9/3/2023

Assignment 1

Performance Evaluation of AI and ML Models

MUHAMMAD SHAAF
STE-AI

ARIMA (AutoRegressive Integrated Moving Average)

Data Type

For time series data, ARIMA is employed. A time series is a collection of data points that are gathered or recorded over a regular time period. It frequently contains timestamps, making it by nature time-ordered.

Task

The ARIMA model's main function is time series forecasting. This entails foreseeing future data points in a time series based on trends and connections seen in the past.

Applications

Forecasting financial data

ARIMA models are used to forecast stock values based on past price information. These forecasts are used by traders and investors to make decisions.

Forecasting demand

Retailers use ARIMA to estimate product demand, optimise inventory levels, and schedule production. Using ARIMA, businesses may forecast demand changes along the supply chain, facilitating effective logistics and inventory control.

Economic projections

GDP Growth Prediction: The ARIMA model is used by economists and decision-makers to forecast GDP growth rates, inflation, and other economic indicators.

Forecasting the unemployment rate: ARIMA models can be used to forecast the unemployment rate and examine trends in the labour market.

Forecasting energy

Utility firms use ARIMA to forecast electricity consumption, which helps with the planning of power generation.

Renewable Energy Production

In order to ensure effective grid operation, ARIMA models are used to anticipate renewable energy production, such as solar or wind power.

Forecasting the climate and the weather

Forecasting of Temperature and Precipitation: ARIMA models can be used to forecast temperature and precipitation patterns, which is helpful for weather forecasts and climatological research.

Marketing and Sales

Sales forecasting: ARIMA models assist companies in forecasting future sales trends, enabling efficient resource allocation and marketing. Retailers utilise ARIMA to analyse consumer behaviour, including churn prediction and purchasing trends.

Healthcare

ARIMA can be used to forecast disease outbreaks based on epidemiological trends and historical health information. In order to estimate patient admission rates and allocate resources appropriately, hospitals use ARIMA.

Traffic Volume Forecasting

To aid in traffic management and infrastructure development, transportation authorities employ ARIMA to forecast traffic volume patterns.

Advertising and marketing

Ad Campaign Performance Prediction: ARIMA uses historical campaign data to forecast the performance of advertising campaigns.

Manufacturing and quality assurance

Based on previous data, ARIMA models may foresee manufacturing process problems, assisting in the maintenance of product quality. Manufacturing companies utilise ARIMA to forecast production numbers and create production schedules.

Training and Validation

Time series data from the past are used to train ARIMA models. In the training phase, the correct model order, denoted by (p, d, q), is determined.

AutoRegressive Order

The value of the autoregressive order parameter, p, controls how many lag observations should be included in the model. It illustrates the connection between the present value and earlier values.

Integration Order

Integration is a differencing technique used to make the time series stationary, that is, with a constant mean and variance. During training, the order of differencing (d) is selected.

Moving Average Order

The value of the parameter q (Moving Average Order) defines how many lag forecast mistakes should be included in the model. It illustrates the connection between the present value and previous forecast blunders.

Statistical methods like the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), which balance model complexity and goodness of fit, are frequently used to select model orders (p, d, and q).

Validation

Out-of-sample testing is used to validate ARIMA models. The forecasts of the model are contrasted with the actual values using a piece of the time series data designated as a validation set. Common evaluation metrics are as follows:

- The mean of the absolute discrepancies between the expected and actual values is known as the mean absolute error (MAE).
- The average of the squared discrepancies between projected and actual values is called the mean squared error (MSE).
- Root Mean Squared Error (RMSE): The square root of Mean Square Error, which is frequently chosen since it more severely penalises significant errors.

Categorizing the Domain

ARIMA is categorized as a supervised machine learning model rather than a deep learning or knowledge-driven model.

Model Complexity

There are only a few parameters (p, d, and q) that need to be determined for ARIMA models, making them rather straightforward. Deep learning models, in comparison, have intricate neural structures with numerous layers and parameters. While ARIMA is built for more straightforward time series patterns, deep learning models may capture complicated patterns and representations in data due to their complexity.

Data Requirements

Usually, univariate time series data with a single dependent variable and a temporal component are subject to ARIMA models. In contrast, deep learning models may work with a variety of data kinds, including multimodal, unstructured, and structured data. For training, deep learning models like neural networks need a large amount of labelled data, whereas ARIMA mostly uses historical time series data.

Model Interpretability

The components of ARIMA models (such as the autoregressive and moving average terms) and how they affect future forecasts are interpreted in a reasonably simple manner. Due to their intricate designs, deep learning models, particularly deep neural networks, are frequently referred to as "black boxes" since it is difficult to understand how they function.

Training Procedure

ARIMA models rely on iterative algorithms to determine the right model orders (p, d, and q) and statistical techniques like maximum likelihood estimation to estimate model parameters. Deep learning models employ backpropagation and gradient-based optimisation methods, which are unique to the discipline of deep learning.

Feature Engineering

Feature engineering in the context of ARIMA primarily involves preprocessing and transformation of the time series data to make it suitable for modeling. While ARIMA doesn't involve creating explicit features as in traditional machine learning, it does require several crucial steps to prepare the data for forecasting.

Differencing (Integration Order, d)

A key feature engineering step in ARIMA is differencing, which makes the time series stationary. The term "stationarity" refers to a time series' statistical characteristics, such as its mean and variance, being constant across time. For ARIMA models, stationarity is a fundamental presumption.

The number of times the data must be differentiated in order to reach stationarity is indicated by the letter "d" in the differencing order. First-order differencing is the method used to compare two successive observations, and it is continued until the data become stationary.

Mathematically, differencing at order 'd' is defined as: **Y(t) - Y(t-d)**

By removing trends and seasonality, differencing renders the time series stationary. ARIMA can assume that the statistical characteristics of the data remain consistent throughout time thanks to stationarity, which makes the modelling process easier. For predicting to be accurate, this premise must be true.

The autocorrelation between data is frequently reduced through differencing, which makes it simpler for ARIMA to represent the link between the current observation and previous values.

Seasonal Decomposition

Seasonal decomposition is a method that can be used with time series data that exhibit seasonal trends. The time series is divided into three key categories in this step: trend, seasonality, and residual (error).

The long-term direction or trend of the data is represented by the trend component. The seasonal component, such as daily, weekly, or yearly seasonality, captures the repeating patterns over a given period. The data's noise or unpredictability is represented by the residual component.

By removing seasonality from the data, seasonal decomposition makes it simpler for ARIMA to identify the remaining patterns. In order to better properly predict the underlying non-seasonal patterns, ARIMA divides the data into trend, seasonal, and residual components.

Autoregressive (AR) and Moving Average (MA) Components (Order p and q)

The orders "p" and "q" in the ARIMA model stand for the autoregressive (AR) and moving average (MA) components, respectively.

AR Component (p): The link between the present observation and earlier observations made at lags of 'p' is modelled by this component. When choosing a model, the 'p' value is established using autocorrelation plots (ACF).

The MA Component (q): The link between the current observation and the historical forecast mistakes (residuals) at lags of 'q' is modelled by this component. The partial autocorrelation plots (PACF) provide another source of information for the 'q' value.

These elements are necessary for identifying the data's time-dependent patterns. Regardless of whether they are autoregressive (AR) or moving average (MA) in nature, ARIMA may accurately identify and describe the distinct time-dependent patterns in the data by selecting the proper orders. Any serial correlations in the data that were not completely eliminated by differencing are taken into consideration by ARIMA with the aid of the MA component. To forecast accurately, this is crucial.

Model Orders Selection

A critical stage in developing the ARIMA model is deciding on the proper orders (p, d, and q). To choose the orders that best characterise the data, it entails examining ACF and PACF plots, running statistical tests, and employing information criterion like AIC and BIC.

As the depth of differencing ('d') and the number of delays ('p' and 'q') employed in the model are determined by the choice of orders, it has an impact on the feature engineering process. The underlying data patterns are captured by ARIMA without overfitting thanks to the careful choice of model orders ('p,' 'd,' and 'q'). Forecasts may be erroneous as a result of overfitting.

Data Scaling and Transformation

You may use scaling or transformation techniques to make sure that values are inside a specific range depending on the data and its scale. The behaviour of the model may be improved as a result. When dealing with data that contains extreme values or non-linear growth patterns, data scaling or transformation helps stabilise the behaviour of ARIMA. The model can provide more accurate forecasts with stable data.

For data exhibiting exponential growth patterns, common transformations include logarithmic or Box-Cox transformations.