Решение пробного задания по предсказанию серий.

Код: <https://github.com/ikamensh/time_series>

Данные: 250 активов, цены которых меняются каждый час. Цены не меняются по субботам, и данные за субботы отсутствуют. Всего 45 недель данных по всем активам.

Функция потерь: Absolute mean error (выбрана потому, что эта самая ошибка составит убытки при неправильных решениях. Именно ее необходимо минимизировать; выбор квадратной или другой функции потерь даст худшую на практике модель).

Модели, которые я попробовал: LSTM, CNN + LSTM, catboost (представляя последовательность за неделю как 120 features). Хорошие результаты дала LSTM c 2мя слоями + бегущее среднее за последние 6 недель.

Управление рисками: В зависимости от желания балансировать максимальную прибыть и минимизировать риски, можно выбирать разное количество активов, которые предсказываются быть лучшими (верх отсортированного списка). Результаты следующие:

**averaging 1 of the best predictions gives profit of:**

our model allows for the profit of: 277493.0

**averaging 2 of the best predictions gives profit of:**

our model allows for the profit of: 237819.5

**averaging 5 of the best predictions gives profit of:**

our model allows for the profit of: 153954.8

**averaging 15 of the best predictions gives profit of:**

our model allows for the profit of: 123630.3333333

Просто суммируя заранее известные максимумы, видим максимально возможную прибыть (не реалистично для реальной системы, т к используются данные «из будущего»)

maximum possible profit is: 1015296.0

Выбирая случайные активы мы теряем деньги:

expected profit of a random choice: -28273.956

Если бы это была оплачиваемая работа и код бы реально решал какие активы купить, его можно было бы улучшить следующим образом:

1. Попробовать классические модели предсказания последовательностей, такие как:

GARCH, ARIMA, Markov chains и другие

1. Использовать Grid Search для оптимизации гиперпараметров
2. Посмотреть чужие решения аналогичных проблем и почерпнуть решения там

Код: <https://github.com/ikamensh/time_series>

Запускается так:

1. read\_data.py
2. forecast.py

необходимы библиотеки:

* keras
* tensorflow
* sklearn
* pickle
* …