



УНИВЕРСИТЕТ
ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА

Временные ряды



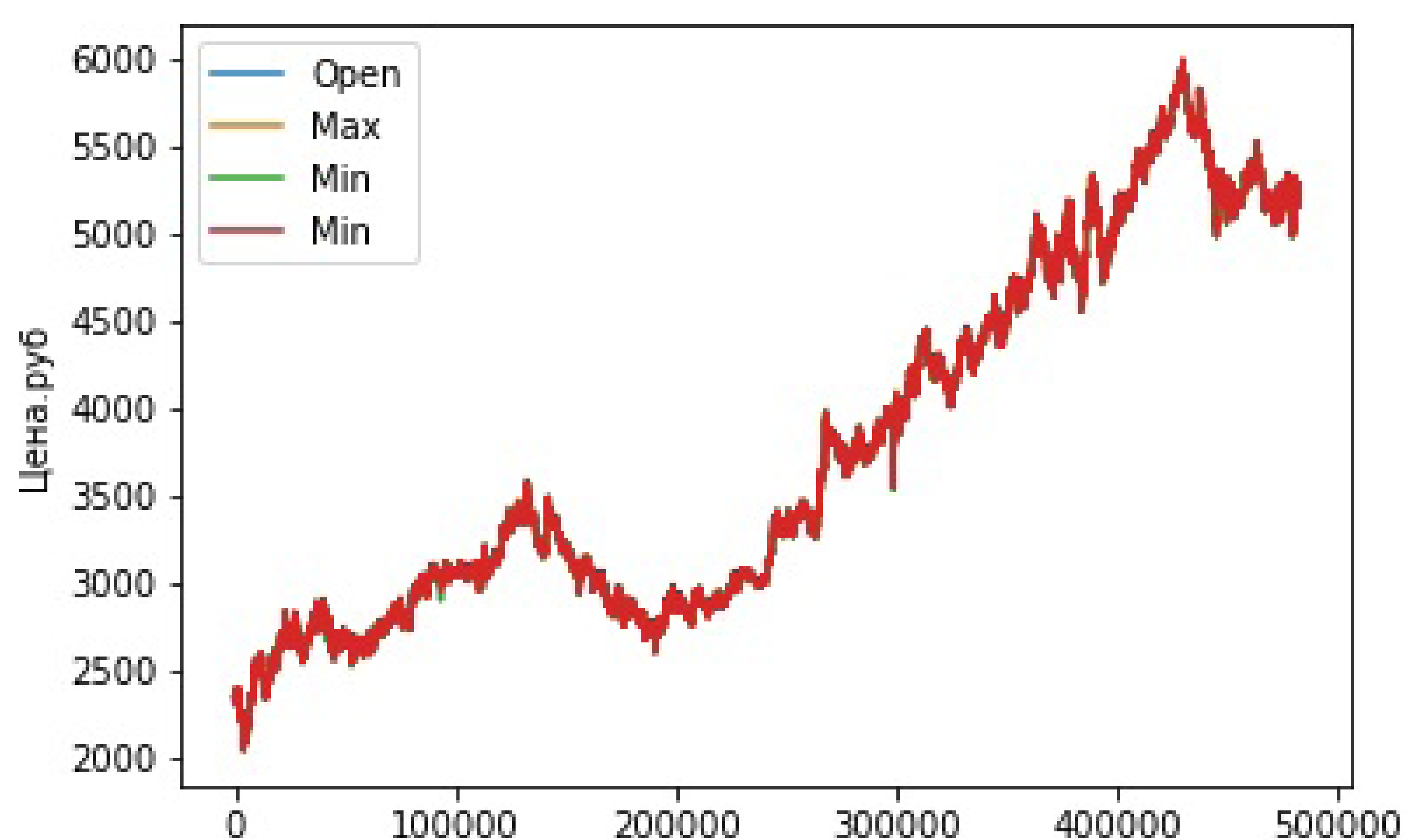


Временные ряды

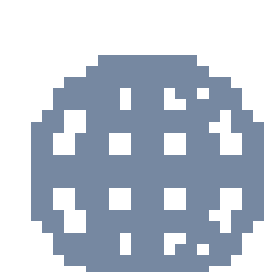
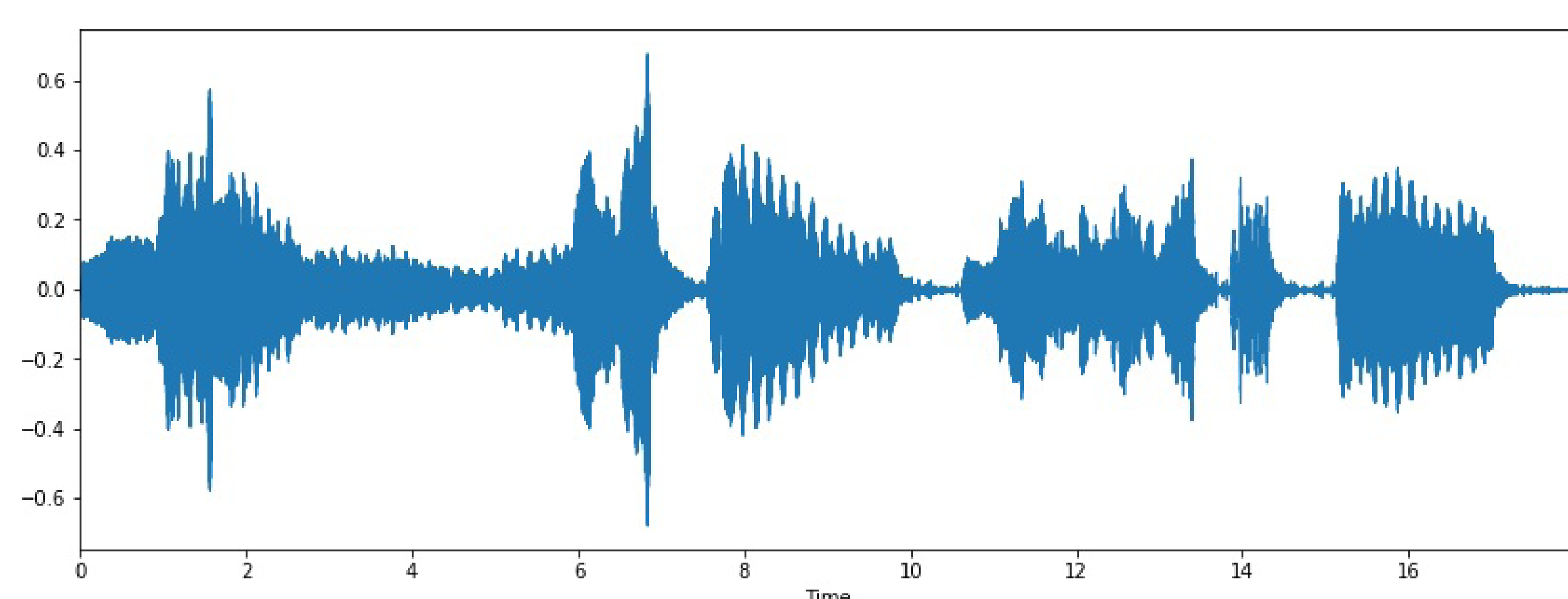
Временной ряд — собранный в разные моменты времени статистический материал о значении каких-либо параметров (в простейшем случае одного) исследуемого процесса.

Примеры:

Изменение цены акций, курса рубля/доллара, стоимости бензина и т. д.



1. Аудиосигнал



Временные ряды

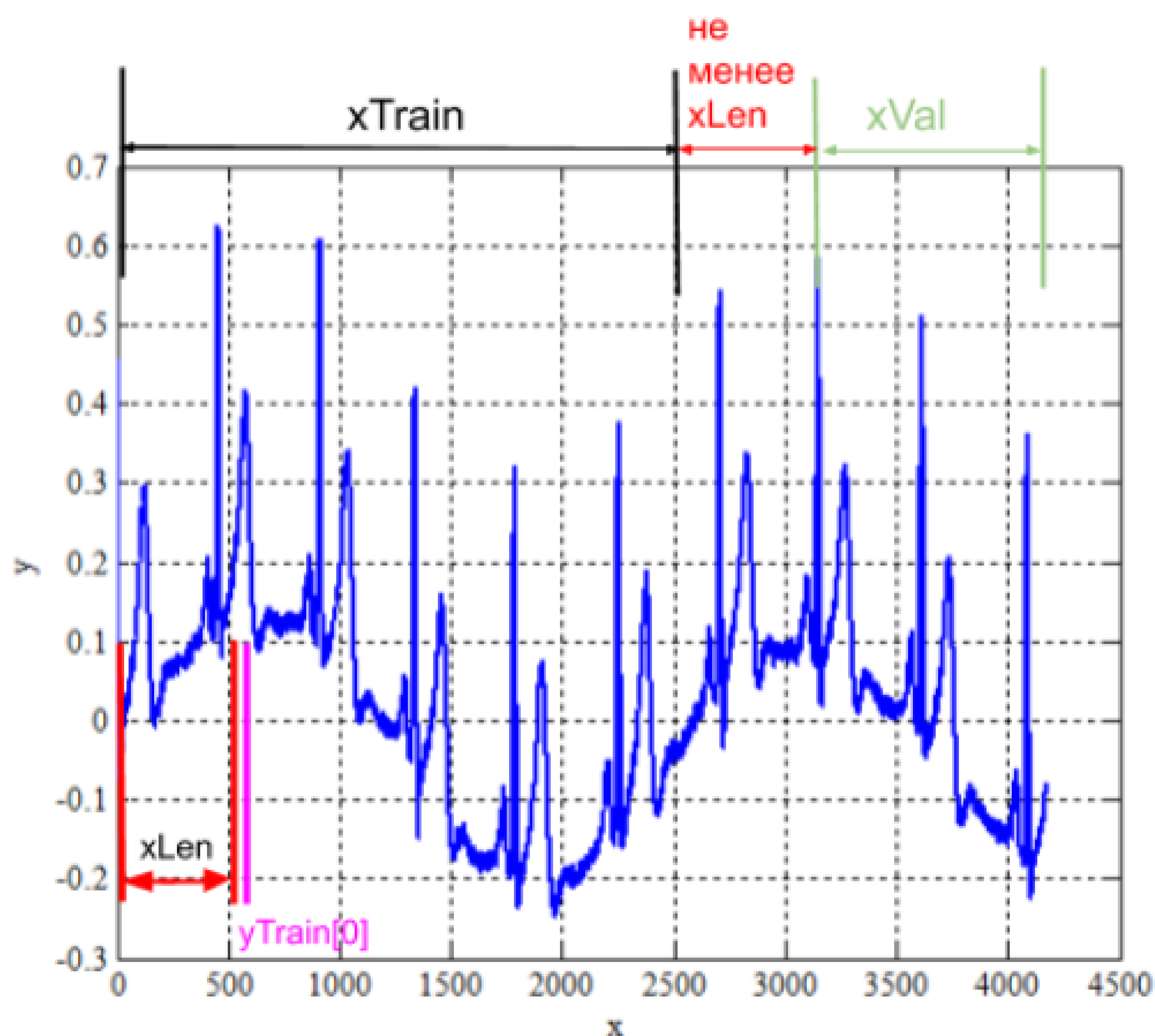
2. Производственные процессы
3. Прогноз погоды
4. Видео (временной ряд из картинок)

Отличия временных рядов в основном в том, что им нужна особая обработка (подготовка) данных перед подачей в нейронную сеть.

Применение временных рядов:

- задачи классификации
- задачи предсказания (предсказание погоды, цены акций и т. д.)

Создание обучающей выборки



Временные ряды

Дан некий временной ряд, на основе которого мы хотим построить

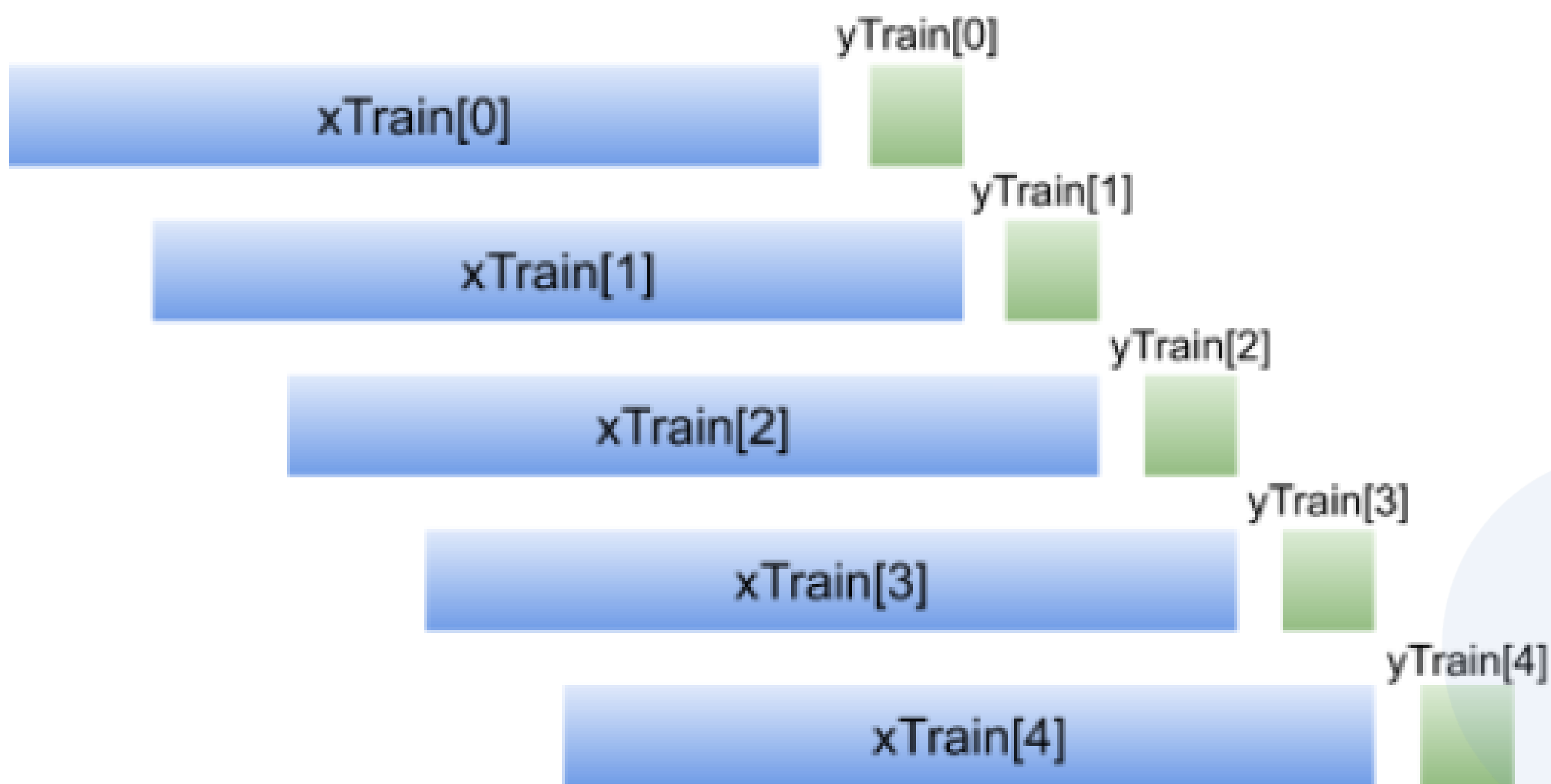
DATE	TIME	OPEN	MAX	MIN	CLOSE	VOLUME
11.01.2010	10:30:00	1744.32	1749	1740	1743	14849
11.01.2010	10:31:00	1743.01	1750	1742.13	1746.04	23797
11.01.2010	10:32:00	1746.05	1750	1739.86	1739.86	17694
11.01.2010	10:33:00	1740.01	1740.01	1736.8	1737.83	6207
11.01.2010	10:34:00	1737.83	1740.26	1737.11	1739	3894
11.01.2010	10:35:00	1740	1742	1738.06	1739	6901
11.01.2010	10:36:00	1738.99	1739.97	1736.8	1738	7271
11.01.2010	10:37:00	1738	1738.68	1736.8	1737.6	1203
11.01.2010	10:38:00	1737.98	1738	1736.8	1736.8	4303
11.01.2010	10:39:00	1736.8	1737.9	1736.6	1737.59	3968
11.01.2010	10:40:00	1736.69	1739	1736.62	1739	3932
11.01.2010	10:41:00	1738.95	1741.7	1738.24	1741	13044
11.01.2010	10:42:00	1740.99	1741.96	1738	1738	5934
11.01.2010	10:43:00	1739.7	1739.7	1737.22	1737.9	2719
11.01.2010	10:44:00	1737.86	1739.21	1736.83	1739.09	8094
11.01.2010	10:45:00	1739.09	1742	1738.9	1742	5160
11.01.2010	10:46:00	1741.89	1743.12	1741	1743	3591
11.01.2010	10:47:00	1743	1744	1741.08	1741.08	6803
11.01.2010	10:48:00	1742	1743.99	1741	1741.02	4765

обучение нейронной сети.

Данные могут иметь следующий вид:

На примере показаны данные изменения цены акций (дата, цена открытия сделки, максимальная цена, минимальная цена, цена закрытия и размер). Для обучения будет использоваться только цена закрытия сделки.

Зададим длину xLen = 500 значений (длина отрезка, по которому анализируем).



Временные ряды

С 1 по 500 значение будут x_{Train} первого элемента, а значение 501 будет y_{Train} первого элемента. Затем смещаемся на единицу, и теперь со 2 по 501 значения будут x_{Train} второго элемента, а значение 502 будет y_{Train} второго элемента и т. д. по всей длине обучающей выборки. Часть данных выделяем на тренировочные (x_{Train}), а часть на проверочные (x_{Val}). Они не должны пересекаться, и между ними должно быть расстояние не меньше, чем длина x_{Len} . Если тренировочные данные и проверочные будут пересекаться, то сеть будет выдавать недействительные показания точности.

Архитектуры

Для решения задач временных рядов применяются следующие архитектуры:

- Полносвязная (Dense)
- Одномерная свёртка (Conv1D)
- Рекуррентная (LSTM, GRU)
- Сложные смешанные сети

Нормализация данных

Во временных рядах мы можем нормировать и x_{Train} , и y_{Train} . Возможно применение следующей нормализации:

1. $N(0, 1)$ — StandardScaler
2. $[-1; 1]$ — MinMaxScaler
3. $[0; 1]$ — MinMaxScaler

Благодаря нормализации данных на вход сети будут подаваться не сами значения, а разница между последовательностью значений. Так сети работают гораздо лучше.

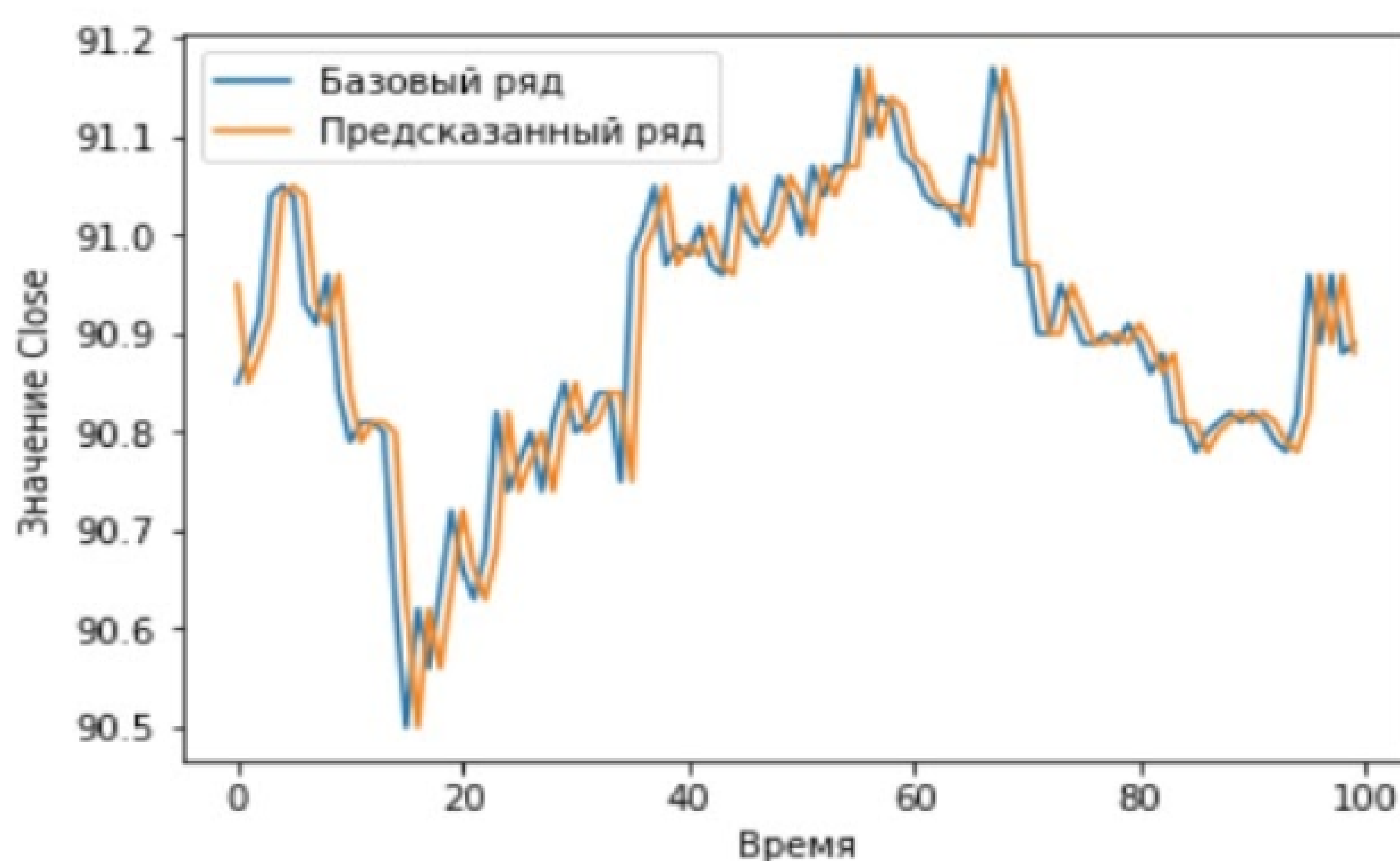
Временные ряды

Автокорреляция

Автокорреляция — это основная проблема временных рядов.

При работе с временными рядами мы используем такие функции ошибки, как mse (mean square error) и mae (mean absolute error), которые определяют разность между двумя значениями.

Когда мы имеем дело с временными рядами и последовательность данных имеет небольшое отклонение, то сеть начинает «хитрить» и выдаёт предсказанное значение, равное предыдущему. На графике это выглядит так:

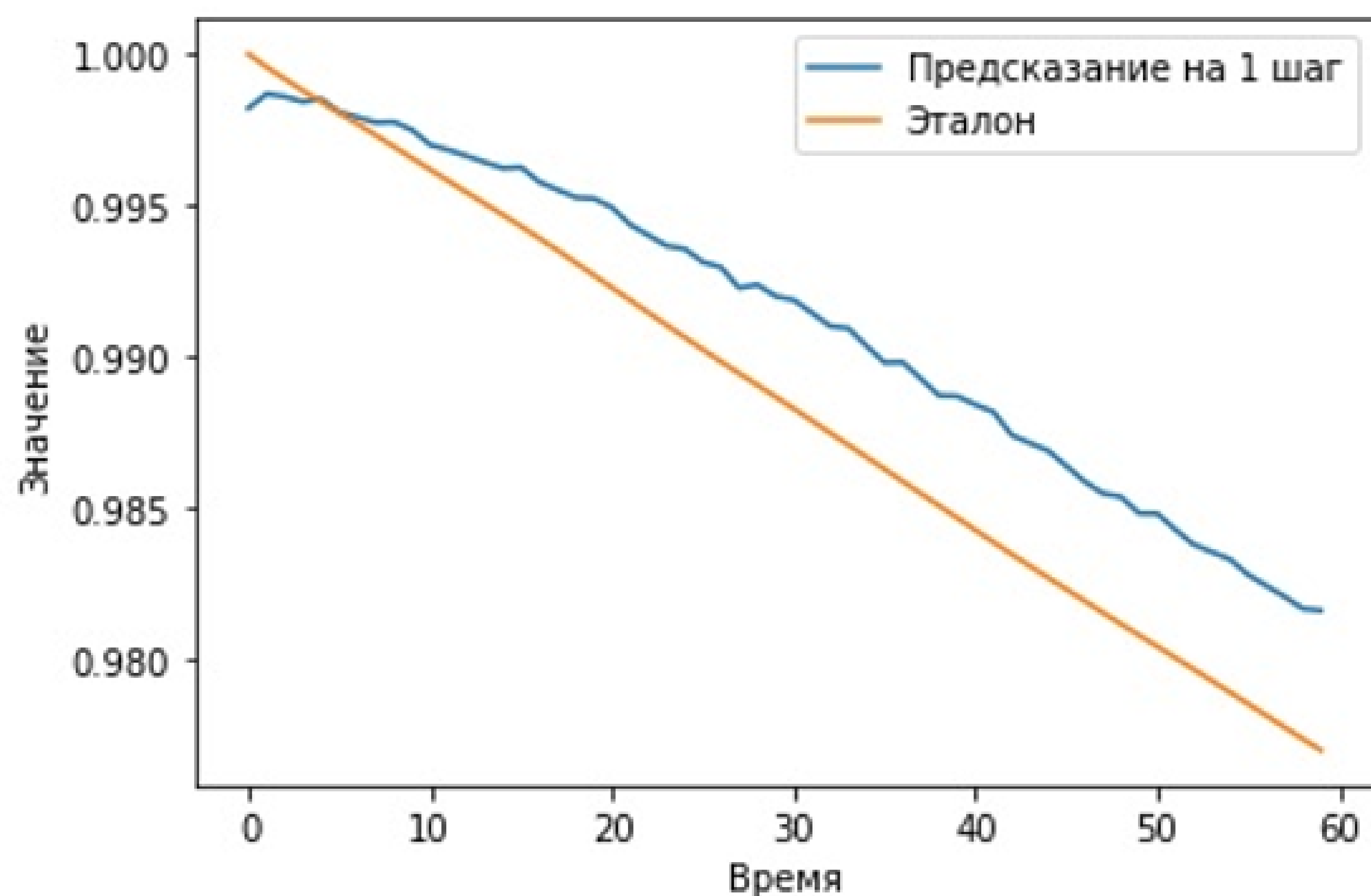


Предсказание в точности повторяет контур базового ряда, но отстаёт на один шаг. При этом значение ошибки сети будет очень низким, а на деле, предсказание будет неудовлетворительным.

С помощью графика корреляции можно определить отклонение.

График корреляции выглядит следующим образом (оранжевая линия — это идеальный график корреляции, синяя линия — корреляция нашего предсказания):

Временные ряды



Основная задача состоит в том, чтобы график корреляции стремился к эталонному.

Рассмотрим пример предобработки базы, разделения на проверочную и тестовую выборки и нормализации данных:

```
xLen = 300 #Анализируем по 300
прошедшим точкам
valLen = 30000 #Используем 30.000 записей
для проверки

trainLen = data.shape[0]-valLen # Размер тренировочной
выборки

#Делим данные на тренировочную и тестовую выборки
xTrain,xTest = data[:trainLen], data[trainLen+xLen+2:]

#Масштабируем данные (отдельно для X и Y), чтобы их легче
было скормить сетке
xScaler = MinMaxScaler()
xScaler.fit(xTrain)
xTrain = xScaler.transform(xTrain)
xTest = xScaler.transform(xTest)
```

Временные ряды

Выбираем размер $xLen$, по которому будет анализироваться предсказание следующего значения. Выделяем на проверочную выборку 30000 записей. Всего датасет имеет примерно 481000 записей.

Разделяем данные на тренировочную и тестовую выборки. Не забываем, что расстояние между тренировочной и тестовой не менее $xLen$, в нашем случае это $xLen+2$.

Далее делаем нормализацию данных при помощи метода `MinMaxScaler` библиотеки `sklearn`.

Данные подготовили, дальше обучаем сеть на выбранной архитектуре и смотрим график корреляции. На выходном слое сеть имеет один нейрон и активационную функцию *linear* (не всегда).

С помощью нейронных сетей можно добиться хороших результатов в прогнозировании курса акций, прогноз будет гораздо точнее отражать поведение временного ряда, чем при использовании таких методов прогнозирования, как тренды и ARIMA.

