

Паспорт

оценочных материалов для проведения текущего контроля и
промежуточной аттестации обучающихся по дисциплине (модулю)

Б1.В.02 Машинное обучение

Перечень оценочных материалов и индикаторов достижения компетенций,
сформированность которых они контролируют¹

Наименование оценочного средства	Коды индикаторов достижения формируемых компетенции	Номер приложения ²
Лабораторные работы	ИД-1 ПК-3, ИД-2 ПК-3, ИД-3 ПК-3, ИД-1 ПК-4, ИД-1 ПК-5, ИД-2 ПК-5	1
Экзамен	ИД-1 ПК-3, ИД-2 ПК-3, ИД-3 ПК-3, ИД-1 ПК-4, ИД-1 ПК-5, ИД-2 ПК-5	2

Разработал: _____ А.Д. Обухов

Утверждено на заседании кафедры «Вычислительная техника»
протокол №3 от «11» октября 2021 года

Заведующий кафедрой _____ К.В.Святов

¹ Перечисляются все оценочные материалы, указанные в рабочей программе дисциплины.

² Указывается порядковый номер приложения, в котором размещены оценочные средства. Нумерация изменяется в зависимости от имеющихся оценочных средств.

I. Текущий контроль

Приложение 1

Лабораторные работы

1. Процедура выполнения лабораторных работ

Количество проводимых лабораторных работ в течение всего периода освоения дисциплины	3 работы
Формат проведения результатов	Бумажный / Электронный
Методические рекомендации (при необходимости)	Критерии оценки оценивания мероприятий текущего контроля успеваемости (лабораторных работ): лабораторная работа выполнена в полном объеме; по лабораторной работе представлен отчет, содержащий необходимые расчеты, выводы, оформленный в соответствии с установленными требованиями; на защите лабораторной работы даны правильные ответы не менее чем на 50% заданных вопросов

2. Шкала оценивания с учетом срока сдачи³

Процент правильных ответов	Балл
20	1
40	2
50	3
65	4
80	5

3. Перечень лабораторных работ

Лабораторная работа №1 ЛИНЕЙНАЯ РЕГРЕССИЯ

Общее задание

Перед выполнением лабораторной работы необходимо загрузить набор данных в соответствии с вариантом на диск

1. Написать программу, которая разделяет исходную выборку на обучающую и тестовую (training set, validation set, test set)

³ За несвоевременную сдачу обучающемуся могут быть начислены штрафные баллы.

2. С использованием библиотеки scikit-learn (<http://scikit-learn.org/stable/>) обучить модель линейной регрессии по обучающей выборке (пример: http://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.html#sphx-glr-auto-examples-linear-model-plot-ols-py)
3. Проверить точность модели по тестовой выборке
4. Построить модель с использованием полиномиальной функции (пример: http://scikit-learn.org/stable/auto_examples/model_selection/plot_underfitting_overfitting.html#sphx-glr-auto-examples-model-selection-plot-underfitting-overfitting-py). Построить графики зависимости ошибки от степени полиномиальной функции.
5. Построить модель с использованием регуляризации (пример: http://scikit-learn.org/stable/modules/linear_model.html). На основе экспериментов подобрать параметры для регуляризации.

Варианты

Массивы данных берутся из UCI Machine Learning Repository <https://archive.ics.uci.edu/ml/index.php>

Вариант определяется набором данных, который можно загрузить по ссылке выше:

1. Condition Based Maintenance of Naval Propulsion Plants
2. UJIIndoorLoc
3. Insurance Company Benchmark (COIL 2000)
4. KDD Cup 1998 Data
5. Forest Fires
6. Concrete Compressive Strength
7. Concrete Slump Test
8. Communities and Crime
9. Parkinsons Telemonitoring
10. YearPredictionMSD
11. Relative location of CT slices on axial axis
12. Individual household electric power consumption
13. Energy efficiency
14. 3D Road Network (North Jutland, Denmark)
15. ISTANBUL STOCK EXCHANGE
16. Buzz in social media
17. Physicochemical Properties of Protein Tertiary Structure
18. Gas Sensor Array Drift Dataset at Different Concentrations
19. SkillCraft1 Master Table Dataset
20. SML2010
21. Bike Sharing Dataset
22. Combined Cycle Power Plant
23. BlogFeedback

Лабораторная работа №2 ОСНОВЫ НЕЙРОННЫХ СЕТЕЙ

Общее задание

Перед выполнением лабораторной работы необходимо загрузить набор данных в соответствии с вариантом на диск

1. Написать программу, которая разделяет исходную выборку на обучающую и тестовую (training set, validation set, test set), если такое разделение не предусмотрено предложенным набором данных.
2. Произвести масштабирование признаков (scaling).
3. С использованием библиотеки scikit-learn (<http://scikit-learn.org/stable/>) обучить 2 модели нейронной сети (Perceptron и MLPClassifier) по обучающей выборке. Перед обучением необходимо осуществить масштабирование признаков. Пример MLPClassifier: http://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html Пример и описание Perceptron: http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Perceptron.html
4. Проверить точность модели по тестовой выборке.
5. Провести эксперименты и определить наилучшие параметры коэффициента обучения, параметра регуляризации, функции оптимизации. Данные экспериментов необходимо представить в отчете (графики, ход проведения эксперимента, выводы).

Варианты

Массивы данных берутся из UCI Machine Learning Repository <https://archive.ics.uci.edu/ml/index.php>

Вариант определяется набором данных, который можно загрузить по ссылке выше:

1. Abalone
2. Adult
3. Artificial Characters
4. ser Knowledge Modeling Data (Students' Knowledge Levels on DC Electrical Machines)
5. EEG Eye State
6. seismic-bumps
7. banknote authentication
8. Weight Lifting Exercises monitored with Inertial Measurement Units
9. REALDISP Activity Recognition Dataset
10. mage Segmentation
11. ISOLET
12. sEMG for Basic Hand movements
13. Letter Recognition
14. Dataset for Sensorless Drive Diagnosis
15. Phishing Websites
16. Multiple Features
17. Diabetic Retinopathy Debrecen Data Set
18. Page Blocks Classification
19. Optical Recognition of Handwritten Digits
20. Pen-Based Recognition of Handwritten Digits
21. Smartphone-Based Recognition of Human Activities and Postural Transitions
22. Indoor User Movement Prediction from RSS data
23. Spambase

Лабораторная работа №3

АЛГОРИТМЫ КЛАСТЕРИЗАЦИИ ДАННЫХ

Общее задание

Перед выполнением лабораторной работы необходимо загрузить набор данных в соответствии с вариантом на диск.

1. Произвести масштабирование признаков (scaling).
2. С использованием библиотеки scikit-learn (<http://scikit-learn.org/stable/>) написать программу с использованием алгоритмов кластеризации данных, позволяющую разделить исходную выборку на классы, соответствующие предложенной вариантом задаче (<http://scikit-learn.org/stable/modules/clustering.html>).
3. Провести эксперименты и определить наилучший алгоритм кластеризации, параметры алгоритма. Необходимо использовать не менее 3-х алгоритмов. Данные экспериментов необходимо представить в отчете (графики, ход проведения эксперимента, выводы).

Варианты

Массивы данных берутся из UCI Machine Learning Repository <https://archive.ics.uci.edu/ml/index.php>

Вариант определяется набором данных, который можно загрузить по ссылке выше:

1. Sponge
2. Water Treatment Plant
3. Synthetic Control Chart Time Series
4. Character Trajectories
5. Plants
6. Libras Movement
7. KEGG Metabolic Relation Network (Directed)
8. SMS Spam Collection
9. seeds
10. Human Activity Recognition Using Smartphones
11. User Knowledge Modeling
12. NYSK
13. Activities of Daily Living (ADLs) Recognition Using Binary Sensors
14. Dresses_Attribute_Sales
15. Wholesale customers
16. StoneFlakes
17. Gesture Phase Segmentation
18. AAAI 2014 Accepted Papers
19. Dow Jones Index
20. AAAI 2013 Accepted Papers
21. wiki4HE
22. Folio
23. Mice Protein Expression
24. Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson's Disease

II. Промежуточная аттестация

Приложение 2

Экзамен

1. Процедура проведения

Общее количество вопросов к экзамену	15 вопросов
Количество вопросов в билете	2 вопроса
Наличие задач в билете	Нет
Формат проведения	Устно и письменно
Методические рекомендации (при необходимости)	

2. Шкала оценивания с учетом текущего контроля работы обучающегося в семестре

Критерии оценки уровня сформированности компетенций по дисциплине	Балл
Оценка «отлично» выставляется обучающемуся, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, свободно справляется с задачами, вопросами и другими видами применения знаний, не затрудняется с ответом при видоизменении заданий, использует в ответе материал рекомендуемой литературы, правильно обосновывает принятое решение, владеет разносторонними навыками и приемами выполнения практических заданий.	Отлично
Оценка «хорошо» выставляется обучающемуся, если он твердо знает программный материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответах на вопросы, правильно применяет теоретические положения при решении практических заданий, владеет необходимыми навыками и приемами их выполнения.	Хорошо
Оценка «удовлетворительно» выставляется обучающемуся, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала, испытывает затруднения при выполнении практических работ.	Удовлетворительно
Оценка «неудовлетворительно» выставляется обучающемуся, который не знает значительной части программного материала, допускает существенные ошибки в ответах на вопросы, неуверенно, с большими затруднениями выполняет практические	Неудовлетворительно

задания.	
----------	--

3. Вопросы и задачи к экзамену

Теоретические вопросы

1. Линейная регрессия. Стоимостная функция. Градиентный спуск.
2. Логистическая регрессия. Стоимостная функция. Градиентный спуск.
3. Регуляризация параметров линейной регрессии. Метод Лассо.
4. Регуляризация параметров линейной регрессии. Гребневая (ridge) регрессия.
5. Основы нейронных сетей. Метод обратного распространения ошибки.
6. Сверточные сети. Обучение, использование.
7. Масштабирование признаков.
8. Проблемы переобучения и недообучения в машинном обучении.
9. Выбор параметров и архитектуры нейронных сетей.
10. Оценка параметров модели в машинном обучении (precision, recall, F1).
Асимметричные (skewed) классы.
11. Способность нейронных сетей к аппроксимации функций (на пример, XNOR через AND, OR, NOT).
12. Предобработка данных. Проблемы, подходы, методы.
13. Оценка моделей машинного обучения. Подходы к повышению точности моделей.
14. Кластеризация данных. Алгоритм K-means.
15. Методы понижения размерностей. Принципиальный анализ компонент.