

Описание

ARIMAX (AutoRegressive Integrated Moving Average eXtended) — это математическая модель для анализа временных рядов, объединяющая в себе интегрированную авторегрессию, скользящее среднее и возможность учета дополнительных внешних факторов.

Модели ARIMA применяются для решения задач, в которых требуется построить прогноз на основе имеющихся данных, то есть вычислить последующие значения ряда на основе предыдущих. Временным рядом могут быть любые данные в разрезе времени, например, продажи товаров, количество заказов, поток клиентов и т.д.

Важно: для получения прогнозируемых данных требуется предварительное обучение узла.

Порты

Вход

-  **Входной источник данных** — таблица данных. Входные данные должны соответствовать следующим требованиям:
 - Поля, которые будут использоваться в качестве входных или выходных, не должны содержать пропущенные значения. Если это требование не выполнено, то в момент активации узла будет выдана ошибка.
 - Поле, соответствующее временному ряду, должно иметь назначение *Прогнозируемое*, тип данных *Вещественный* и вид данных *Непрерывный*. Такое поле должно быть только одно.
 - Поля, соответствующие дополнительным факторам, должны иметь назначение *Входное*. Ограничений на тип данных в этих полях нет, вид данных может быть любой, кроме *Неопределённого*. Данные поля могут отсутствовать или присутствовать в любом количестве.

Выход

-  **Выход модели** — таблица данных, которая содержит следующие поля:
 - Имя_поля|Прогноз — прогнозируемые значения исходного временного ряда.
 - Имя_поля|Ошибка аппроксимации — остатки модели, отклонения между прогнозируемыми и фактическими значениями ряда. Поле присутствует, если установлен флаг *Рассчитать ошибку аппроксимации*.

- Имя_поля|Нижняя граница — нижняя граница доверительного интервала. Поле присутствует, если установлен флаг *Рассчитать доверительный интервал*.
 - Имя_поля|Верхняя граница — верхняя граница доверительного интервала. Поле присутствует, если установлен флаг *Рассчитать доверительный интервал*.
-  **Коэффициенты модели** — таблица данных, которая содержит следующие коэффициенты:
 - Тип;
 - Параметр;
 - Лаг;
 - Имя входного поля;
 - Уникальное значение;
 - Коэффициент;
 - Стандартное отклонение;
 - T-статистика;
 - P-value.
-  **Сводка** — переменные:
 - Всего примеров (TotalSamples);
 - Всего отобранных примеров (TotalSelectedSamples);
 - Примеров в обучающем множестве (TrainSamples);
 - Среднеквадратическая ошибка на обучающем множестве (TrainRMSError);
 - Средняя абсолютная ошибка на обучающем множестве (TrainAvgError);
 - Средняя относительная ошибка на обучающем множестве (TrainAvgRelError);
 - Информационный критерий Акаике (AIC);
 - Информационный критерий Акаике скорректированный (AICc);
 - Информационный критерий Байеса (BIC);
 - Коэффициент детерминации (R2);
 - Скорректированный коэффициент детерминации (AdjustedR2);
 - Число степеней свободы модели (ModelDF);
 - Число степеней свободы остатков (ResDF).

Мастер настройки

Шаг 1. Настройка входных столбцов

На первом этапе необходимо задать назначение столбцов входного набора данных. Для каждого из столбцов можно выбрать один из вариантов назначения:

-  **Прогнозируемое** — для данных, соответствующих временному ряду.
-  **Входное** — для данных, соответствующих дополнительным входным факторам.
-  **Не задано** — для данных, не участвующих в построении модели. Устанавливается по умолчанию для остальных столбцов.

Шаг 2. Настройки нормализации

Для моделей ARIMA нормализация для прогнозируемых данных обычно не требуется. Рекомендуется не применять нормализацию для данных временного ряда, для данных внешних факторов не менять настройки по умолчанию.

Шаг 3. Настройки ARIMAX

Структура модели ARIMAX

- **Определить структуру автоматически** — при установке данного флага активируется автоподбор параметров модели. Параметры в процессе вычисления подбираются таким образом, чтобы минимизировать значение AIC.
- **Порядок AR части** — задает порядок (p) авторегрессионной части. Определяет число предыдущих значений ряда, учитываемых при построении модели. Устанавливается значение целого типа больше 0.
- **Порядок интегрирования** — задает порядок (d) разностей ряда при необходимости привести исходный ряд к стационарному. Устанавливается значение целого типа больше 0.
- **Порядок MA части** — задает порядок (q) части скользящего среднего. Определяет размер скользящего окна для сглаживания исходного ряда. Устанавливается значение целого типа больше 0.
- **Включить расчет сезонности** — установка данного флага позволяет задать параметры для сезонной составляющей:
 - **Порядок сезонной AR части** — устанавливает значение целого типа от 0 и выше.
 - **Порядок сезонного интегрирования** — устанавливает значение целого типа от 0 и выше.
 - **Порядок сезонной MA части** — устанавливает значение целого типа от 0 и выше.
 - **Период сезонной составляющей** — устанавливает положительное значение целого типа.
- **Включить константу в модель** — значение логического типа. По умолчанию включено.

Прогнозирование временных рядов

- **Горизонт прогноза** — задает количество значений, которые будут спрогнозированы и добавлены в выходной набор в конце исходного временного ряда. Устанавливает значение целого типа ≥ 1 .
- **Рассчитать ошибку аппроксимации** — флаг, установка которого добавляет в выходной набор столбец со средними отклонениями прогнозируемых значений от фактических.
- **Рассчитать доверительный интервал** — позволяет вручную задать настройку следующего параметра:
 - **Уровень доверия прогноза, %** — значение вещественного типа 0 до 100, по умолчанию 95.

Важно: для модели ARIMAX есть ограничение - построение прогноза с учетом внешних факторов возможно только на 1 период вперед. Для построения прогноза на 2, 3 и более периодов внешние факторы не учитываются, т.е. используется модель ARIMA.

