

Cours PCD – Labo 8 : Lissage exponentiel et prévision

Objectifs

- Comparer trois méthodes de lissage exponentiel en les utilisant pour la prévision (*forecasting*).
- Tester différents paramètres et indiquer les meilleurs.
- Entraîner les modèles sur les 80% points de données initiaux, les tester sur les 20% finaux.
- Utilise RMSE et MAPE comme mesures d'évaluation.

Données

- Statistique économique australienne : on considère les dépenses mensuelles de la population lorsqu'elle mange dehors (au café, restaurant, *take-away*), en milliards de dollars.
- Ces données vous sont fournies au format CSV dans le fichier '**auscafe.csv**'.
- Source des données : Australian Bureau of Statistics, cat. n. 8501.0 ; utilisées par Rob J. Hyndman et George Athanasopoulos dans [Forecasting: Principles and Practice](#), et fournies dans le package R correspondant (variable globale 'auscafe' du [package fpp2](#)) ; converties par APB.

Indications

1. Utilisez le package Python **statsmodels** pour l'implémentation des méthodes de lissage exponentiel et surtout pour l'optimisation (entraînement) de leurs paramètres.
2. Prenez exemple sur le code fourni dans la documentation de ce package : https://www.statsmodels.org/stable/examples/notebooks/generated/exponential_smoothing.html
3. Commencez par lire les données du fichier dans un objet *pandas.Series* muni d'un index qui est une plage temporelle (*range*). Séparez les données d'entraînement et celles de test.
4. Entraînez (avec *.fit()*) un modèle de type « Simple Exponential Smoothing », et utilisez-le pour prédire les données de test. Que valent la RMSE et la MAPE ? Veuillez afficher sur un graphique la courbe des données d'origine, la courbe lissée par le modèle, et la prédiction.
5. Dans une deuxième étape, entraînez un ensemble de modèles de type « Holt », en demandant au système d'optimiser tous les paramètres (avec : `initialization_method="estimated"`). Veuillez comparer les performances (RMSE et MAPE) de quatre modèles correspondant aux valeurs True ou False des paramètres « exponential » et « damped_trend ». Quelle est la combinaison qui conduit au meilleur modèle ? Pour ce modèle, veuillez afficher la courbe des données d'origine, la courbe lissée par le modèle, et la prédiction.
6. *Optionnel* : même question que (5) en utilisant un modèle de Holt-Winters, implémenté dans la classe « ExponentialSmoothing ». Veuillez tester toutes les combinaisons des valeurs possibles des paramètres 'trend', 'seasonal' (valeurs 'add' ou 'mul') et 'damped_trend', 'use_boxcox' (valeurs True ou False). Quels sont les meilleurs scores RMSE et MAPE obtenus ?