вопрос №20 На каком языке программирования можно производить анализ данных?

- **На любом языке программирования.**
- Только на Python, так как он изначально был разработан для подобных задач.
- Обычно на Python, но если данных много или требуется высокая производительность, то необходимо использовать C++.

Вопрос №21 Какой из алгоритмов машинного обучения не чувствителен к масштабу признаков?

- Нейронные сети.
- К-средних.
- Линейная регрессия.
- **Деревья решений.**

Вопрос №22 Выберите истинное утверждение.

- **Обученная модель машинного обучения должна работать, в том числе, и на данных, которые не видела.**
- Лучшим методом машинного обучения является нейронная сеть. Данный факт подтверждается числом публикаций и упоминаний в СМИ.
- Деревья решений не «страдают» от переобучения.

Вопрос №23 При построении бинарного классификатора, Вы получили точность 99%. Хороший ли это результат?

- Да, хороший.
- Нет, плохой.
- **Ответ зависит от поставленной задачи.**

Вопрос №24 Необходимо ли требовать определение границ применимости от разработчика модели?

- Нет, нужды в этом нет никакой, так как цена ошибки, как правило не большая.
- **Да, необходимо, так как модель может показывать удовлетворительные результаты только на очень ограниченном наборе данных.**
- Только если решаем задачу для отраслей типа медицины.

Вопрос №25 Вы обучили нейронную сеть классифицировать цифры по их изображению (MNIST). Что Вы получите, если подадите на вход данной сети свое чёрно-белое фото?

- Будет выведена ошибка, так как модель обучали для работы с изображениями цифр.
- В ответ получим NaN, так как на вход подали то, что модель никогда не видела.
- **Нейросеть «скажет», что это одна из цифр.**

Вопрос №26 Что такое кластеризация?

- **Это многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем, упорядочивающая объекты в сравнительно однородные группы.**
- Это многомерная процедура, выполняющая визуализацию данных с одновременных разбиением на однородные группы.
- Это двумерная процедура, выполняющая визуализацию данных с одновременных разбиением на однородные группы.

Вопрос №27 Необходимо ли масштабировать признаки перед кластеризацией методом к-средних?

- **Зависит от выбранной меры расстояния.**
- Нет, не нужно, метод не чувствителен к разному масштабу признаков.
- Обязательно нужно, так как расстояние Эвклида, которое всегда используется в методе к-средних, чувствительно к разному масштабу признаков.

Вопрос №28 Какой геометрический смысл расстояния Эвклида?

- Расстояние городских кварталов (сумма расстояний по каждой координате).
- **Гипотенуза прямоугольного треугольника.**
- Угловое расстояние между двумя векторами.

Вопрос №29 Какой метод всегда дает правильное число кластеров?

- К-средних и метод локтя.
- К-средних и метод Силуэта.
- Использование спектрального метода кластеризации?
- **Универсального метода не существует.**

Вопрос №30 Что такое ChatGPT?

- **Это модель, обученная прогнозировать слова по поданному контексту (набору слов/тексту), «завернутая» в пользовательский интерфейс.**
- Это универсальная модель машинного обучения, позволяющая решать широкий спектр задач со 100% качеством.
- Это искусственный интеллект.
- Это бот в телеграмм.
- Данный вопрос не разбирался в рамках курса.

Вопрос №31 Что теряется при извлечении числовых признаков из текста методом TF-IDF?

- Потерь нет, метод позволяет преобразовывать текст в матрицу признаков и обратно.
- **Теряется порядок слов, особенно если речь идет об униграммах (одиночных словах).**

Вопрос №32 Возможно ли обучить большую языковую модель (типа GPT3) в домашних условиях?

- Нет, это невозможно, так как требует огромных объемов оперативной памяти и множества графических процессоров.
- **Да, это возможно, однако потребуется оптимизировать алгоритмы и программный код.**
- Данный вопрос не разбирался в рамках курса.
- Нет, это невозможно, так как обучение данных моделей является недостижимым технологическим уровнем.

Вопрос №35 Что такое временной ряд?

- **Это последовательность данных, собранных в разные моменты времени.**
- Это множество данных, собранных в разные моменты времени.
- Данный вопрос не разбирался в рамках курса.

Вопрос №36 Как доказать, что временной ряд не является детерминированным?

- Необходимо задаться моделью, подобрать значения ее коэффициентов и доказать их статистическую значимость.
- Можно обучить нейронную сеть классифицировать временные ряды на случайные и детерминированные на большом наборе синтетических данных. Применение нейронной сети к заданному временному ряду позволит судить о его детерминированности.
- **В общем случае нельзя доказать, что ряд является случайным.**

Вопрос №37 Какой метод машинного обучения можно использовать для создания классификатора изображений?

- Сверточная нейронная сеть, так как она разработана для решения задач, связанных с изображениями.
- Полносвязная нейронная сеть.
- Дерево решений.
- **Все перечисленные выше.**

Вопрос №38 Возможно ли взять уже обученную нейронную сеть и использовать ее выход (или выход с некого слоя) для решения своей задачи?

- Нет, каждая нейронная сеть разрабатывается для решения конкретной задачи, таким образом, необходимо обучать свою модель.
- **Да, можно. Более того, данный подход позволяет существенно сэкономить на**

**обучении модели.**

Вопрос №39 Почему графические процессоры используют для обучения нейронных сетей?

- Вопрос поставлен некорректно. Графические процессоры используют только для работы с графикой.
- **Произведение вычислений в нейронной сети можно свести к последовательности матричных операций, в которых графические процессоры более эффективны чем центральные процессоры.**
- Данный вопрос не разбирался в рамках курса.

Вопрос №40 Можно ли так изменить изображения, чтобы нейросеть давала неправильный, но нужный нам ответ?

- Будучи искусственным интеллектом, нейронные сети практически никогда не ошибаются, а следовательно это невозможно.
- **Да, это возможно.**
- Нет, это невозможно.