Вики

Модель авторегрессии скользящего среднего (ARIMA)

Синонимы: Модель Бокса-Дженкинса, АРИСС, Методология Бокса-Дженкинса, Autoregressive Integrated Moving Average, Box-Jenkins model
Разделы: Алгоритмы

Модель авторегрессии скользящего среднего и систематический подход к ее построению был предложен в 1970-х годах Джорджем Боксом и Гвилимом Дженкинсом. Она предназначена для анализа стационарных временных рядов на основе оценки линейной зависимости прогнозируемых значений от исторических.

Для использования модели временной ряд должен быть стационарным, т.е. его среднее и дисперсия должны быть постоянны.

Модель Бокса-Дженкинса предполагает, что временной ряд содержит три составляющие: авторегресионную, интегрированную и скользящее среднее, которые в модели обозначены p, d и q соответственно:

- Величина p называется порядком авторегрессии. Она позволяет ответить на вопрос, будет ли очередной элемент ряда близок к значению X, если к нему были близки p предыдущих значений.
- Величину d называют порядком интегрирования. Она показывает, насколько элемент ряда близок по значению к d предыдущим значениям, если разность между ними минимальна.
- Параметр q — порядок скользящего среднего. Позволяет установить погрешность модели как линейную комбинацию наблюдавшихся ранее значений ошибок.

Авторегрессия — это составляющая модели временного ряда, в которой его прогнозируемое значение может быть выражено в виде линейной комбинации исторических значений этого же ряда и случайной ошибки.

Обычно модель упоминается, как $ARIMA(p, d, q)$, где p, d и q — целые неотрицательные числа, характеризующие порядок для частей модели (соответственно авторегрессионной, интегрированной и скользящего среднего).

Для временного ряда $X(t)$ модель может быть записана в виде:

$$(\Delta^d X_t) = \sum_{t=1}^{p} a_t \left(\Delta^d X_{t-1}\right) + \varepsilon_t + \sum_{j=1}^{q} b_j \left(\Delta^d \varepsilon_{t-j}\right),$$

где

- Δ^d — оператор разности порядка d (последовательное взятие d раз разностей первого порядка — сначала от самого ряда, затем от полученных разностей первого порядка, затем от второго порядка и т. д.);
- a_t — коэффициенты авторегрессионной части модели, $\varepsilon(t)$ — значения ошибки (полагаются независимыми одинаково распределенными случайными величинами из нормального распределения с нулевым средним);
- b_j — коэффициенты скользящего среднего.

Существует также расширенная модель ARIMAX, которая учитывает внешние факторы при построении прогноза.

Модель Бокса-Дженкинса широко применяются при прогнозировании временных рядов. Основная задача при этом заключается в оценке параметров модели. Методология построения ARIMA-модели исследуемого временного ряда включает следующие основные этапы:

- построение пробной модели;
- оценивание параметров модели и проверка адекватности модели;
- использование модели для прогнозирования.

В Loginom модель ARIMA реализована как отдельный обработчик, который позволяет производить прогнозирование временных рядов. Временным рядом могут быть любые данные в разрезе времени, например, продажи товаров, количество заказов, поток клиентов и т.д.

Построение модели в Loginom для прогноза объема продаж сезонных товаров зимнего спорта по месяцам подробно описано в статье «Пример использования ARIMA для прогноза продаж».