

Множество обучающее (Training set)

Синонимы: Выборка обучающая, Множество тренировочное

В машинном обучении обучающее множество представляет собой набор структурированных данных, используемый для обучения аналитических моделей — нейронных сетей, деревьев решений, карт Кохонена и др. Каждая запись обучающего множества представляет собой обучающий пример.

Фактически, обучающее множество представляет собой функцию, заданную таблично парами входных и выходных векторов $[(X_1 Y_1), (X_2 Y_2), \dots, (X_k Y_k)]$. После обучения модель должна реализовывать эту функцию.

Например, обучающее множество для обучения умножению и сложению будет содержать четыре переменных — две входных и две выходных. Входными переменными будут два числа (аргумента) x_1 и x_2 , которые требуется умножить или сложить, а выходными переменными будут y_1 и y_2 — правильные результаты сложения и умножения. Для решения задачи классификации, обучающее множество должно состоять из примеров, для которых априорно задана метка класса.

Обучающее множество должно удовлетворять нескольким **требованиям**:

- отражать правила и закономерности исследуемого процесса, которые должна обнаружить модель и по которым должно строится отображение вход-выход;
- быть репрезентативным, т.е. содержать достаточное количество уникальных примеров, как можно более полно отражающих закономерности исследуемого процесса;
- удовлетворять определенным критериям качества данных;
- не содержать дубликатов и противоречий, пропусков и аномальных значений.

Разделение исходного набора данных на обучающую и тестовую выборки в Logitom осуществляет обработчик разбиение на множества.