

Введение в методы машинного обучения

Кумсков М.И.

Аннотация

Спецкурс в кратком и сжатом изложении описывают широко применяемые на практике методы анализа данных, включая алгоритмы машинного обучения, конструирования и обучения нейросетей.

Рассматриваются различные постановки задач поиска зависимостей и распознавания образов, проводится изложение математических моделей и разбор подходов к их решению, включая:

- Регрессионные модели;
- Оценки прогностической устойчивости;
- Выбор подмножества признаков объектов в модель, такие эволюционные алгоритмы как метод группового учета аргументов (МГУА) и генетические алгоритмы;
- Кластерный анализ;
- Метод главных компонент;
- Методы классификации;
- Деревья решений.

Описаны достоинства и недостатки таких алгоритмов кластерного анализа как построение остовного дерева, K-средних, иерархический, EM-алгоритм, ФОРЭЛЬ. На этой основе показаны способы введения ограничений - правил задания областей допустимых значений входных объектов, поступающих для прогнозирования. Рассмотрены методы сокращения размерности признаков пространств – метод главных компонент, задача многомерного шкалирования.

Представлены основные архитектуры нейронных сетей и известные подходы к проведению их обучения. Описан эффект переобучения нейросетей и подходы для преодоления эффекта. Приведены прикладные задачи, решаемые нейросетями и платформы, на которых проводится практическое решение таких задач, включая платформы TensorFlow и Keras.