Synopsis

Analyzing the stock market and predicting their prices plays a key role both in the global economy and for private traders to make a profit. Empirically, people figured out the main approaches to understanding the market and learned how to predict price movements based on a huge variety of certain indicators. However, humanity has not yet reached a complete understanding, and to this day this area of research and development of technologies is more relevant than ever.

In our study, we will focus on the question: how to determine the true distribution with which the price of an arbitrary asset is formed and use it to generate all kinds of price movement trajectories in the near future. If we assume that such a generator exists, then from all trajectories with the yield of interest to us, we can calculate the probability of making a profit for a fixed trading strategy. This technology will optimize strategies and provide a new tool for risk management. Moreover, this generator can be used in case of lack of data, for example, from charts of a larger timeframe, to generate a chart of a smaller timeframe that is close to the true one.

For our task, we can use price charts of any asset. However, finance is notorious for the lack of a large amount of data, which can significantly affect the solution. Therefore, it is better to use charts of real assets for future research tests, and build training on an artificially generated dataset. From stochastic differential equations, the well-established hypothesis is known that pricing is subject to the Ornstein-Uhlenbeck process:

$dx_t = -\gamma x_t dt + \sigma dW_t$, where $x_t = x(t)$, $\sigma < 2$, $0 > \gamma >> -0.1$, W_t - Wiener process.

In discreate form: $x_t - x_s = -\gamma x_t \Delta t + \sigma(W_t - W_s)$, where $W_t - W_s \sim N(0, t - s)$,

hence $\Delta x = -\gamma x_t \Delta t + \sigma\sqrt{\Delta t}N(0, 1)$, with $N(0, 1)$ - normal distribution.

Using this function, we can generate artificial closing prices of the trading day of an arbitrary asset. Note that the combination of trading sessions for American securities gives 11 hours of information per day, 5 days per week, excluding holidays (~8 days per year). Therefore, information for one year in the hourly timeframe will give $(52 * 5 - 8) * 11 = 2772$ points. For our research, the generation of 20 such time series is quite suitable, which corresponds to 20 years of information about the closing prices of an arbitrary asset on an hourly timeframe.

If we transform the discrete form of the process by using natural logarithm and try to optimize it, we will get something very similar to an optimization task:

$min_G max_D\ log(D(x)) + log(1 - D(G(z)))$, where D - discriminator and G - is generator of the Generative Adversarial Network. That is why we will use this model for finding a solution to our problem.

We will need to find a suitable GAN network architecture and tune its hyperparameters. First, let's try to figure out how this model behaves with stationary series. To do this, we generate data by the same process, but with the parameters $\gamma = 1$ and $\sigma = 2$. We select the optimal architecture and hyperparameters using statistical hypothesis testing for the similarity of distributions and by the first 4 moments, as well as two-point statistics. Once the model is sufficiently accurate and stable, we will go through the same steps for the artificially generated price data of an arbitrary asset. After setting up the model at this stage, we will move on to the final stage of the project: testing the model on the price chart of a real asset, for example, Apple.

The complexity of the project is significant and at the moment it has practically been possible to proceed to the second stage of the study.