

ОБЗОР МЕТОДОВ РЕШЕНИЯ ЗАДАЧИ ОБУЧЕНИЯ С УЧИТЕЛЕМ ДЛЯ РАЗРАБОТКИ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ

Р С Гредасов, В С Кедрин

В данной работе рассматриваются методы решения задачи обучения с учителем для разработки моделей прогнозирования. Задача прогнозирования является основой для финансового планирования в экономике и торговле. В настоящее время для большинства экономических субъектов с учетом применения современных информационных технологий осуществляется накопление исторических значений экономических и физических показателей в базах данных. Вместе с тем, развитие программно-аппаратных средств предоставляет все более мощные вычислительные платформы, на которых возможна реализация моделей прогнозирования с помощью методов машинного обучения в основе которых лежат алгоритмы оптимизации. Базовыми методами машинного обучения для построения моделей прогнозирования являются методы обучения с учителем. Для обзора подобных методов нужно:

1. Получить и форматировать для эффективного моделирования данные обучающей выборки.
2. Обучить на основе полученных данных модель прогнозирования.
3. Проверить работоспособность модели на тестовой выборке.
4. Провести анализ результатов с помощью ROC-диаграммы.

Какой из методов будет наиболее эффективен, по сравнению с другими? Какой метод целесообразно использовать для решения тех или иных задач? Ответы на эти вопросы можно получить в результате экспериментов по проектированию моделей машинного обучения в рамках решения конкретных задач прогнозирования.

Для выполнения поставленных задач лучше всего подходит язык программирования Python, так как он обладает библиотеками, необходимыми для изучения возможностей тех или иных методов. В частности, будут использоваться следующие библиотеки:

- Sclearn [1] - библиотека для машинного обучения и анализа данных
- Matplotlib [2] - библиотека для визуализации данных двумерной и трехмерной графикой
- Pandas [3] - библиотека для обработки и анализа данных

Работая над темой “Обзор методов решения задачи обучения с учителем для разработки моделей прогнозирования”, были рассмотрены следующие методы машинного обучения:

1. Линейная регрессия - это модель регрессии, то есть эта модель принимает функции и прогнозирует непрерывный выход.

2. Логистическая регрессия - то же, что и линейная регрессия, но применима для задачи классификации. Она так же рассчитывает линейный выход, за которым следует функция сохранения на выходе регрессии.

3. Дерево решений - алгоритм на основе дерева, используемый для решения задач регрессии и классификации. Дерево решений выводится из независимых переменных, в каждой внутренней его вершине записано условие, а в каждом листе дерева — прогноз.

4. Случайный лес - это модель ансамбля, в которой несколько деревьев решений объединяются для получения более эффективной модели. Идея заключается в том, чтобы обучить несколько алгоритмов, а затем усреднить полученные от них ответы для задачи регрессии и для задачи классификации.

5. Градиентный бустинг строит модель предсказания в форме ансамбля слабых предсказывающих моделей, обычно деревьев решений. Используя градиентный спуск и обновляя предсказания, основанные на скорости обучения, можно найти значения, на которых ошибка минимальна.

Вывод: исходя из известной на данный момент информации об алгоритмах решения задач классификации и регрессии, по моему мнению, наиболее предпочтительным методом для представленных выше задач является метод градиентного бустинга, так как, несмотря на перечисленные недостатки, он всё ещё является крайне эффективным инструментом благодаря адаптивной технике построения композиции и гибкости в плане использования базовых алгоритмов.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Sklearn // URL: <https://scikit-learn.org/stable/> (дата обращения: 16.04.2022)
2. Matplotlib // URL: <https://matplotlib.org/> (дата обращения: 15.04.2022)
3. Pandas // URL: <https://pandas.pydata.org/> (дата обращения: 15.04.2022).