



КАРТА МАШИННОГО ОБУЧЕНИЯ

Специально для моих
учеников

NEUVERSO

ОГЛАВЛЕНИЕ

01.

АНАЛИЗ ДАННЫХ

02.

ПОДГОТОВКА ДАННЫХ

03.

МАШИННОЕ ОБУЧЕНИЕ

04.

ЛИНЕЙНЫЕ МОДЕЛИ

05.

КЛАССИФИКАЦИЯ

06.

INTERPRETABLE ML

07.

UNSUPERVISED ML

08.

НЕЙРОННЫЕ СЕТИ

Анализ данных

КЛЮЧЕВЫЕ ТЕМЫ

1

Понимание структуры данных

- Типы переменных (числовые, категориальные)
- Размерность данных
- Баланс классов

2

Описательная статистика

- Среднее, медиана, мода
- Дисперсия, стандартное отклонение
- Квантили
- Корреляция (Пирсона, Спирмена)

3

Анализ распределений

- Нормальное распределение
- Асимметрия
- Выбросы
- Логарифмирование

4

Визуализация данных

- Гистограммы
- Boxplot
- Scatter plot
- Heatmap корреляций

5

Анализ категориальных признаков

- Частотный анализ
- Crosstab

6

Проверка гипотез

- t-test
- Chi-square test
- ANOVA

Вопросы к блоку:

- Как понять, что данные несбалансированы?
- Когда использовать медиану вместо среднего?
- Что показывает корреляционная матрица?
- Как проверить статистическую значимость различий?
- Какие признаки стоит удалить до обучения модели?
- Зачем нам знать нормальное или ненормальное ли распределение?
- Когда нужно логарифмировать данные?



Подготовка данных

КЛЮЧЕВЫЕ ТЕМЫ

1 Проблемы в данных

- Выбросы
- Пропуски
- Шум
- Дубликаты

2 Форматы данных

- Числовые
- Категориальные
- Текстовые
- Временные ряды

3 Масштабирование

- Стандартизация
- Нормализация

4 Feature engineering

- Генерация новых признаков
- Отбор признаков

5 Разделение данных

- Train/test
- Утечка данных

Вопросы к блоку:

- Как определить и обработать выбросы?
- Когда нужно масштабирование?
- В чем разница между One-Hot и Label Encoding?
- Почему нельзя обучать модель без разделения данных?
- Как избежать утечки данных (data leakage)?



Машинное обучение

КЛЮЧЕВЫЕ ТЕМЫ

1

Типы обучения

- С учителем (Supervised)
- Без учителя (Unsupervised)
- С подкреплением (Reinforcement Learning)

2

Примеры задач

- Системы рекомендаций
- Credit scoring системы
- Чат-боты для бизнеса
- Natural Language Processing

3

Проблемы обучения

- Переобучение
- Недообучение

4

Bias-Variance tradeoff

- Bias
- Variance
- Компромисс Bias-Variance

Вопросы к блоку:

- В чем разница между supervised и unsupervised?
- Что такое overfitting?
- Как обнаружить underfitting?



Линейные модели

КЛЮЧЕВЫЕ ТЕМЫ

1

Задача регрессии

- Линейная регрессия
- Функция потерь (MSE, R^2)

2

Задача оптимизации

- Формулировка задачи
- Метод наименьших квадратов
- Градиентный спуск

3

Регуляризация

- L1 (Lasso)
- L2 (Ridge)
- Elastic Net

4

Анализ β -коэффициентов

- Смысл β_0
- Смысл β_i
- Величина коэффициента и масштаб признаков
- Стандартизированные коэффициенты
- p-value и статистическая значимость

Вопросы к блоку:

- Как работает градиентный спуск?
- В чем разница L1 и L2?
- Что происходит при слишком большом learning rate?
- Почему нельзя сравнивать коэффициенты без масштабирования?
- Что показывает p-value для β -коэффициента?



Классификация

КЛЮЧЕВЫЕ ТЕМЫ

1

Задача классификации

- Бинарная классификация
- Мультиклассовая классификация
- Мультилейбл классификация
- Линейно разделимые данные

2

Алгоритмы классификации

- Logistic Regression
- K-Nearest Neighbors (KNN)
- Decision Tree
- Ансамблевые модели

3

Метрики оценки

- Accuracy
- Precision, Recall, F1-score
- ROC-AUC
- Confusion Matrix

4

Работа с несбалансированными данными

- Oversampling
- Undersampling
- SMOTE
- Class weights

Вопросы к блоку:

- В чем разница между бинарной и мультиклассовой классификацией?
- Когда Accuracy вводит в заблуждение?
- Что важнее: Precision или Recall в задаче fraud detection?
- Как работать с сильным дисбалансом классов?
- Как понять, что модель переобучилась?



Interpretable ML

КЛЮЧЕВЫЕ ТЕМЫ

1

Зачем нужна интерпретируемость

- Прозрачность моделей
- Доверие пользователей
- Регуляторные требования
- Отладка и поиск ошибок

2

Типы интерпретации

- Глобальная интерпретация (вся модель)
- Локальная интерпретация (конкретное предсказание)

3

Важность признаков

- Встроенная важность (деревья)
- Permutation importance
- Ограничения importance-метрик

4

Локальные методы объяснения

- SHAP
- LIME
- Partial Dependence Plot (PDP)
- ICE plot

Вопросы к блоку:

- В чем разница между глобальной и локальной интерпретацией?
- Почему SHAP считается более устойчивым методом?
- Когда feature importance может вводить в заблуждение?
- Почему корреляция не означает причинность?
- В каких задачах интерпретируемость критична?



Unsupervised ML

КЛЮЧЕВЫЕ ТЕМЫ

1

Кластеризация

- K-Means
- Иерархическая кластеризация
- DBSCAN

2

Снижение размерности

- Principal Component Analysis
- t-SNE / UMAP

3

Подготовка и анализ данных

- Масштабирование и нормализация
- Обработка выбросов
- Метрики оценки кластеров (silhouette score, Davies-Bouldin index)
- Интерпретация кластеров (что они значат для бизнеса / задачи)

Вопросы к блоку:

- Как выбрать число кластеров?
- В чем разница K-Means и DBSCAN?
- Что делает PCA математически?
- Почему высокая размерность - проблема?



Нейронные сети

КЛЮЧЕВЫЕ ТЕМЫ

1 Основы нейронных сетей

- Искусственный нейрон
- Линейная комбинация признаков
- Функция активации
- Линейная разделимость

2 Функции активации

- Подбор подходящей функции
- Проблема затухающего градиента

3 Обучение нейросети

- Forward pass
- Backpropagation
- Batch / Mini-batch / Stochastic

4 Практические аспекты

- Выбор архитектуры
- Подбор гиперпараметров
- Интерпретация весов
- Диагностика обучения (графики loss)
- Работа с GPU

Вопросы к блоку:

- Почему нужна нелинейность?
- Как работает backpropagation?
- Что делает Dropout?
- Почему ReLU лучше sigmoid активации в глубоких сетях?



БЛАГОДАРЮ ЗА ЧТЕНИЕ!

Привет! Меня зовут Элина и я преподаю data science уже 3 года, консультирую компании и занимаюсь автоматизацией.

СВЯЗАТЬСЯ СО МНОЙ :

[CLICK](#)