

Домашнее задание по машинному обучению 2025г

Постановка задачи

Рассмотрим частицу единичной массы, движущуюся в одмерном потенциальном поле с локальной ямой и вязким трением. Потенциал имеет вид:

$$V(x) = -\frac{V_0}{1 + \left(\frac{x}{x_0}\right)^2},$$

где $V_0 > 0$ - глубина ямы, $x_0 > 0$ - характерная ширина ямы.

Уравнение движения с коэффициентом трения γ :

$$\ddot{x}(t) = -\gamma\dot{x}(t) - \frac{dV}{dx}$$

В зависимости от начальных условий и параметров системы возможны три режима:

- oscillating — частица захвачена в яме и совершает незатухающие (или слабо затухающие) колебания.
- converging — частица постепенно теряет энергию и останавливается вблизи центра ямы ($x=0$).
- diverging — частица обладает достаточной энергией, чтобы уйти далеко от ямы и не вернуться за время наблюдения.

Ваша задача — построить модель машинного обучения, которая по начальным параметрам системы предсказывает режим движения.

Описание датасета

Датасет синтетический, сгенерированный численным моделированием уравнения движения. Каждая строка — один запуск симуляции.

Признак	Тип	Описание
x0	float	Начальная координата ($x(0)$)
v0	float	Начальная скорость ($\dot{x}(0)$)
V0	float	Глубина потенциальной ямы
x0_scale	float	Ширина ямы (параметр (x_0))
gamma	float	Коэффициент трения

Признак	Тип	Описание
trajectory_type	string	Метка класса: "oscillating", "converging", "diverging"

- **Размер:** 20 000 примеров.
- **Шум:** ко всем признакам добавлен гауссовский шум (~3% от типичного масштаба).

Что нужно сделать

Выполните полный цикл машинного обучения:

1. Анализ данных

- Визуализируйте распределения признаков и целевой переменной.
- Постройте scatter-plot'ы (например, `v0` vs `x0`, раскрашенные по классам).
- Найдите и визуализируйте возможные выбросы (например, через boxplot или PCA).
- Вычислите доли классов — сбалансирован ли датасет?

2. Предобработка

- Обработайте данные.
- Закодируйте целевую переменную.
- При необходимости — масштабируйте признаки.
- Разделите данные на обучающую (70%), тестовую (30%) выборки.

3. Выбор и обучение моделей

Реализуйте и сравните как минимум три подхода, например:

- Логистическая регрессия,
- Случайный лес,
- XGBoost / LightGBM,
- Полносвязная нейросеть (MLP).

4. Подбор гиперпараметров

- Используйте кросс-валидацию (`StratifiedKFold`) и `RandomizedSearchCV` или `Optuna` и т.д.
- Зафиксируйте лучшую конфигурацию.

5. Анализ результатов

- Оцените модели по: **accuracy**, **balanced accuracy**, **F1-макро**, **матрица ошибок**.
- Визуализируйте предсказания на 2D-срезах (например, `x0` — `v0` при фиксированных `v0`, `x0_scale`, `gamma`).
- Проанализируйте важность признаков.
- Ответьте на вопросы:
 - Как модель использует энергию (`E_0`)?

- Часто ли она путает `oscillating` и `converging` ? Почему?
- Насколько модель устойчива к выбросам?