



Sèries Temporals

Aina Palacios

Aina Palacios

- Enginyera de Telecomunicacions especialitzada en Audiovisuals
- Màster en Tecnologies Avançades especialitzada en deep learning en Multimèdia!
- Experiència en programació web i machine learning.
- Mentora a IT Academy de **Vuejs**



<https://www.linkedin.com/in/ainapc/>



ainaPali#2617

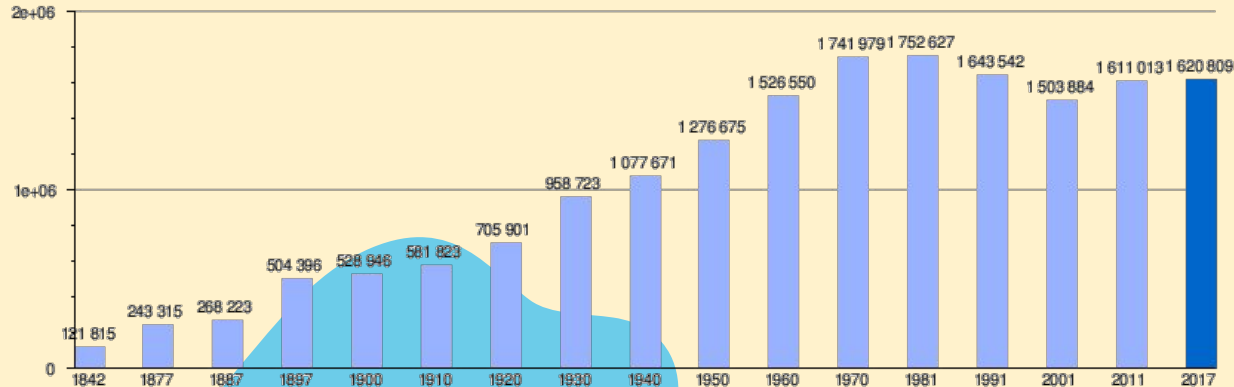
Sèrie temporal

Una col·lecció d'observacions d'una variable recollides seqüencialment en el temps.

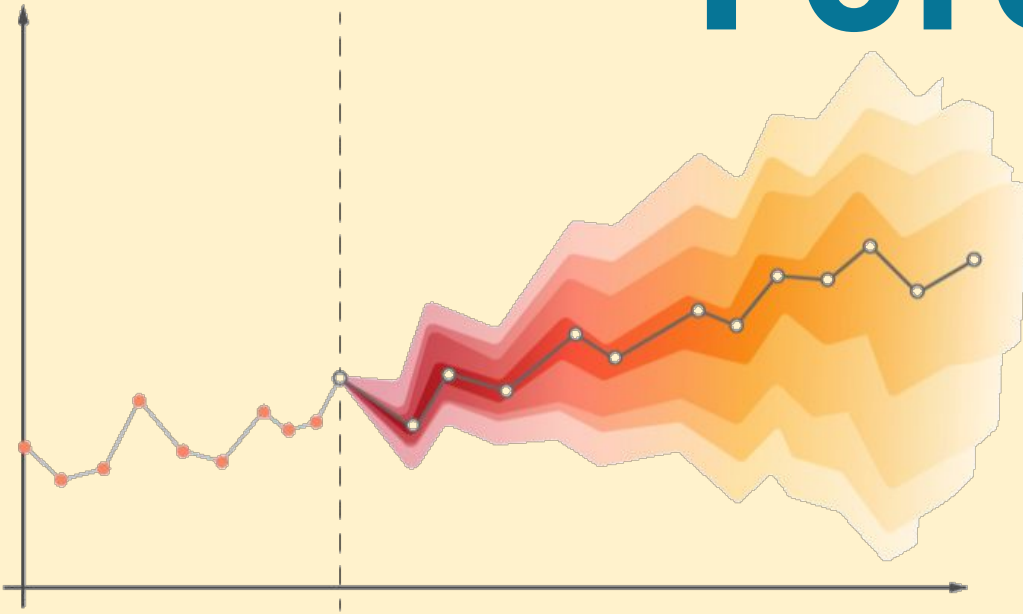
Important: La gràcia de les sèries temporals és que les mostres contínues no són independents entre si i s'ha d'analitzar tenint en compte l'ordre temporal.

Tipus: Deterministes i estocàstiques.

- Estocàstica: Es pot determinar de forma parcial amb elements del passat gràcies al fet que tenen una distribució de probabilitat condicionada als valors passats.



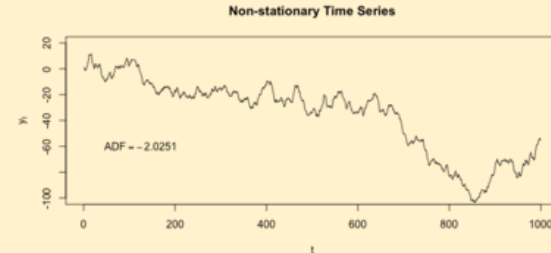
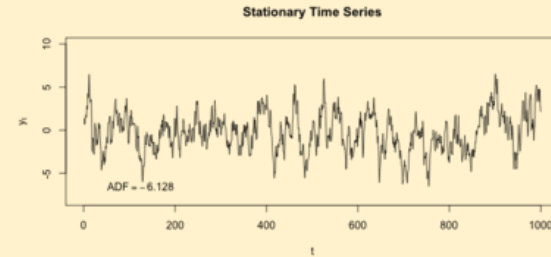
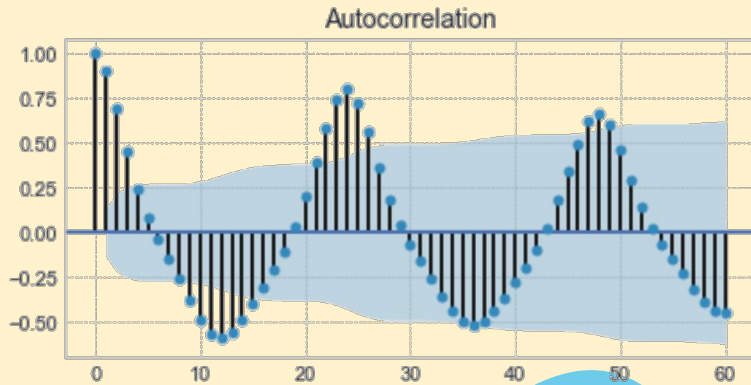
Temporal Series Forecasting



El nostre objectiu en Data Science i Machine Learning és fer prediccions científiques utilitzant dades històriques temporals! Així com que els nostres models siguin capaços de predir noves observacions.

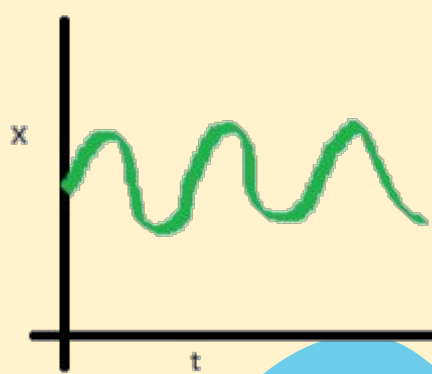
Característiques de les sèries temporals

- **Autocorrelació:** Grau de similitud d'una mateixa sèrie temporal amb una versió temporal diferent.
- **Efecte estacional:** Fluctuacions periòdiques. També es pot extreure de l'autocorrelació

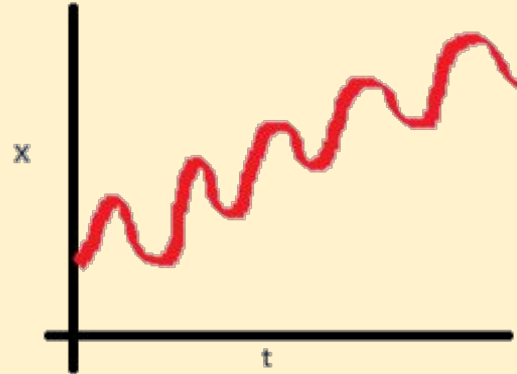


- **Sèries Estacionàries:** Són aquelles que no modifiquen la mitjana i la variància. -> Normalment les variables no són estacionàries, però ens interessa preprocessar-les per millorar les prediccions.
 - Podem aplicar el test de Dickey-Fuller per saber si són estacionàries.
- **Tendència:** Canvi a llarg termini de la mitjana
- **Component aleatòria:** Valors aleatoris no estacionaris o de tendència. Són els que necessitem predir
- **Anàlisi univariant i multivariant:** Predirem sobre la mateixa variable o utilitzarem altres variables en la predicció.

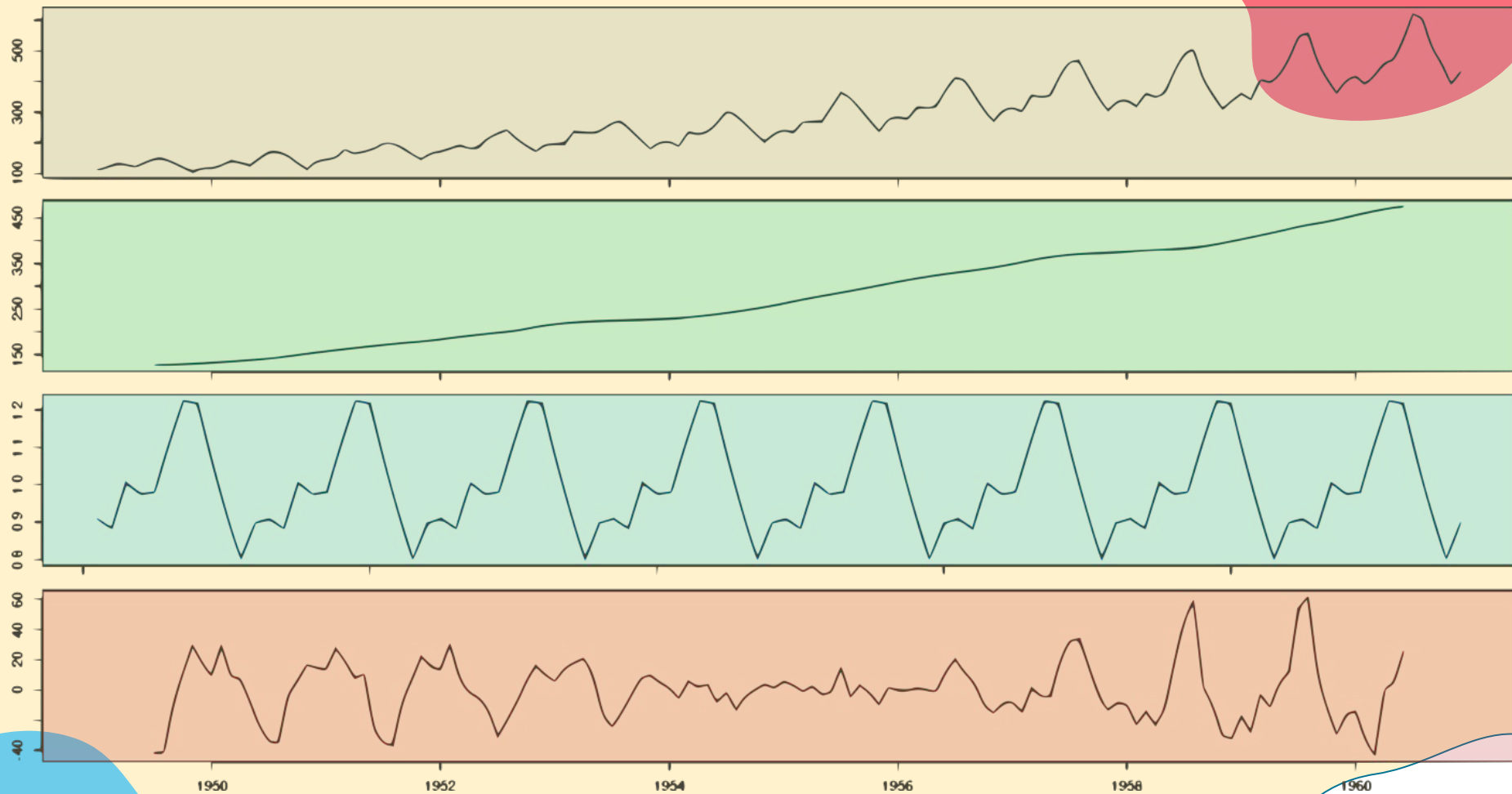
$$X = E + T + A$$



Stationary series



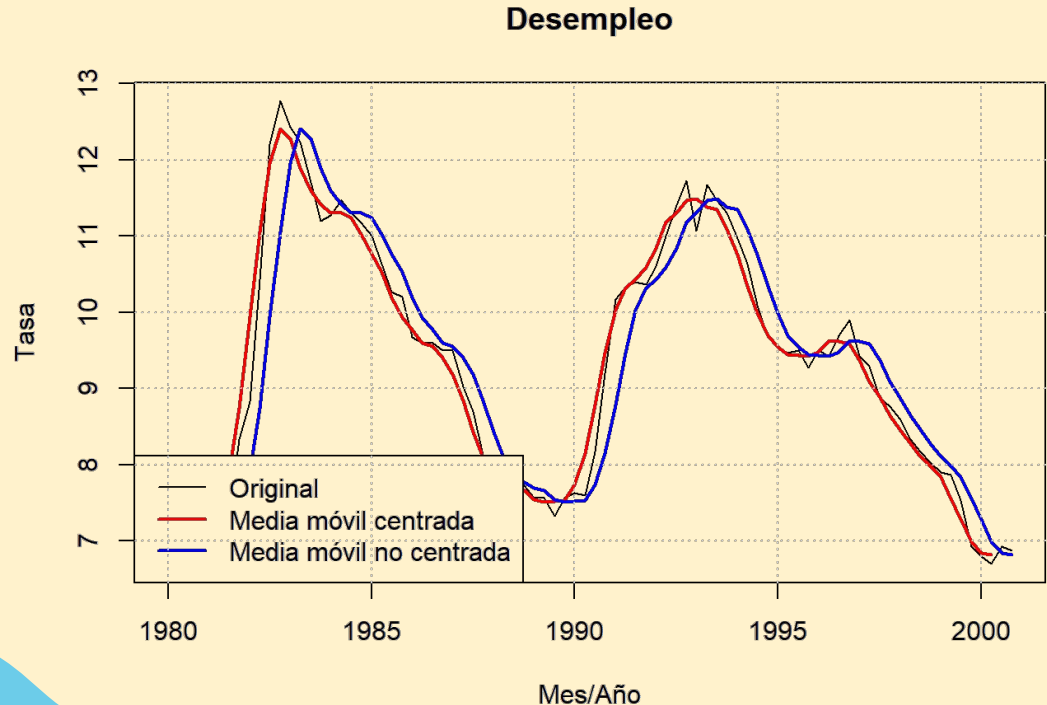
Non-Stationary series



Mètodes de suavitzat

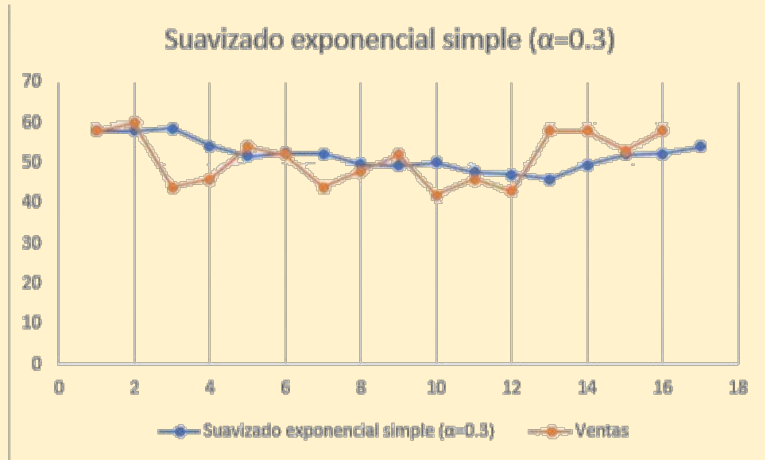
Per preparar les dades per ser predites, abans podem aplicar preprocessats que milloraran les nostres prediccions.

- **Mitges mòbils:** Les centrades per representar tendència i les aritmètiques per models amb mitja constant. Utilitza una part de les mostres petita, fa mitjana i suavitza la variable



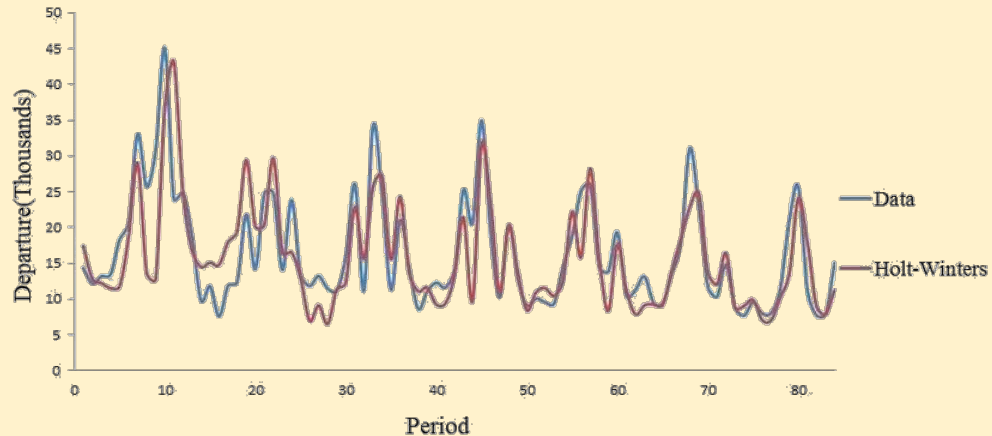
Mètodes de suavitzat

- **Allisat exponencial simple:** Allisa oscil·lacions i dóna més pes a les més properes en temps.

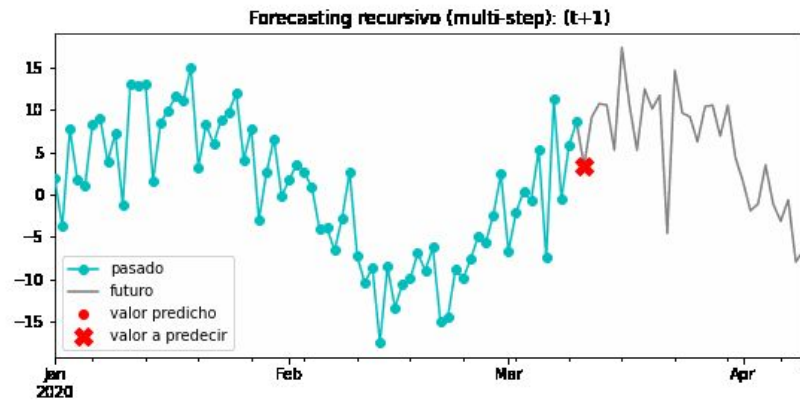


celeberrima.com

- **Mètode de Holt-Winters:** Suavitzat que té en compte l'estacionalitat.



Mètodes de predicció



| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|----|

Prediccions one-step

Quan ens interessa predir un element de la següent sèrie temporal.

Aquí podem utilitzar qualsevol model de regressió

Time series

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|----|

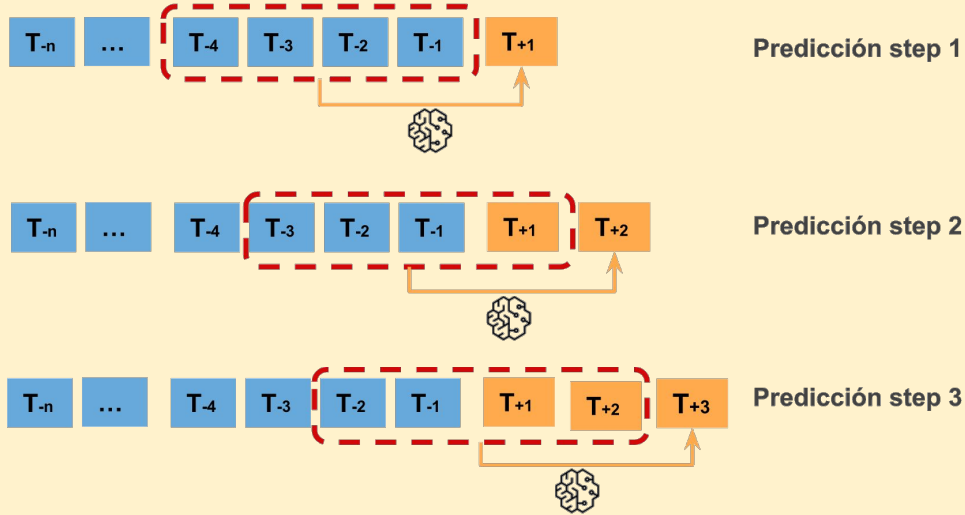
Exogenous variable

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| a | b | c | d | e | f | g | h | i | j |
|---|---|---|---|---|---|---|---|---|---|

| X | | | | | | y |
|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | f | 6 |
| 2 | 3 | 4 | 5 | 6 | g | 7 |
| 3 | 4 | 5 | 6 | 7 | h | 8 |
| 4 | 5 | 6 | 7 | 8 | i | 9 |
| 5 | 6 | 7 | 8 | 9 | j | 10 |

Recursive multi-step forecasting

Cada predicció fa ús de la predicció anterior

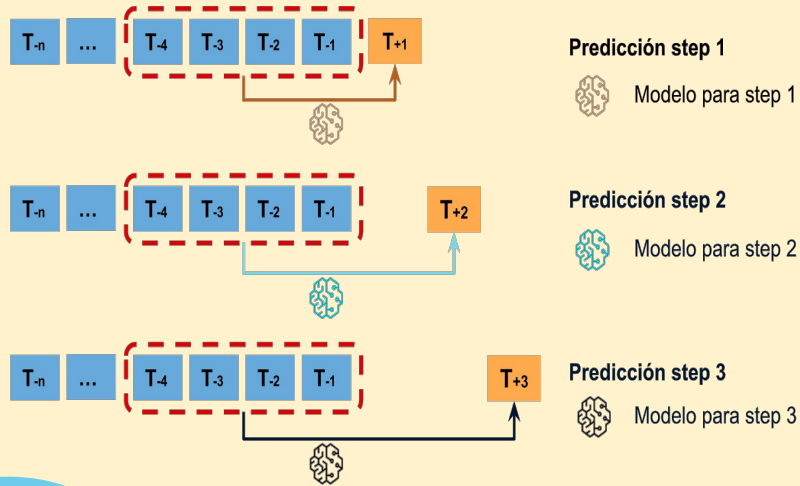


Prediccions multi-step

Normalment, ens interessarà predir sobre el temps i que tingui en compte les característiques de la sèrie temporal.

Direct multi-step forecasting

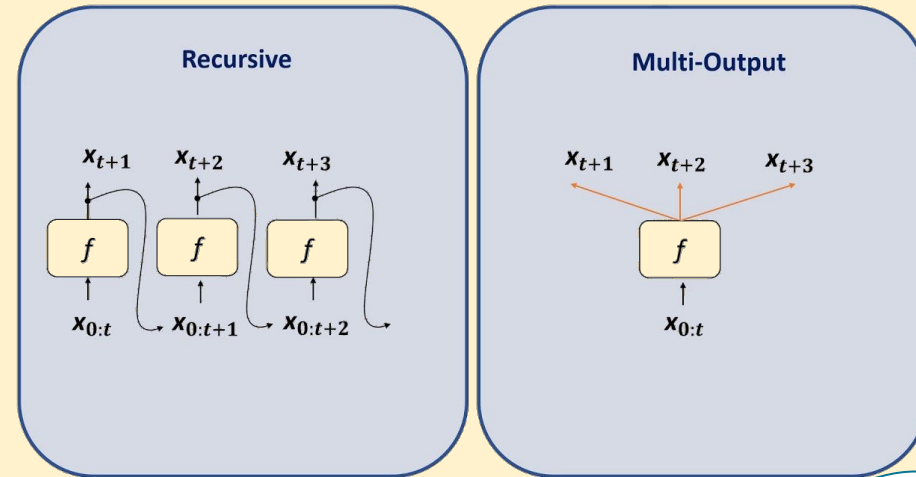
Entrenar un model diferent en cada step.
Prediccions independents una de l'altre. Més cost computacional.



Multiple output forecasting

Predir de forma simultània varis elements de la seqüència.

Multi-Step Forecasting



Forecasting recursive autoregressive

- Predir de la mateixa variable un resultat futur

Forecasting recursive autoregressive amb variables exògenes

- Utilitzar variables que coneixem per predir resultats

Forecasting recursive autoregressive amb variables personalitzades

- Afegir variables personalitzades de la sèrie temporal

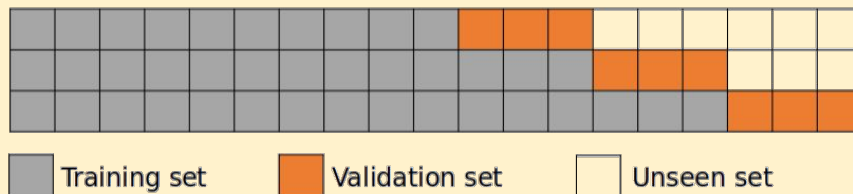
<https://nix-united.com/blog/find-out-how-to-use-machine-learning-for-time-series-forecasting/>

Backtesting

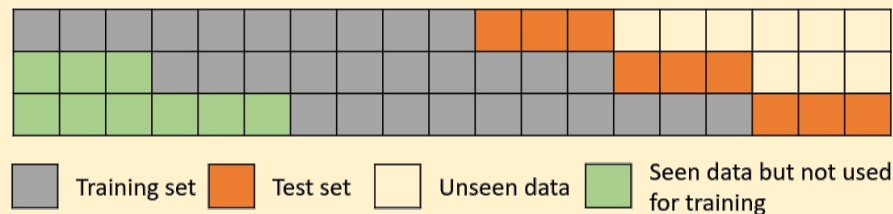
És un mètode de validació amb les dades històriques per avaluar el model.

- Amb reentrenament: El model entrena cada cop utilitzant totes les dades
- Amb reentrenament i training constant: El nombre de dades al train és el mateix (time series cross-validation).
- Sense reentrenament: Només s'entrena un cop les dades, molt més ràpida.

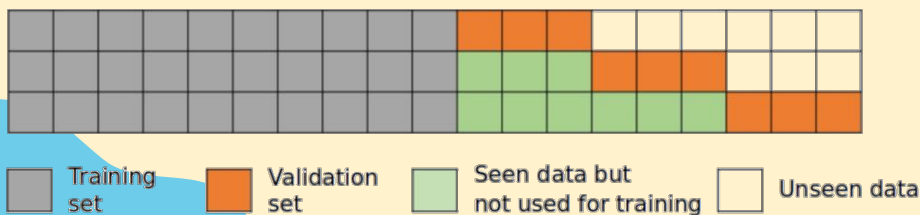
Time series backtesting with refit



Time series backtesting with refit and fixed train size



Time series backtesting without refit



Mètriques utilitzades

Time Series Forecast Error Metrics

Scale-Dependent Metrics

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Square Error (RMSE)

Percentage-Error Metrics

- Mean Absolute Percentage Error (MAPE)
- Symmetric Mean Absolute Percentage Error (SMAPE)

Relative-Error Metrics

- Median Relative Absolute Error (MdRAE)
- Geometric Mean Relative Absolute Error (GMRAE)

Scale-Free Error Metrics

- Mean Absolute Scaled Error (MASE)

Com de bé prediu la serie temporal:
Residual Diagnostics

Outliers i anomalies

- **Identificació:** test d'hipòtesis o anàlisis
 - <https://neptune.ai/blog/anomaly-detection-in-time-series>
- **Què fer?**
 - Eliminar
 - Quantile based flooring and capping
 - Suavitzat -> mean imputation
 - Replacement
 - Estudi del data set

Bones pràctiques

1. Entendre la teva base de dades i quina classe de sèrie temporal tens
2. Seleccionar features
3. Pre-processat
4. Treballar amb les anomalies i els outliers
5. Evitar l'overfitting -> Ens interessa tècniques com validació interna (Backtesting) i fer samples de la data.
6. Evita dades infinites -> No necessites totes les dades històriques, i si la teva feina és predir el futur pròxim, a vegades podem sobredimensionar-nos.
7. Treballar amb dada estacionaria -> Si tens dada estacionària, treballa amb un dataset que contingui 'estacionalitat de les dades.

Recursos

<https://bookdown.org/content/2274/series-temporales.html>

<https://neptune.ai/blog/time-series-prediction-vs-machine-learning>

<https://nix-united.com/blog/find-out-how-to-use-machine-learning-for-time-series-forecasting>

SKLEARN:

<https://www.cienciadedatos.net/documentos/py27-forecasting-series-temporales-python-scikitlearn.html>

Recursos d'interès:

- Sèries temporals i NN:
<https://www.aprendemachinelearning.com/pronostico-de-series-temporales-con-redes-neuronales-en-python/>
- Com no utilitzar ML per sèries temporals:
<https://towardsdatascience.com/how-not-to-use-machine-learning-for-time-series-forecasting-avoiding-the-pitfalls-19f9d7adf424>

Ja hem acabat!

Gràcies a tots!

