Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный технический университет»
Кафедра прикладной математики

Паспорт зачета

по дисциплине «Введение в анализ данных и машинное обучение в биологии», 2 семестр

1. Методика оценки

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

Тест состоит из 10 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 при решении задач. 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 18 до 20 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 при решении задач. 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 13 до 17 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. 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 10 до 12 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, при решении задач 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 10 баллов.

3. Шкала оценки

Зачет считается сданным, если сумма баллов по всем заданиям билета составляет от 10 до 20 баллов включительно. Сумма менее 10 баллов признается неудовлетворительным результатом промежуточной аттестации по дисциплине.

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

4. Примерный тест для зачета

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

Тест

по дисциплине «Введение в анализ данных и машинное обучение в биологии»

Вопрос 1. Когда появилось понятие искусственных нейронных сетей?
- Более 100 лет назад
- Менее 20 лет назад
- 50-70 лет назад
- 20-50 лет назад
- 70-100 лет назад

(один вариант)

Вопрос 2. Какие типы задач решают с использованием искусственных нейронных сетей?
- Синтез речи
- Распознавание речи
- Обработка естественного языка
- Классификация изображений
- Предсказание временных рядов

(возможно нескольких вариантов)

Вопрос 3. В каких задачах машинного обучения, как правило, используется обучение с учителем?
- Регрессии
- Кластеризации
- Прогнозирования
- Классификации

(возможно нескольких вариантов)

Вопрос 4. Для задачи классификации верно
- Количество классов известно заранее
- Есть объекты, для которых класс заранее известен
- Объект может принадлежать только одному классу

(возможно нескольких вариантов)

Вопрос 5. Для обучения нейронных сетей, используют
Метод сопряженных градиентов
- Алгоритм обратного распространения ошибки
- Алгоритм прямого распространения ошибки
- Метод градиентного спуска

(возможно нескольких вариантов)

Вопрос 6. Функцию активации применяют к
- Сумме входов нейрона
- Входам нейрона
- Взвешенной сумме входов нейрона
- Входам нейрона умноженным на веса

(возможно нескольких вариантов)

Вопрос 7. Глубокими нейронными сетями, называют сети
- В которых много связей
- В которых много слоев

 В которых много нейронов
 В которых нелинейная функция активации
(один вариант)

Вопрос 8. Преимуществом свёрточных нейронных сетей является
 Большее количество связей
 Меньшее количество связей
 Меньшее количество нейронов
 Большее количество нейронов
(один вариант)

Вопрос 9. Использование предобученной нейронной сети позволяет
 Увеличить размер обучающей выборки
 Уменьшить размер обучающей выборки
 Уменьшить количество нейронов
 Сократить время на обучение
(возможно нескольких вариантов)

Вопрос 10. Сколько параметров будет у нейронной сети, если у нее 100 входов, 5 выходов и один скрытый слой из 100 нейронов (считаем что bias не используется)?
Ответ (число):

5 Перечень тем текущего контроля самостоятельной работы

1 Методы анализа символьных последовательностей (регулярные выражения, токенизация, частотный анализ).
2 Алгоритмы представления последовательностей (one-hot кодирование, k-меры, поиск шаблонов).
3 Методы снижения размерности для (PCA, t-SNE, UMAP): принцип, цель применения, интерпретация.
4 Кластеризация: K-means, иерархическая кластеризация, алгоритмы Лейдена и Лувена.
5 Аугментация. Взвешивание классов. Стратегии (на уровне данных и на уровне алгоритма) для обучения модели на биомедицинских данных с сильным дисбалансом классов (например, 1% больных, 99% здоровых)
6 Рекуррентные сети для анализа последовательностей (RNN, LSTM, GRU). Проблема исчезающего градиента.
7 Векторизация входов для рекуррентных сетей: основные способы представления биологических последовательностей на вход RNN.
8 Анализ временных рядов: стационарность, автокорреляция. Модель ARIMA.
9 Свёрточные нейросети для анализа последовательных сигналов (1D-CNN, CNN-LSTM).
10 Применение машинного обучения в кинетическом моделировании (постановка обратной задачи, выбор подходящей модели).