Introduction aux séries temporelles

UQAC | Université du Québec à Chicoutimi

Définition

Les séries temporelles (*time series*) constituent une forme importante de données structurées dans de nombreux domaines différents, tels que la finance, l'économie, l'écologie, la neurosciences et la physique. Tout ce qui est enregistré de manière répétée à de nombreux moments dans le temps forme une série temporelle. De nombreuses séries temporelles ont une fréquence fixe, que nous avons une observation à des intervalles réguliers selon une certaine règle, comme toutes les 15 secondes, toutes les 5 minutes ou une fois par mois. Les séries temporelles peuvent également être irrégulières sans unité de temps fixe. La façon dont les données de séries temporelles sont organisées dépend de l'application, et on voit souvent les 3 cas de figures suivants :

- *Timestamps*: instants spécifiques dans le temps
- Intervalle de temps: Indiqué par un *timestamps* de début et de fin.
- Périodes fixes: par mois, année ou trimestre. Les périodes fixes peuvent être considérées comme des cas spéciaux d'intervalles de temps.

UQAC

Rééchantillonnage (*Resampling*)

Le rééchantillonnage dans les séries temporelles consiste à ajuster les données à une fréquence différente. Il existe deux types de rééchantillonnage:

- *down-sampling* : agrégation ou sélection des données à une fréquence plus basse (par exemple, de données quotidiennes à mensuelles)
- *up-sampling*: interpolation des données à une fréquence plus élevée (par exemple, de données mensuelles à quotidiennes).

Chevauchement (*Time-shift*)

Une opération très commune sur des séries temporelle est le chevauchement (*time-shift*). Cela est utile quand on veut voir la différence entre des valeurs pour une variable données pour au même moment mais à une autre période. Par exemple, si on veut comparer nos ventes totales obtenue jusqu'à maintenant à celle qu'on avait à pareil date au mois dernier (ou l'année dernière !).

Chevauchement (*Time-shift*)

Exemple

Un exemple concret d'utilisation du time-shift est le calcul du retour sur investissement (ROI) de l'achat d'une action.

$$\text{ROI} = \left(\frac{\text{Gain financier}}{\text{Coût de l'investissement initial}} \right) \times 100\%$$

Calculons le ROI sur 1 an pour toutes les occurences du DataFrame et créons un graphique à ligne afin de voir l'évolution de celui-ci.

Fenêtre glissante (Rolling Window)

Un autre type d'opérations propre aux séries temporelles est la fenêtre glissante (rolling window). Cet opération est appliqué à un sous-ensemble continu de données qui se déplace progressivement à travers l'ensemble des données.

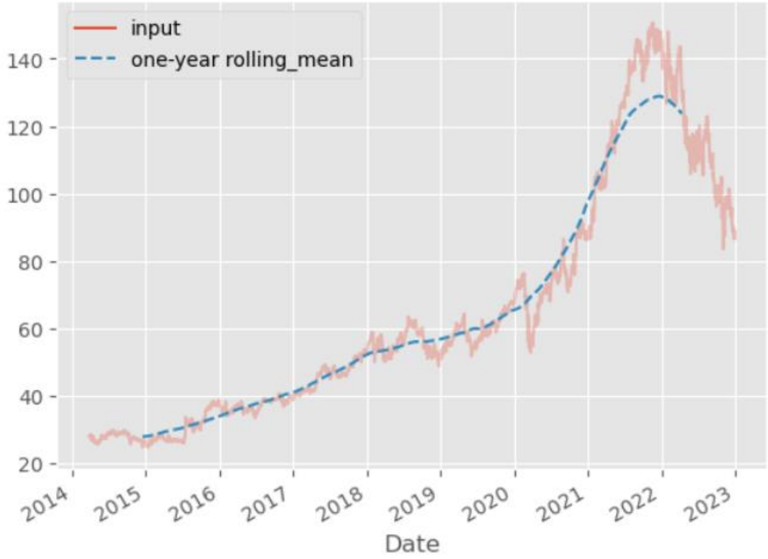
Pour mieux comprendre, imaginez une série temporelle qui représente des données sur une période de temps donnée, par exemple, les ventes quotidiennes d'un produit sur un an. Si vous utilisez une fenêtre glissante de 30 jours, l'algorithme effectuera une opération (comme la moyenne, la somme, ou tout autre calcul) sur les données des 30 premiers jours, puis glissera d'un jour, recalculera l'opération sur les 30 jours suivants, puis glissera à nouveau, et ainsi de suite, jusqu'à ce que toute la série temporelle soit parcourue.

L'application d'une fenêtre glissante peut aider à lisser les fluctuations dans les données, mettant en évidence des tendances à plus long terme.

Fenêtre glissante (Rolling Window)

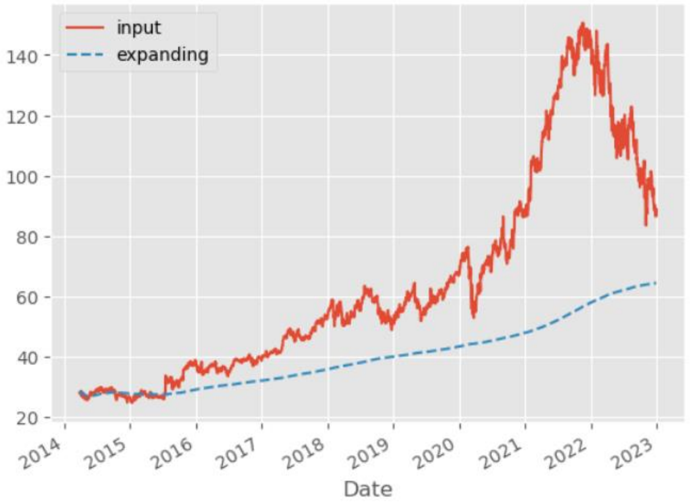
Exemple

Calculons des *rolling windows* qui font la moyenne annuelle des valeurs de l'action de Google.



Expanding window

Contrairement à la *rolling window*, où la taille de la fenêtre reste fixe et se déplace à travers les données, une *expanding window* augmente progressivement en taille au fur et à mesure qu'elle se déplace à travers les données jusqu'à englober les données au complet.



Expanding window

Une alternative à avoir le même poids pour chacune des valeurs que l'on a dans une fenêtre est d'avoir une fonction qui pondère les données pour mettre plus de poids sur les données récente. La méthode *ewm* permet de réaliser ce type de fenêtre sur les séries temporelles.

Voici le calcul qu'effectue cette fonction lorsqu'utilisé pour calculer une moyenne:

$$\mathrm{EMWA}_t = (1-\alpha) \times \mathrm{EMWA}_{t-1} + \alpha \times X_t$$

Où:

- EMWA_t est la moyenne mobile pondérée exponentiellement à l'instant t
- EMWA_{t-1} est la moyenne mobile pondérée exponentiellement à l'instant $t-1$.
- X_t est la valeur actuelle de la série temporelle à l'instant t
- α est le facteur de pondération Il existe plusieurs argument afin de spécifier comment on veut que α soit calculé.

Introduction à l'extraction de caractéristiques temporelles

Les caractéristiques temporelles d'une *time-serie* incluent divers aspects qui permettent de comprendre son comportement, son évolution et ses tendances au fil du temps. Voici quelques-unes des caractéristiques temporelles importantes d'une série temporelle :

- Tendance: la direction générale dans laquelle les données évoluent au fil du temps. Une tendance peut être ascendante, descendante ou horizontale (stationnaire).
- Saisonnalité: variations régulières ou périodiques dans les données qui se répètent à des intervalles fixes, comme quotidiennement, mensuellement ou annuellement. La saisonalité peut être causée par des facteurs tels que les saisons, les jours de la semaine, ou d'autres cycles réguliers.
- Résidus: fluctuations aléatoires ou irrégulières dans les données qui ne peuvent pas être expliquées par la tendance ou la saisonnalité.

Introduction à l'extraction de caractéristiques temporelles

Il existe plusieurs façons décomposer une série temporelle pour obtenier les caractéristiques énoncées plus haut. Dans ce cours-ci, je traiterai seulement des deux modèles suivants:

- décomposition additive
- décomposition multiplicative

Introduction à l'extraction de caractéristiques temporelles

Dans une décomposition additive, la série est représentée comme la **somme** de la tendance, de la saisonnalité et du résidu.

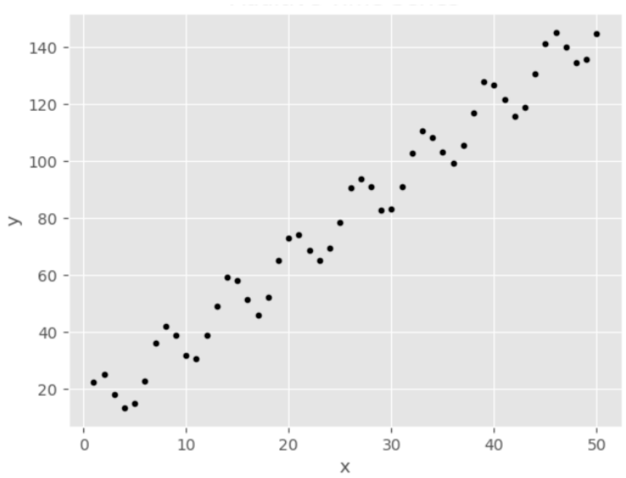
$$Y_t = T_t + S_t + R_t$$

Dans une décomposition additive, la série est représentée comme la **multiplication** de la tendance, de la saisonnalité et du résidu.

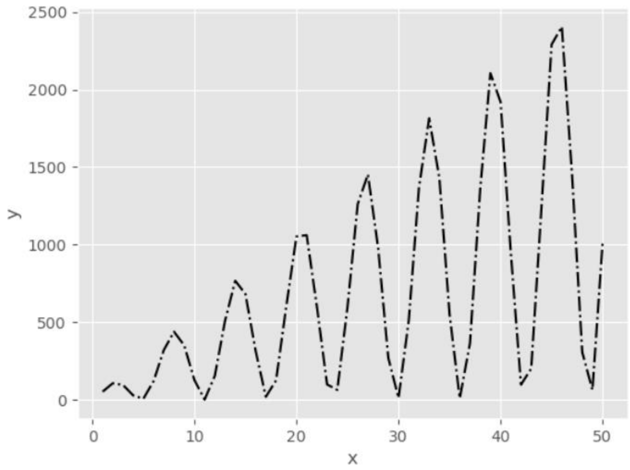
$$Y_t = T_t \times S_t \times R_t$$

Introduction à l'extraction de caractéristiques temporelles

La décomposition additive est souvent utilisée lorsque les variations saisonnières restent constantes en termes absolus tout au long de la série, tandis que la décomposition multiplicative est utilisée lorsque les variations saisonnières augmentent proportionnellement au fil du temps.



Additive



Multiplicative

Stationnarité

La stationnarité est un concept important en statistiques et en séries temporelles. Une série temporelle est dite stationnaire si ses propriétés statistiques, telles que la moyenne, la variance et la structure de corrélation, restent constantes au fil du temps. En d'autres termes, une série temporelle stationnaire ne présente pas de tendance à long terme ni de saisonnalité.

Stationnarité

Dickey-Fuller augmenté (ADF)

Est un test statistique utilisé pour vérifier la stationnarité

Hypothèses (simplifiées):

- H0: La série temporelle n'est pas stationnaire
- HA: La série temporelle est stationnaire

Stationnarité

Modification pour obtenir la stationnarité

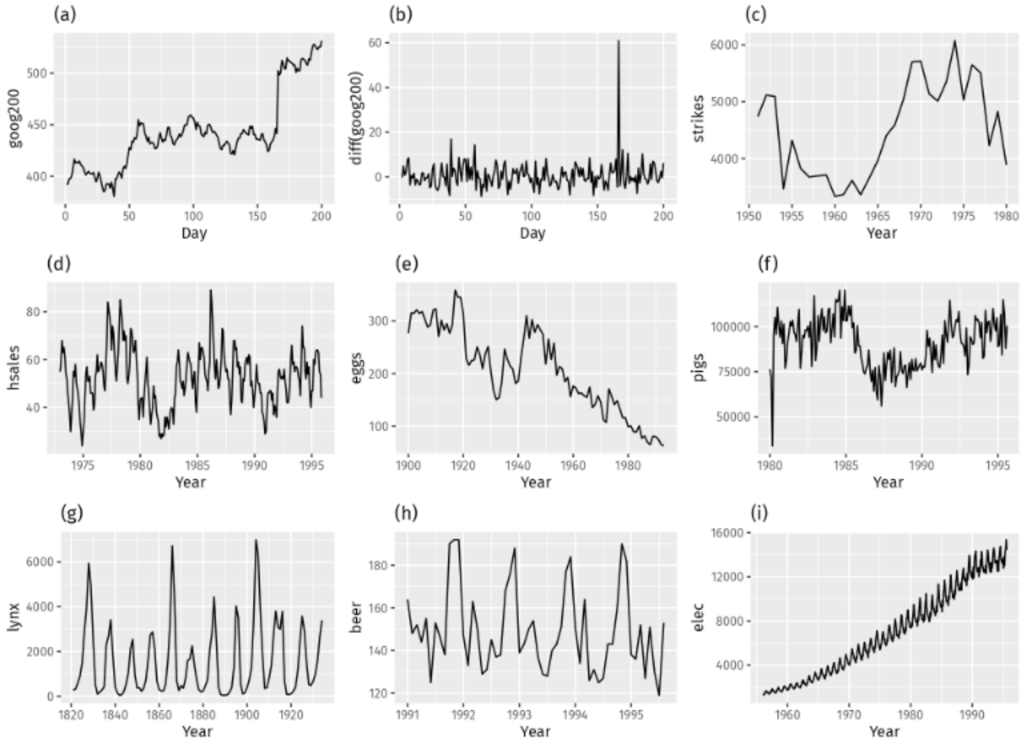
Différenciation : En prenant la différence entre des valeurs successives, on peut éliminer la tendance à long terme. (avec shift)

Transformation Logarithmique : Utile pour réduire les variations proportionnelles à un facteur constant.

Moyenne Mobile : En utilisant une moyenne mobile (rolling window avec aggrégation de moyenne) pour lisser la série.

Découpage en Sous-Périodes : Parfois, diviser la série en sous-périodes peut rendre les propriétés statistiques plus constantes.

Stationnarité



https://app.wooclap.com/ Code: ROHIAD

UQAC