

УДК 004.032.2

ИССЛЕДОВАНИЕ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ С УЧИТЕЛЕМ НА ПРИМЕРЕ ПРОГНОЗИРОВАНИЯ ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ

Раздьяконов А.А. (Самарский университет)

Научный руководитель – кандидат физико-математических наук Савельев Д.А.
(Самарский университет)

Аннотация

В работе изучены субстраты моделей машинного обучения с учителем, а именно: алгоритмы регрессии и классификации. На основе данных о курсе валют, котировок акций цен на нефть, полученных с сайта Банка России, сервиса Финам и платформы «Независимая статистика и аналитика США» соответственно, было обучено пять типов моделей в разных конфигурациях, включая нейронные сети.

Введение. На сегодняшний день в целях помощи человеку в его информационной деятельности и принятии решений активно развиваются методы обработки информации. Одним из направлений такой деятельности является изучение вопроса прогнозирования различных показателей.

Количество новых брокерских счётов на Московской бирже за прошлый год увеличилось в объёмах сильнее, чем за все прошлые года вместе взятые, а именно на несколько миллионов. Общее же их число всё сильнее приближается к отметке в 10 миллионов. Только это можно считать ярким доказательством того, особенно в последние годы, что присущими атрибутами межличностных отношений и коммуникаций являются финансы и экономика.

Изменения в сфере финансов происходят нелинейно, и иногда может показаться, что цены на акции формируются совершенно случайным образом. Для анализа необходимо предварительно собирать большое количество данных, желательно в их исторической ретроспективе. Вышеперечисленное фактически указывает на то, что самым оптимальным вариантом увеличения доходов является использование алгоритмов машинного обучения.

Основная часть. Для прогнозирования любого рода показателей нельзя просто использовать правила типа: если произошло событие А, то следует исход Б, тогда как событие В влечёт за собой исход Г. Такая система «если-то» не учитывает состояние и изменения среды в динамике, а тем более пласт данных, накопленных за предыдущее время. Так что для задачи прогнозирования следует использовать именно модели машинного обучения, которые могут в процессе обучения улавливать какие-то паттерны в данных, учитывать их нелинейность и фактически, в какой-то степени, имитировать мыслительные процессы и причинно-следственную связь человека.

Так называемое обучение с учителем подразумевает существование большой выборки данных, на основании которой нужно что-то предсказать, и некоторой готовой гипотезы. Данные уже отсортированы по категориям или имеются конкретные значения величин. Такие алгоритмы решают 2 основные задачи: задача регрессии и задача классификации.

Алгоритм регрессии подразумевает получение на вход набора параметров (свойств/«фич») объектов и целевых переменных, которые соответствуют искомому значению, которое необходимо предсказать. С помощью матричных вычислений определяются веса каждого параметра объекта, что в итоге приводит к функции нескольких переменных, которая стремится максимально близко описать уже имеющиеся точки и прогнозировать значения новых.

Разница между задачами классификации и регрессии лишь в том, что в рамках первой предсказывается какая-то дискретная величина/категория. Самым наглядным примером является классификация изображений: на вход алгоритму поступает какая-то фотография, а алгоритм определяет, кто же или что же на ней изображено. Существует множество алгоритмов, лежащих в основе решения той или иной задачи классификации.

Наиболее распространённые следующие: алгоритм ближайших соседей, наивный байесовский классификатор и алгоритм Деревьев решений.

Нейронная сеть является математической моделью с набором множества вычислительных узлов. В рамках одиночного существования каждый узел является весьма маленьким вычислительным центром, который не может эффективно решать сложные задачи. Но как только эти нейроны объединяются в слои, а слои в свою очередь в нейронные сети, намного вырастает вычислительная сложность. Так, например, нейронная сеть с одним скрытым слоем, нелинейной функцией активации и с конечным числом нейронов может приблизить любую функцию, насколько «сложной» ни была входная выборка данных.

Эксперимент заключался в исследовании двух экономических показателей: котировок акций и курса валюты. Первое исследование проводилось в корреляции с ценой на нефть и курс валют, тогда как второе – разрезе задержек курса валюты и также цены на нефть. Для оценки полученных моделей и их сравнения использовались метрики, которые в той или иной степени сравнивают предсказанные данные тестовых признаков с реальными значениями целевых переменных.

Выводы. Для объективной оценки качества предсказаний, которые были получены с помощью созданных в процессе экспериментов моделей, было проведено сравнение предсказанных экономических показателей с реальными значениями. Данная информация была получена с сайтов Финам (для котировок акций) и Банка России (для курса валюты). Сравнение метрик качества моделей показало следующее: закономерным образом качество предсказаний алгоритмов классификации выше, так как результатом является лишь класс, говорящий о росте/падении цены экономического показателя. То есть пользователь получает более «размытый» ответ, что, хоть и несёт за собой несомненную пользу и применимость, но меньшую, нежели результаты работы алгоритмов регрессии, обладающими худшими метриками качества, но прогнозирующими непрерывную, точную величину конкретного показателя.

Раздьяконов А.А. (автор) _____

Савельев Д.А. (научный руководитель) _____