Домашнее задание 1.

Каждое из этих заданий предоставляет возможность применить различные методы машинного обучения, которые мы рассмотрели на лекции, оценить их производительность и провести анализ результатов.

1. Используйте набор данных с изображениями рукописных цифр (например, MNIST) и обучите модель классификации (например, метод опорных векторов) для распознавания цифр.
2. Загрузите временной ряд (например, данные о ценах на акции) и создайте модель регрессии, для прогнозирования будущих значений.
3. Используйте набор данных о потреблении электроэнергии и обучите регрессионную модель для прогнозирования будущего потребления.
4. Решите задачу регрессии на наборе данных о продажах автомобилей (например, набор данных Auto MPG) и предскажите расход топлива.
5. Используйте набор данных о рецептах блюд (например, набор данных Recipe Ingredients) и создайте модель классификации, которая будет предсказывать тип кухни по перечню ингредиентов.
6. Загрузите набор данных о посетителях веб-сайта (например, набор данных Online Shoppers) и обучите модель классификации для предсказания, будет ли посетитель совершать покупку.
7. Используйте набор данных о текстах отзывов (например, набор данных IMDB Reviews) и создайте модель классификации для определения тональности текста (положительная/отрицательная).
8. Решите задачу классификации изображений на наборе данных с фотографиями животных (например, набор данных CIFAR-10).
9. Используйте данные о температуре в разных городах и создайте модель регрессии для прогнозирования температуры в будущем.
10. Загрузите набор данных о продажах товаров в супермаркете и создайте модель регрессии для прогнозирования продаж.
11. Решите задачу регрессии на наборе данных о расходе электроэнергии в домах и предскажите будущий расход энергии.
12. Используйте данные о зарплате работников и создайте модель регрессии для прогнозирования зарплаты на основе опыта работы и образования.
13. Загрузите набор данных о населении городов и создайте модель регрессии для прогнозирования роста населения в будущем.
14. Загрузите набор данных о стоимости медицинских процедур и создайте модель регрессии для определения стоимости процедур на основе их характеристик.
15. Загрузите данные о пассажирах самолетов и создайте модель классификации для определения класса обслуживания (первый класс, бизнес-класс, эконом-класс) по характеристикам пассажиров.

Открытые наборы данных можно найти на следующих ресурсах:

1. UCI Machine Learning Repository: Это один из самых популярных репозиториев для машинного обучения. Здесь вы найдете множество датасетов для различных задач, включая классификацию и регрессию. Сайт: <http://archive.ics.uci.edu/ml/index.php>
2. Kaggle: Kaggle - это платформа для соревнований и обмена данными. Она предоставляет доступ к широкому спектру датасетов и соревнований по машинному обучению. Сайт: <https://www.kaggle.com/datasets>
3. OpenML: OpenML - это открытая платформа для обмена данными и экспериментов в машинном обучении. Здесь вы найдете множество датасетов и инструментов для работы с ними. Сайт: <https://www.openml.org/>
4. Google Dataset Search: Google предоставляет поисковик для поиска открытых датасетов. Вы можете воспользоваться им для поиска данных по вашим интересам. Сайт: <https://datasetsearch.research.google.com/>
5. Awesome Public Datasets: Это репозиторий на GitHub, содержащий список открытых датасетов по разным темам. Сайт: <https://github.com/awesomedata/awesome-public-datasets>
6. Data.gov: Этот ресурс предоставляет доступ к огромному количеству открытых данных, включая данные от правительственных организаций США. Сайт: <https://www.data.gov/>
7. Eurostat: Если вам нужны данные о европейских странах, то Eurostat предоставляет открытые данные о различных аспектах экономики, общества и окружающей среды. Сайт: <https://ec.europa.eu/eurostat/data/database>
8. World Bank Data: World Bank Data предоставляет доступ к данным о развитии стран мира. Здесь вы найдете данные о показателях экономического и социального развития. Сайт: <https://datacatalog.worldbank.org/>
9. Amazon AWS Public Datasets: Amazon предоставляет доступ к ряду открытых датасетов, которые можно использовать через их облачный сервис. Сайт: <https://registry.opendata.aws/>
10. DataHub: DataHub - это платформа для обмена данными, где вы также найдете множество открытых датасетов. Сайт: <https://datahub.io/>