Hautes Études des Technologies de l'Information et de la Communication

Réalisation du modèle ARIMA en Python

Cahier des charges et rapport

HETIC

Réalisé par:

-ABBACHE Aghiles.
-BOUZBID Dalil.
-TAHAR Ahmed Amar.

Spécialité : Master Data & Ia
Année universitaire : 2022/2023
Module : Fondements mathématiques et IA
Semestre : 02
Professeur : Monsieur HORRAIRY Hakim

Êtes-vous fatigué de dépendre de bibliothèques pour ARIMA et les prévisions ? Avez-vous déjà envisagé de l'écrire à partir de zéro en utilisant NumPy et Pandas pour obtenir une meilleure compréhension ? Ne cherchez plus, car voici la solution qui vous donnera le contrôle ultime sur vos données.

ARIMA (Auto Regressive Integrated Moving Average) est un modèle incroyablement puissant utilisé pour les prévisions de séries temporelles. Les trois principales composantes de l'ARIMA sont la stationnarité des données, l'autorégression (AR) et la moyenne mobile (MA). En commençant par différencier les données pour les rendre stationnaires, nous pouvons ensuite faire des estimations à l'aide de la régression automatique, utiliser la moyenne mobile sur les erreurs générées, dé-différencier le résultat et vérifier sa précision.

Ne vous contentez pas d'utiliser les bibliothèques existantes ; prenez les choses en main et développez une compréhension plus approfondie en créant à partir de zéro ! Plongeons ensemble dans l'exploration de tous les aspects de cette méthode de modélisation complexe.

Après une analyse approfondie, il est évident qu'il existe un modèle discernable de tendance et de saisonnalité dans les données. La tendance générale à l'augmentation ou à la diminution peut être observée comme récurrente chaque année.

Partie 1 : Rendre les données stationnaires

Il existe plusieurs méthodes pour rendre les données stationnaires, notamment des techniques logarithmiques et de différenciation. Une méthode antérieure de soustraction des valeurs n'a pas permis d'obtenir la stationnarité en raison d'une tendance saisonnière ; les données ont donc été décalées de 12 unités et soustraites à nouveau.

Partie 2 : Modèle autorégressif

$$y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots\dots\dots\dots + \varphi_p y_{t-p} + \varepsilon_t$$

Lors de l'analyse d'une valeur à l'instant t, nous supposons qu'elle dépend des p valeurs décalées précédentes de manière linéaire et qu'un terme d'erreur lui est associé. Cette méthode présente des similitudes avec la régression linéaire où X représente les p valeurs décalées tandis que y représente la valeur au moment t. L'ordre de p peut être déterminé à l'aide des graphiques ACF et PACF ou considéré comme un hyper-paramètre si l'on dispose de suffisamment de données.

Partie 3 : Moyenne mobile

Maintenant que nous avons généré les coefficients et l'ordonnée à l'origine, nous pouvons obtenir nos prédictions. La différence d'erreur entre la valeur réelle et la valeur prédite constituera nos résidus.

Comme pour p, q est utilisé ici pour désigner le nombre d'observations décalées. Là encore, on peut considérer qu'il s'agit d'une régression linéaire avec q erreurs décalées comme X et l'erreur comme y.

Nous combinons maintenant les deux modèles AR et MA pour obtenir nos prévision:

$$ARMA(p,q) : Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t$$