Национальный исследовательский университет ИТМО
Факультет информационных технологий и программирования
Прикладная математика

Регрессия временных рядов

Отчет по лабораторной работе №4

Работу выполнили:
Обиджанов Алишер
Казаков Андрей
Кузнецов Павел

Преподаватель:
Свинцов М.В

Санкт-Петербург
2024

1 Построение авторегрессионной модели

Авторегрессионная (AR- модель) - модель временного ряда, в которой значения ряда зависят от предыдущих его значений. В нашем случае мы рассматриваем AR-3 модель такого вида:

$$x_t = a_0 + \sum_{i=1}^{3} a_i x_{t-i} + \epsilon_t$$

Где a_i - коэффициенты (параметры) модели, ϵ_t - белый шум.

Нам необходимо рассмотреть стационарную модель, условие стационарности - все корни характеристического уравнения лежат вне единичной окружности. Характеристическое уравнение имеет вид:

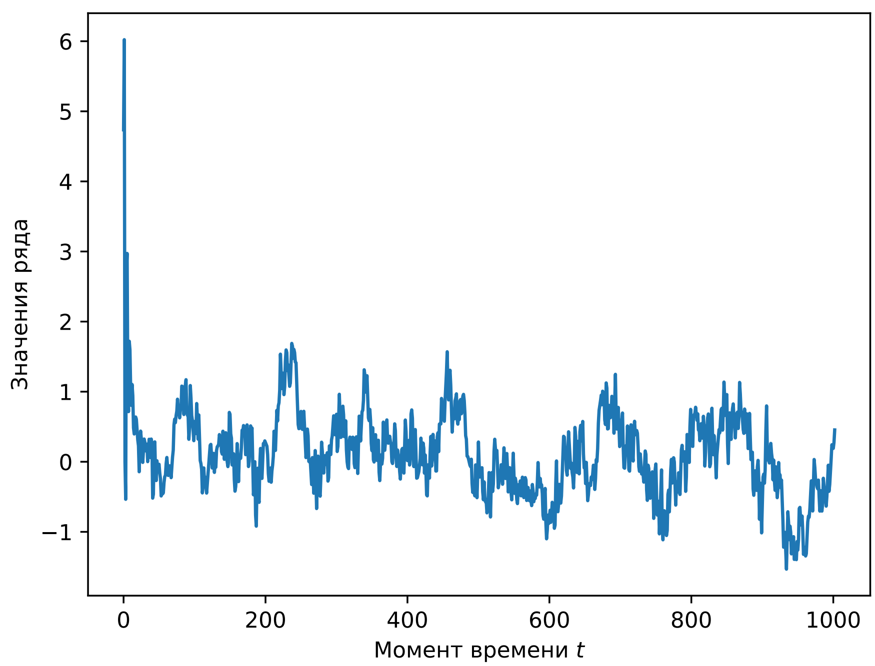
$$1 - \sum_{i=1}^{3} a_i z^i = 0$$

2 Демонстрация регрессии ряда

Воспользовавшись генератором авторегрессионных моделей, мы получили пример следующего ряда:

$$x_t = 0.326 + 0.938x_{t-1} - 0.505x_{t-2} + 0.521x_{t-3} + \epsilon_t$$

График первых 1003 значений ряда с первыми тремя значениями $t_0 = 4.736, t_1 = 6.022, t_2 = -0.006$:



Как можно наблюдать из графика, ряд быстро сходится к некоторому "равновесию"вокруг значения 0, то есть ряд обладает свойством стационарности. Колебания зависят сугубо от заданной дисперсии.

3 Обучение модели

Построение матрицы лагов - метод подготовки данных для прогнозирования временных рядов. Этот метод основан на идее использования предыдущих значений временного ряда для предсказания будущих.

Пусть X - матрица входных данных: Каждая строка матрицы представляет собой вектор задержек (лагов) длиной `lag_order`. То есть, каждая строка содержит последовательность предыдущих значений временного ряда.

y - вектор выходных данных. Каждый элемент вектора соответствует значению временного ряда, следующему сразу после соответствующего вектора задержек в матрице X.

Пример: Пусть у нас есть временной ряд [1,2,3,4,5] и `lag_order` = 2. Тогда матрица X и вектор y будут следующими:

$$X = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix}$$

$$y = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$$

Далее инициализируется и обучается модель машины опорных векторов (SVM) с использованием линейного ядра. После обучения модели производится предсказание на тестовой выборке. Также оценивается качество модели с использованием среднеквадратичной ошибки

4 График

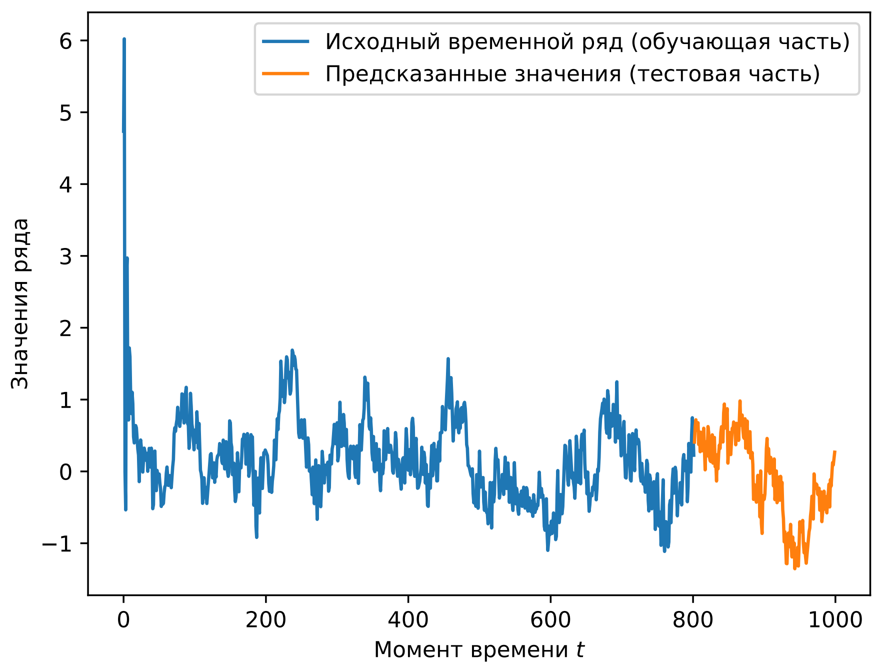


Рис. 1:

Можно заметить, что предсказанные значения тесно соответствуют исходному временному ряду. Однако дисперсия предсказанных значений оказывается немного меньше, чем у исходного временного ряда. Это может указывать на то, что модель в некоторой степени сглаживает колебания и шум временного ряда.

5 Эксперименты с различными ядрами

Эксперименты с различными ядрами и гиперпараметрами будем выполнять в цикле, где перебираются различные комбинации ядер(linear, rbf, poly), значений параметра штрафа C и степени полинома (если используется полиномиальное ядро). Для каждой комбинации параметров обучается модель машины опорных векторов (SVM), и оценивается её качество на тестовой выборке с использованием среднеквадратичной ошибки (MSE).

Ядра и гиперпараметры:
Kernels = ['linear', 'rbf', 'poly']
C values = [0.1, 1, 10]
Degree values = [2, 3, 4]

Результаты:
Best options: 'kernel': 'linear', 'C': 1, 'degree': None
Mean Squared Error for Best Parameters: 0.044743342443301685

Результаты говорят о том, что линейное ядро с высоким значением параметра штрафа C демонстрирует лучшую производительность в данном контексте, и позволяет сделать вывод о том, что выбранная модель хорошо адаптирована к характеру временного ряда.

6 Вывод:

Лабораторная работа позволила успешно построить авторегрессионную модель AR(3) для стационарного временного ряда. Модель, основанная на машине опорных векторов с линейным ядром, продемонстрировала хорошую аппроксимацию и точность прогнозов. Эксперименты с различными ядрами подтвердили эффективность линейного ядра при высоком значении параметра штрафа C