

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Макаренко Елена Николаевна

Должность: Ректор

Дата подписания: 18.01.2024 14:33:07

Уникальный программный ключ:

c098bc0c1041cb374cf926cf17146715d8926ae00adc8e27b55cbe1e2dbd7c78

Лист изменений в рабочую программу дисциплины

Дополнения и изменения в рабочую программу дисциплины

«Машинное обучение» по направлению подготовки 01.03.05 «Статистика»

для 2020 года набора.

В рабочую программу внесены следующие изменения:

- п.3 «Структура и содержание дисциплины» изложить в следующей редакции:

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ					
Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетенции	Литература
	Раздел 1. Введение в машинное обучение				
1.1	Тема "Введение в машинное обучение". Постановки основных классов задач в машинном обучении. Обучение с учителем (supervised learning): регрессия и классификация; обучение без учителя (unsupervised learning): кластеризация, снижение размерности; semi-supervised learning, рекомендательные системы, обработка текстов: тематическое моделирование, построение аннотаций, извлечение ответов на вопросы, машинный перевод; обработка изображений: порождение, преобразование; обучение представлений; обучение с подкреплением. Примеры задач. Виды данных: структурированные таблицы, тексты, изображения, звук, логи. Признаки. /Лек/	8	2	ПК-2 ПК-3	Л1.1 Л1.4 Л1.3 Л1.2 Л1.5Л2.1 Л2.3 Л2.2
1.2	Тема "Статистические оценки и проверка гипотез". Основные понятия математической статистики: статистические оценки (точечные и интервальные), их свойства, проверка гипотез. /Лек/	8	2	ПК-2 ПК-3	Л1.1 Л1.4 Л1.3 Л1.2 Л1.5Л2.1 Л2.3 Л2.2
1.3	Тема. «Введение в машинное обучение: методы, виды, способы машинного обучения. Основы языка Python». Основные библиотеки, инструменты анализа данных и принципы применения инструментов машинного обучения на языке Python и в среде разработки Google Colab /Лаб/	8	2	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3
1.4	Тема. «Введение в машинное обучение: методы, виды, способы машинного обучения. Основы языка Python». Основы работы с Pandas. Предобработка данных. Анализ данных с помощью Pandas. Визуализация данных.(С использованием языка Python и в среде разработки Google Colab) /Лаб/	8	2	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3Л2.1 Л2.2 Л2.3
1.5	Тема "Статистические оценки и проверка гипотез". Основные понятия математической статистики: статистические оценки (точечные и интервальные), их свойства, проверка гипотез.(С использованием языка программирования Python и в среде разработки Google Colab) /Лаб/	8	2	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3

1.6	Тема "Введение в машинное обучение". Постановки основных классов задач в машинном обучении. Обучение с учителем (supervised learning): регрессия и классификация; обучение без учителя (unsupervised learning): кластеризация, снижение размерности; semi-supervised learning, рекомендательные системы, обработка текстов: тематическое моделирование, построение аннотаций, извлечение ответов на вопросы, машинный перевод; обработка изображений: порождение, преобразование; обучение представлений; обучение с подкреплением. Примеры задач. Виды данных: структурированные таблицы, тексты, изображения, звук, логи. Признаки. /Ср/	8	4	ПК-2 ПК-3	Л1.1 Л1.4 Л1.3 Л1.2 Л1.5Л2.1 Л2.3 Л2.2
1.7	Тема "Статистические оценки и проверка гипотез". Основные понятия математической статистики: статистические оценки (точечные и интервальные), их свойства, проверка гипотез. /Ср/	8	4	ПК-2 ПК-3	Л1.1 Л1.4 Л1.3 Л1.2 Л1.5Л2.1 Л2.3 Л2.2
	Раздел 2. Линейные модели				
2.1	Тема "Машинное обучение как математическое моделирование". Статистические модели. Теоретико-вероятностная постановка задачи обучения с учителем. Минимизация ожидаемой ошибки. Пример: задача регрессии, минимизация квадрата отклонения. Регрессионная функция: условное матожидание. Линейная регрессия и метод k ближайших соседей. Переобучение и недообучение. Разложение ошибки на шум, смещение и разброс. Методы оценивания обобщающей способности, кросс-валидация. /Лек/	8	2	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
2.2	Тема «Введение в линейные модели и задача регрессии». Функции потерь. Метрики качества регрессии. Линейная регрессия, метод наименьших квадратов и максимизация правдоподобия. Теорема Гаусса—Маркова. Явный вид решения в методе наименьших квадратов. Ковариационная матрица для коэффициентов. Практические соображения: что делать с категориальными данными? /Лек/	8	2	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
2.3	Тема "Машинное обучение как математическое моделирование". Статистические модели. Теоретико-вероятностная постановка задачи обучения с учителем. Переобучение и недообучение. Метод k-ближайших соседей. Логистическая регрессия. Классификация методом деревьев решений. (С использованием языка программирования Python и в среде разработки Google Colab) /Лаб/	8	2	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
2.4	Тема «Введение в линейные модели и задача регрессии». Линейная регрессия, метод наименьших квадратов и максимизация правдоподобия. Ошибки модели. Метрики качества. (С использованием языка программирования Python и в среде разработки Google Colab) /Лаб/	8	4	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3

2.5	Тема "Машинное обучение как математическое моделирование". Статистические модели. Теоретико-вероятностная постановка задачи обучения с учителем. Минимизация ожидаемой ошибки. No free lunch theorem. Пример: задача регрессии, минимизация квадрата отклонения. Регрессионная функция: условное матожидание. Линейная регрессия и метод k ближайших соседей. Переобучение и недообучение. Разложение ошибки на шум, смещение и разброс. Проклятие размерности. Методы оценивания обобщающей способности, кросс-валидация. /Ср/	8	6	ПК-2 ПК-3	Л1.1 Л1.4 Л1.3 Л1.2 Л1.5Л2.1 Л2.3 Л2.2
2.6	Тема «Введение в линейные модели и задача регрессии». Градиентный спуск, методы оценивания градиента. Функции потерь. Метрики качества регрессии. Линейная регрессия, метод наименьших квадратов и максимизация правдоподобия. Теорема Гаусса—Маркова. Явный вид решения в методе наименьших квадратов. Ковариационная матрица для коэффициентов. Практические соображения: что делать с категориальными данными? Вычислительные соображения: точное решение vs градиентный спуск. Регуляризация. /Ср/	8	6	ПК-2 ПК-3	Л1.1 Л1.4 Л1.3 Л1.2 Л1.5Л2.1 Л2.3 Л2.2
	Раздел 3. Признаковые представления				
3.1	Тема "Линейные модели и задача классификации". Задачи классификации. Общая постановка. Байесовский классификатор. Метод классификации k-ближайших соседей. Линейные методы для классификации. Логистическая регрессия, максимизация правдоподобия, кросс-энтропия. /Лек/	8	2	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
3.2	Тема "Выбор и оценка моделей, работа с признаками". Кросс-валидация: тонкости (отбор переменных, переобучение на валидационное множество). Оценки ожидаемой ошибки для линейной регрессии: AIC и другие. L1 и L2 регуляризация. Методы отбора признаков. Метод главных компонент и singular spectrum analysis. /Лек/	8	2	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
3.3	Тема "Линейные модели и задача классификации". Задачи классификации. Классификация «k-ближайших соседей». Линейные методы для классификации. Логистическая регрессия, максимизация правдоподобия. Кросс-валидация. L1 и L2 регуляризация. Классификация на основе деревьев решений. (С использованием языка программирования Python и в среде разработки Google Colab) /Лаб/	8	4	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
3.4	Тема "Линейные модели и задача классификации". Задачи классификации. Общая постановка. 0-1 ошибка. Байесовский классификатор. Линейные методы для классификации. Логистическая регрессия, максимизация правдоподобия, кросс-энтропия. /Ср/	8	6	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3

3.5	Тема "Выбор и оценка моделей, работа с признаками". Кросс-валидация: тонкости (отбор переменных, переобучение на валидационное множество). Оценки ожидаемой ошибки для линейной регрессии: AIC и другие. L1 и L2 регуляризация. Методы отбора признаков. Метод главных компонент и singular spectrum analysis. Ядровые методы. Ядра и спрямляющие пространства, методы их построения. Операции в спрямляющих пространствах. /Ср/	8	6	ПК-2 ПК-3	Л1.1 Л1.4 Л1.3 Л1.2 Л1.5Л2.1 Л2.3 Л2.2
Раздел 4. Решающие деревья и композиции					
4.1	Тема "Деревья и ансамбли". Ограничения линейных методов (пример: XOR). Решающие деревья. CART. Ансамбли. Бутстреп. Бэггинг. Случайный лес. /Лек/	8	2	ПК-2 ПК-3	Л1.1 Л1.4 Л1.3 Л1.2 Л1.5Л2.1 Л2.3 Л2.2
4.2	Тема "Бустинг" AdaBoost, градиентный бустинг. XGBoost. /Лек/	8	2	ПК-2 ПК-3	Л1.1 Л1.4 Л1.3 Л1.2 Л1.5Л2.1 Л2.3 Л2.2
4.3	Тема "Деревья и ансамбли". Применение различных алгоритмов в построении деревьев решений и ансамблевых методов. (С использованием языка программирования Python и в среде разработки Google Colab) /Лаб/	8	4	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
4.4	Тема "Деревья и ансамбли". Ограничения линейных методов (пример: XOR). Решающие деревья. CART. Ансамбли. Бутстреп. Бэггинг. Случайный лес. /Ср/	8	6	ПК-2 ПК-3	Л1.1 Л1.4 Л1.3 Л1.2 Л1.5Л2.1 Л2.3 Л2.2
4.5	Тема "Бустинг" AdaBoost, градиентный бустинг. XGBoost. /Ср/	8	6	ПК-2 ПК-3	Л1.1 Л1.4 Л1.3 Л1.2 Л1.5Л2.1 Л2.3 Л2.2
Раздел 5. Кластеризация и методы снижения размерности					
5.1	Тема "Кластеризация". K-means. EM-алгоритм. Другие методы кластеризации: иерархическая кластеризация, DBSCAN, Affinity Propagation. /Лек/	8	2	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
5.2	Тема "Снижение размерности". SVD-разложение. Метод главных компонент. t-SNE, UMAP /Лек/	8	2	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
5.3	Тема "Кластеризация". Применение K-means и других методов кластеризации. (С использованием языка программирования Python и в среде разработки Google Colab) /Лаб/	8	2	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
5.4	Тема "Снижение размерности". Метод главных компонент. (С использованием языка программирования Python и в среде разработки Google Colab) /Лаб/	8	2	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
5.5	Тема "Кластеризация". K-means. EM-алгоритм. Другие методы кластеризации: иерархическая кластеризация, DBSCAN, Affinity Propagation. /Ср/	8	8	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
5.6	Тема "Снижение размерности". SVD-разложение. Метод главных компонент. t-SNE, UMAP /Ср/	8	6	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
Раздел 6. Нейронные сети					

6.1	Тема "Введение в нейросети". Нейронные сети: общая архитектура. Реализация XOR с помощью трёх персептронов. Теорема об универсальной аппроксимации. Многослойные сети. Обратное распространение ошибки. Стохастический градиентный спуск. Проблемы: затухающие и взрывающиеся градиенты, невыпуклость функции потерь. /Лек/	8	2	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
6.2	Тема "Современные нейросетевые архитектуры" Нейронные сети в обработке изображений. Фильтры. Сверточные слои. /Лек/	8	4	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
6.3	Тема "Признаковые представления для дискретных входных данных". Практические соображения. Кодирование категориальных данных. Пропущенные значения. Обработка текстов: bag of words, tf-idf, векторные эмбединги. /Лек/	8	2	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
6.4	Тема "Современные нейросетевые архитектуры" Нейронные сети в обработке изображений. Фильтры. Сверточные слои. (С использованием языка программирования Python и в среде разработки Google Colab) /Лаб/	8	4	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
6.5	Тема "Признаковые представления для дискретных входных данных". Практические соображения. Кодирование категориальных данных. Пропущенные значения. Обработка текстов: bag of words, tf-idf, векторные эмбединги. /Ср/	8	10	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
6.6	Тема "Введение в нейросети". Нейронные сети: общая архитектура. Реализация XOR с помощью трёх персептронов. Теорема об универсальной аппроксимации. Многослойные сети. Обратное распространение ошибки. Стохастический градиентный спуск. Проблемы: затухающие и взрывающиеся градиенты, невыпуклость функции потерь. /Ср/	8	10	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
6.7	Тема "Современные нейросетевые архитектуры" Нейронные сети в обработке изображений. Фильтры. Сверточные слои. Нейронные сети и обучение представлений. Обработка последовательностей. Рекуррентные нейронные сети. /Ср/	8	10	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3
6.8	/Экзамен/	8	36	ПК-2 ПК-3	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5Л2.1 Л2.2 Л2.3

2. п.5.1. «Основная литература» код Л1.3 изложить в следующей редакции:

Л1.3	Андрей Бурков	Машинное обучение без лишних слов	Санкт-Петербург: Питер, 2020	https://ibooks.ru/reading.php?short=1&productid=3679 91 неограниченный доступ для зарегистрированных пользователей
------	---------------	-----------------------------------	---------------------------------	--

3. п. 5.4. «Перечень программного обеспечения» изложить в следующей редакции:

Google Colab

4. Задания к лабораторным работам Приложения 1 к рабочей программе изложить в следующей редакции:

Задание к лабораторной работе 1. Введение в машинное обучение: методы, виды, способы машинного обучения. Основы языка Python.

Цель: ознакомиться с основами языка Python, получить умения для выполнения дальнейших лабораторных работ, изучить типизацию данных, научиться пользоваться циклами «for» и «while», рассмотреть «ветвление» в Python, отработать задачи с использованием конструкции «try-except», разобрать функции и пространства имён.

Задания:

Задание выполняется в Google Colab

Используйте предложенный датасет.

1. Откройте блокнот с примером по данной теме.
2. Выполните задание по алгоритму, указанному в блокноте.
3. Сделайте выводы.

Задание к лабораторной работе 2. Введение в машинное обучение: методы, виды, способы машинного обучения. Основы языка Python.

Цель: научиться пользоваться библиотекой Pandas и её встроенными объектами для анализа и визуализации данных в датасетах, получить умения по использованию библиотеки Pandas, сформировать понятия о DataFrame и Series; научиться строить графики с помощью scatter matrix (матрица рассеивания) и matplotlib.

Задания:

Задание выполняется в Google Colab

Используйте предложенный датасет.

1. Откройте блокнот с примером по данной теме.
2. Выполните задание по алгоритму, указанному в блокноте.
3. Сделайте выводы.

Задание к лабораторной работе 3. Статистические оценки и проверка гипотез

Цель: научиться применять алгоритмы Python для проверки статистических гипотез (параметрические/непараметрические, корреляционные тесты и тесты нормальности).

Задания:

Задание выполняется в Google Colab

Используйте предложенный датасет.

1. Откройте блокнот с примером по данной теме.
2. Выполните задание по алгоритму, указанному в блокноте.
3. Сделайте выводы.

Задание к лабораторной работе 4. Машинное обучение как математическое моделирование

Цель: научиться работать с данными при помощи визуальных инструментов и разобрать азы классификации при помощи построения простейшего классификатора со статичными параметрами, а также анализировать данные.

Задания:

Задание выполняется в Google Colab

Используйте предложенный датасет.

1. Откройте блокнот с примером по данной теме.

2. Выполните задание по алгоритму, указанному в блокноте.
3. Сделайте выводы.

Задание к лабораторной работе 5. Введение в линейные модели и задача регрессии

Цель: изучить метод простейший метод классификации данных, «к-ближайших соседей» и научиться производить оценку данных с помощью визуальных инструментов Python, научиться работать с информацией.

Задания:

Задание выполняется в Google Colab

Используйте предложенный датасет.

1. Откройте блокнот с примером по данной теме.
2. Выполните задание по алгоритму, указанному в блокноте.
3. Сделайте выводы.

Задание к лабораторной работе 6. Линейные модели и задача классификации

Цель: понять и научиться применять метод линейной регрессии в машинном обучении.

Задания:

Задание выполняется в Google Colab

Используйте предложенный датасет.

1. Откройте блокнот с примером по данной теме.
2. Выполните задание по алгоритму, указанному в блокноте.
3. Сделайте выводы.

Задание к лабораторной работе 7. Деревья и ансамбли

Цель: познакомиться с методом машинного обучения, построенном на деревьях решений, а также научить строить сами деревья, сформировать понятие случайного леса, а также научиться использовать данную модель для решения задач.

Задания:

Задание выполняется в Google Colab

Используйте предложенный датасет.

1. Откройте блокнот с примером по данной теме.
2. Выполните задание по алгоритму, указанному в блокноте.
3. Сделайте выводы.

Задание к лабораторной работе 8. Кластеризация

Цель: научиться работать с данными при помощи визуальных инструментов и разобрать азы кластеризации, а также анализировать данные.

Задания:

Задание выполняется в Google Colab

Используйте предложенный датасет.

1. Откройте блокнот с примером по данной теме.
2. Выполните задание по алгоритму, указанному в блокноте.
3. Сделайте выводы.

Задание к лабораторной работе 9. Снижение размерности

Цель: научиться работать с данными при помощи методы снижения размерности на основе анализа главных компонент

Задания:

Задание выполняется в Google Colab

Используйте предложенный датасет.

1. Откройте блокнот с примером по данной теме.
2. Выполните задание по алгоритму, указанному в блокноте.
3. Сделайте выводы.

Задание к лабораторной работе 10. Современные нейросетевые архитектуры

Цель: изучить основы работы с OpenCV и показать основные алгоритмы работы с ним.

Задания:

Задание выполняется в Google Colab

Используйте предложенный датасет.

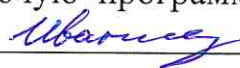
1. Откройте блокнот с примером по данной теме.
2. Выполните задание по алгоритму, указанному в блокноте.
3. Сделайте выводы.

Критерии оценивания


Максимально число баллов – 60.

Максимальное количество баллов по каждой лабораторной работе – 6 баллов.


- 3-6 баллов выставляется, если задание решено полностью, самостоятельно и рационально выбраны инструменты, в представленном решении обоснованно получены правильные ответы, проведен анализ, дана грамотная интерпретация полученных результатов, сделаны выводы, возможны отдельные логические и стилистические ошибки.
- 0-2 балла выставляется, если решение неверно или отсутствует.

Изменения в рабочую программу внесены: составители д.э.н., профессор, Иванченко И.С. 

Согласовано:

Зав. кафедрой: д.э.н., профессор Ниворожкина Л.И. 

Председатель методического совета направления:

к.э.н., доц., Кислая И.А. 

Начальник отдела лицензирования и аккредитации:

Чаленко К.Н.  18.09.2024