Лекция: Деревья Решений

# Введение

В данной лекции мы рассмотрим один из фундаментальных алгоритмов машинного обучения - деревья решений. Этот метод широко применяется как самостоятельно, так и в составе ансамблевых методов.

# 1. Основные понятия и определения

## 1.1 Дерево решений

Определение: Дерево решений - это иерархическая структура, состоящая из узлов принятия решений, где каждый узел представляет собой правило вида 'если-то' на основе значений признаков.

## 1.2 Компоненты дерева решений

• Корневой узел - начальная точка дерева, содержащая все данные

• Внутренние узлы - точки принятия решений по определенному признаку

• Листовые узлы - конечные узлы, содержащие предсказания

• Ветви - связи между узлами, представляющие возможные значения признаков

# 2. Математические основы

## 2.1 Критерии разбиения

1) Индекс Джини:

• Измеряет неоднородность узла

• Значение 0 означает идеальное разделение

• Значение 0.5 означает равномерное распределение классов

2) Энтропия:

• Мера неопределенности в узле

• Значение 0 означает чистый узел

• Максимальное значение при равномерном распределении

# 3. Практическая реализация

## 3.1 Основные параметры модели

1. max\_depth - максимальная глубина дерева

Определяет максимальное количество уровней в дереве. Помогает контролировать сложность модели.

2. min\_samples\_split - минимальное число образцов для разбиения

Определяет минимальное количество образцов, необходимое для разделения узла.

3. min\_samples\_leaf - минимальное число образцов в листе

Определяет минимальное количество образцов, которое должно оказаться в каждом листовом узле.

# 4. Оценка модели

## 4.1 Метрики качества

• Accuracy (точность) - доля правильных предсказаний

• Precision (точность) - доля правильных положительных предсказаний

• Recall (полнота) - доля обнаруженных положительных случаев

• F1-score - гармоническое среднее между precision и recall

# 5. Практические рекомендации

1. Всегда начинайте с простой модели с небольшой глубиной

2. Используйте кросс-валидацию для оценки качества модели

3. Анализируйте важность признаков

4. Проверяйте модель на переобучение

5. Документируйте все эксперименты с параметрами

# Заключение

Деревья решений представляют собой мощный и гибкий инструмент машинного обучения. Их главные преимущества - интерпретируемость и способность работать с различными типами данных. При правильной настройке они могут давать высокие результаты как самостоятельно, так и в составе ансамблевых методов.