



TIME SERIES PREDICTION

Авторы: Георгий Каспарьянц, Анастасия Новикова, Роман Дегтярёв
Руководитель проекта: Польшковский Даниил Александрович

1 Аннотация

Прогнозирование финансовых рынков является одной из самых интересных и перспективных задач современного мира. Но эта задача скрывает под капотом много сюрпризов. Существует много факторов от которых зависит рост/падение цен на бирже и зачастую учесть всех их невозможно. Но существует версия того, что эти факторы неявно учтет наша нейронная сеть в процессе обучения. Нелинейные нейронные сети позволяют с любой степенью точности аппроксимировать произвольную непрерывную функцию. Поскольку временной ряд представляет собой непрерывную функцию, то применение нейронных сетей кажется вполне оправдано и корректно.

2 Данные

Для проверки гипотезы были взяты данные с Московской биржи.

3 Эксперимент

Для начала хотелось бы посмотреть на то, что дают классические подходы в решении данной задачи.

3.1 Случайный лес

Первой моделью был взят случайный лес. В качестве входных данных подавались значения x_i, \dots, x_{i+k} , а целевыми значениями являлись тренд на $k+1$ день: $sgn(x_{i+k+1} - x_{i+k})$. На основе графиков ниже было подобрано оптимальное значение окна и количества деревьев в ансамбле (рис.1-2).

Для лучших результатов были выбраны значения: $k = 14$ и $nestimators = 160$. Модель была обучена на матрице (1451, 28), где в качестве признаков добавили попарную разность между сигналом на $i+1$ и i шагах, а также значение RSI метрики. Последние дали уменьшение $cross_entropy_loss$ в 4 раза. Обученным алгоритмом предсказывалась другая часть (1486, 28).

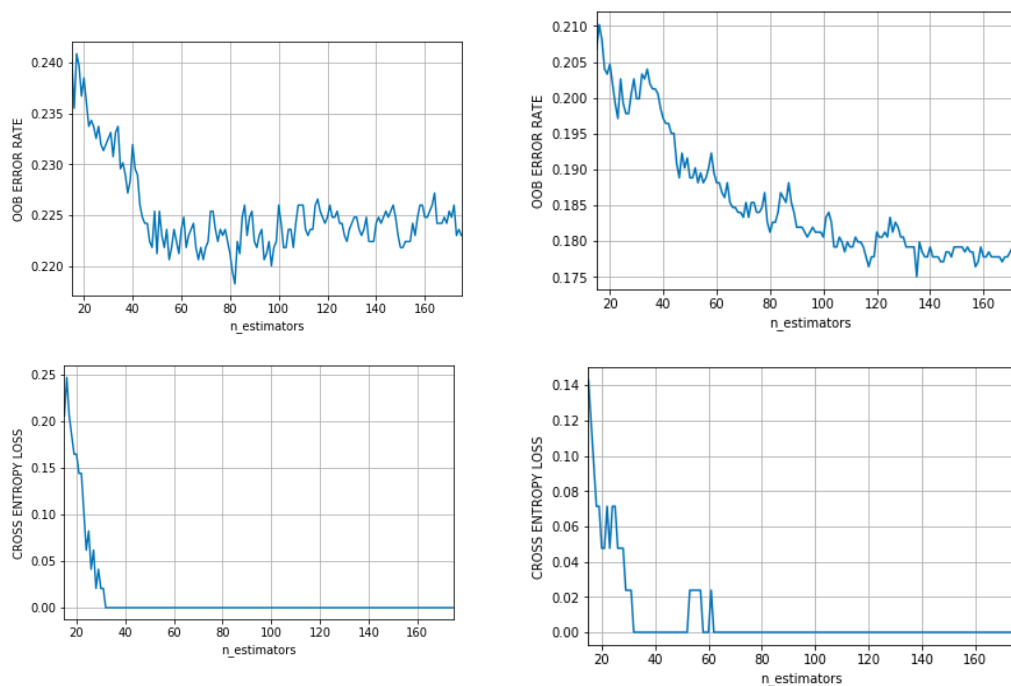
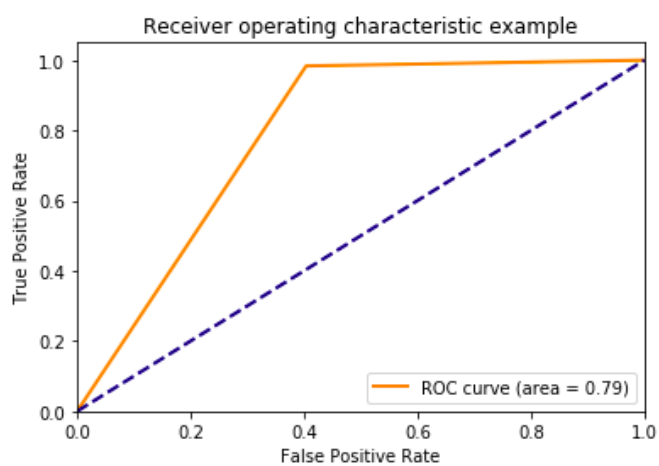


Рис. 1: Значение окна $k = 7$

Рис. 2: Значение окна $k = 14$



CROSS ENTROPY LOSS: 3.00, PRECISION: 0.92, RECALL: 0.98, F1_SCORE: 0.95

Рис. 3: ROC-кривая для теста



В итоге были подсчитаны значения $f1$, $precision$, $recall$, oob_error и построена ROC —кривая для данного случая (рис.3). Можно с уверенностью сказать, что наш алгоритм не переобучился на трейне. Полученные значения достаточно хороши для задач машинного обучения. Вопрос применимости такой модели остаётся открытым. Стоит заметить, что алгоритм на данном этапе протестирован на небольшом объеме данных. Улучшение также можно сделать в плане признаков, не были задействованы Price Rate of Change, Moving Average Convergence Divergence, Williams R, Stochastic Oscillator и другие.