


Series Temporales

Aitor Donado Urquizu



Series Temporales



- Una serie temporal es una sucesión de datos ordenados cronológicamente y espaciados a intervalos iguales o desiguales.
 - Este concepto es muy amplio y, por lo tanto, lo podemos encontrar en muchísimas aplicaciones o casos de uso
 - las predicciones meteorológicas
 - en las cotizaciones de la Bolsa de Valores
 - en análisis de tráfico web (predecir la carga de un servidor o de una página web.)
- 

Forecasting

- Consiste en predecir el valor futuro que va a tener una serie temporal
 - bien sea modelando la serie únicamente en función del comportamiento pasado que ha tenido (modelado autorregresivo)
 - O bien empleando variables externas o **exógenas** a la serie temporal
- Normalmente utilizamos datos históricos con el objetivo de obtener una representación matemática que sea capaz de predecir en valores futuros de esa serie.
- Esta idea está fundamentada en la asunción de que el comportamiento futuro del fenómeno que se está modelando se puede explicar a partir de su comportamiento pasado

Forecasting

- En términos generales podemos decir que el forecast (predicción) se puede dividir en dos partes
 - Los patrones o la varianza explicada y
 - La varianza no explicada
- $\text{Forecast} = \text{Patrones} + \text{Varianza no explicada}$
 - “Patrones” hace referencia a todo aquello que tiene un carácter repetitivo a lo largo del tiempo como pueden ser tendencias, estacionalidades, factores cíclicos... todo aquello que vemos que se va repitiendo en el tiempo
 - el segundo término, varianza no explicada, engloba todo aquello que influye en nuestra variable respuesta pero que no está recogido o explicado por el pasado de la serie temporal, es decir, no está incluido en los patrones

Estrategias de Forecasting

- Modelos estadísticos-econométricos.
 - Autorregresión AR
 - Moving Average MA
 - Autoregressive Moving Average ARMA
 - Autoregressive Integrated Moving Average (ARIMA)
 - Simple Exponential Smoothing SES
 - Holt Winters Exponential Smoothing HWES
- Modelos de ML.
 - Lasso.
 - Ridge.
 - Random Forest.
 - Gradient Boosting.
 - SVM
 - RRNN
 - LSTM
 - Modelos de step único
 - Modelos multi-step recursivo
 - Modelos multi-step directo

Métodos de Machine Learning



Ventajas

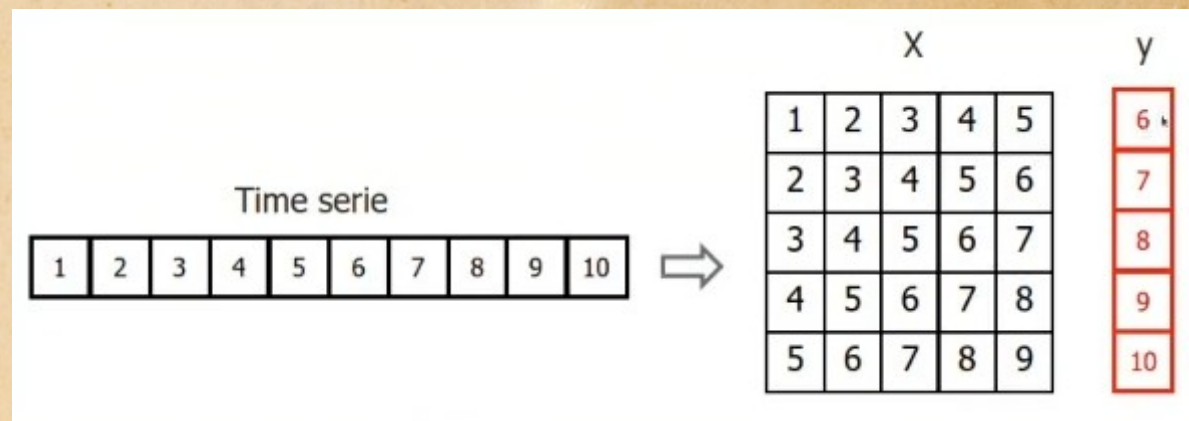
- Facilidad de incluir variables exógenas además de las auto regresivas.
- Es relativamente más sencillo incorporar comportamientos no lineales ya que podemos utilizar modelos como son RandomForest GradientBoosting y Máquinas Vector Soporte que, de forma nativa son capaces de incorporar o aprender relaciones no lineales entre los predictores.
- Elevada escalabilidad ya que son aplicables incluso cuando se dispone del grandes volúmenes de datos

Desventajas

- Necesidad de reestructurar los datos para poder utilizarlos como si se tratara de un problema de regresión.
- Necesidad de programar procesos iterativos en los que cada predicción utiliza las predicciones anteriores cuando intentamos predecir un horizonte de de de de tiempo el que contiene múltiples steps.
- Los modelos de de forecasting requieren de métodos de estrategias de validación específicas BackTesting, Walk-forward validation, Time-series cross validation.

Creación de la matriz de entrenamiento

La principal adaptación que se necesita hacer para aplicar modelos de Machine Learning a problemas de forecasting es transformar la serie temporal en una matriz en la que, cada valor, está asociado a la ventana temporal (lag) que le precede

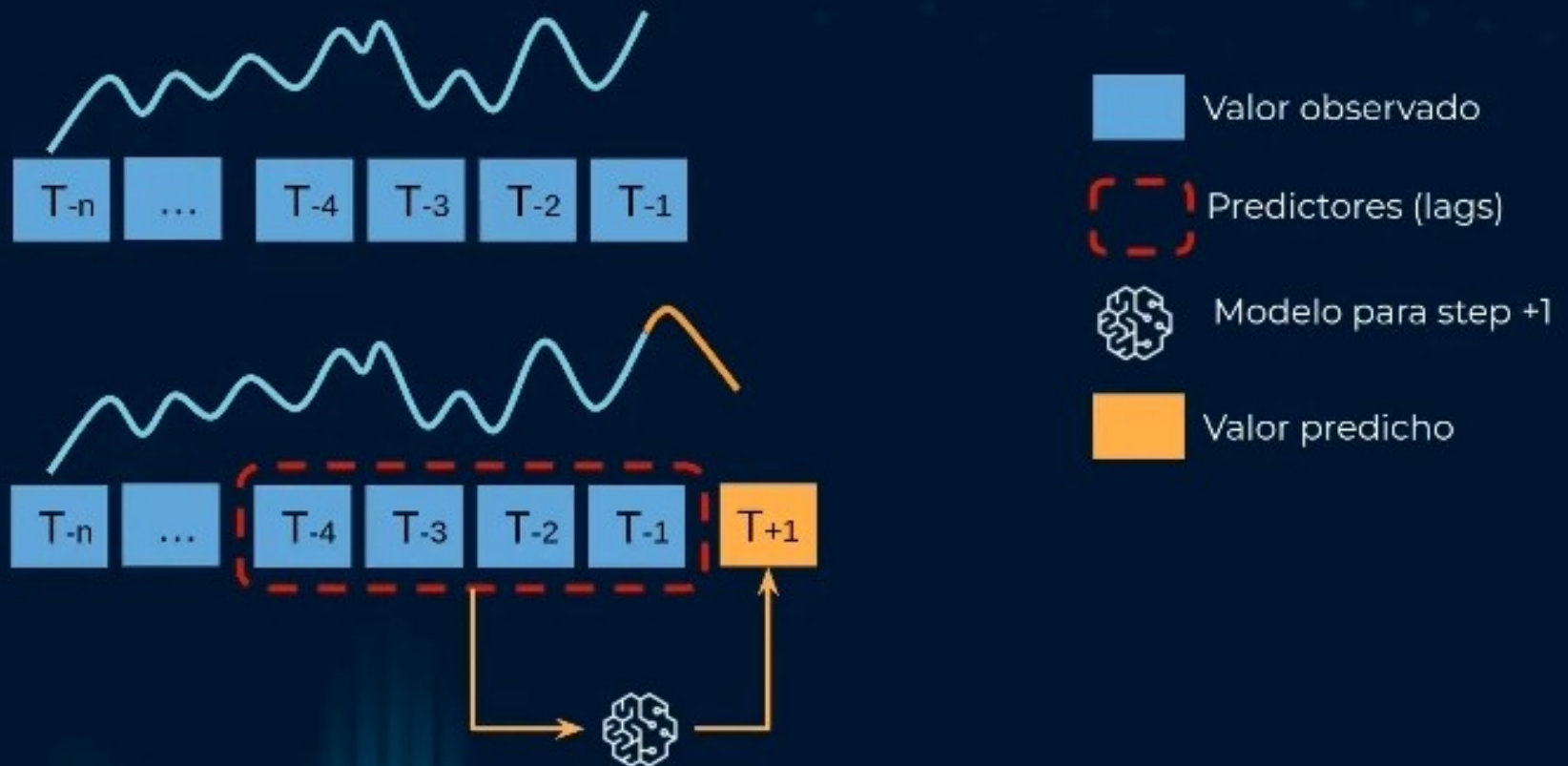


Horizonte temporal

- Predicción Single Step
 - El objetivo es predecir el siguiente valor de la serie temporal
- Predicción Multi Step
 - El objetivo es predecir los n siguientes valores de la serie.

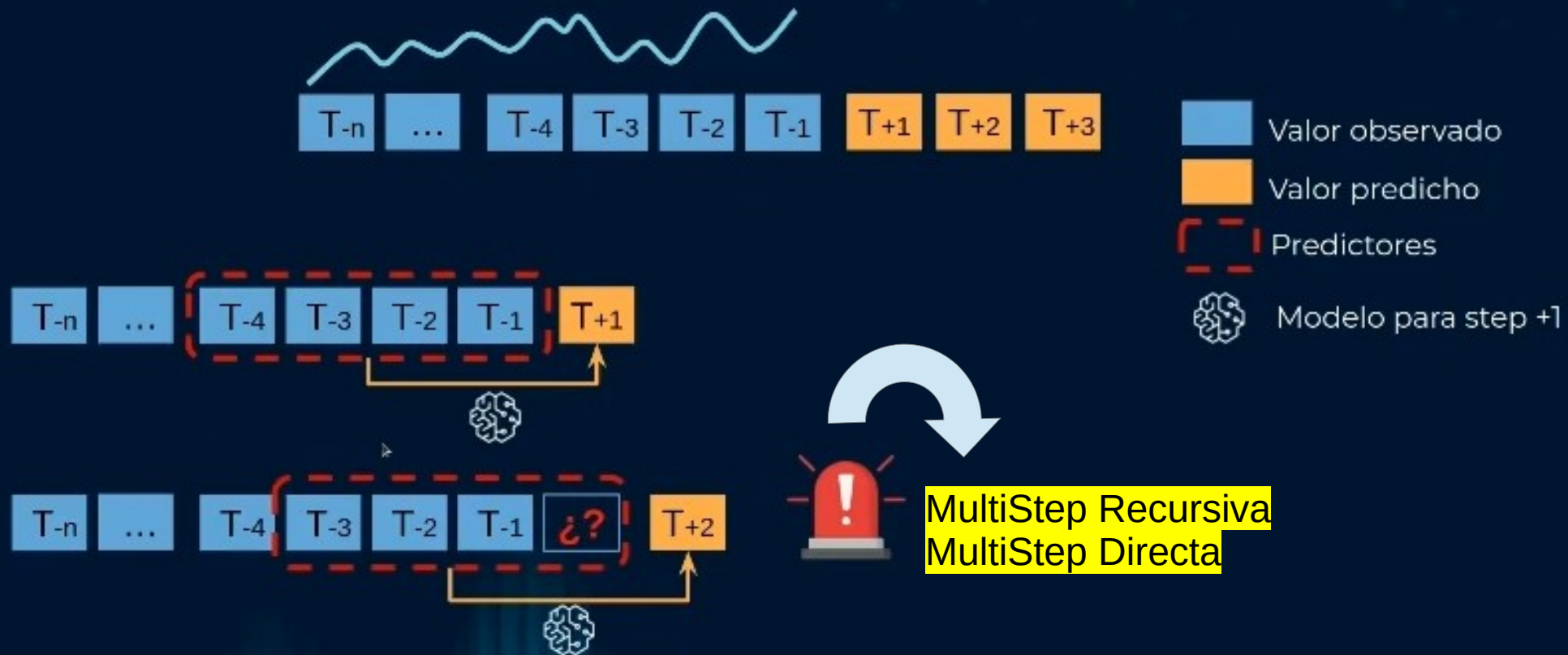
Predicción Single Step

El objetivo es predecir únicamente el siguiente valor de la serie.



Predicción Multi-Step

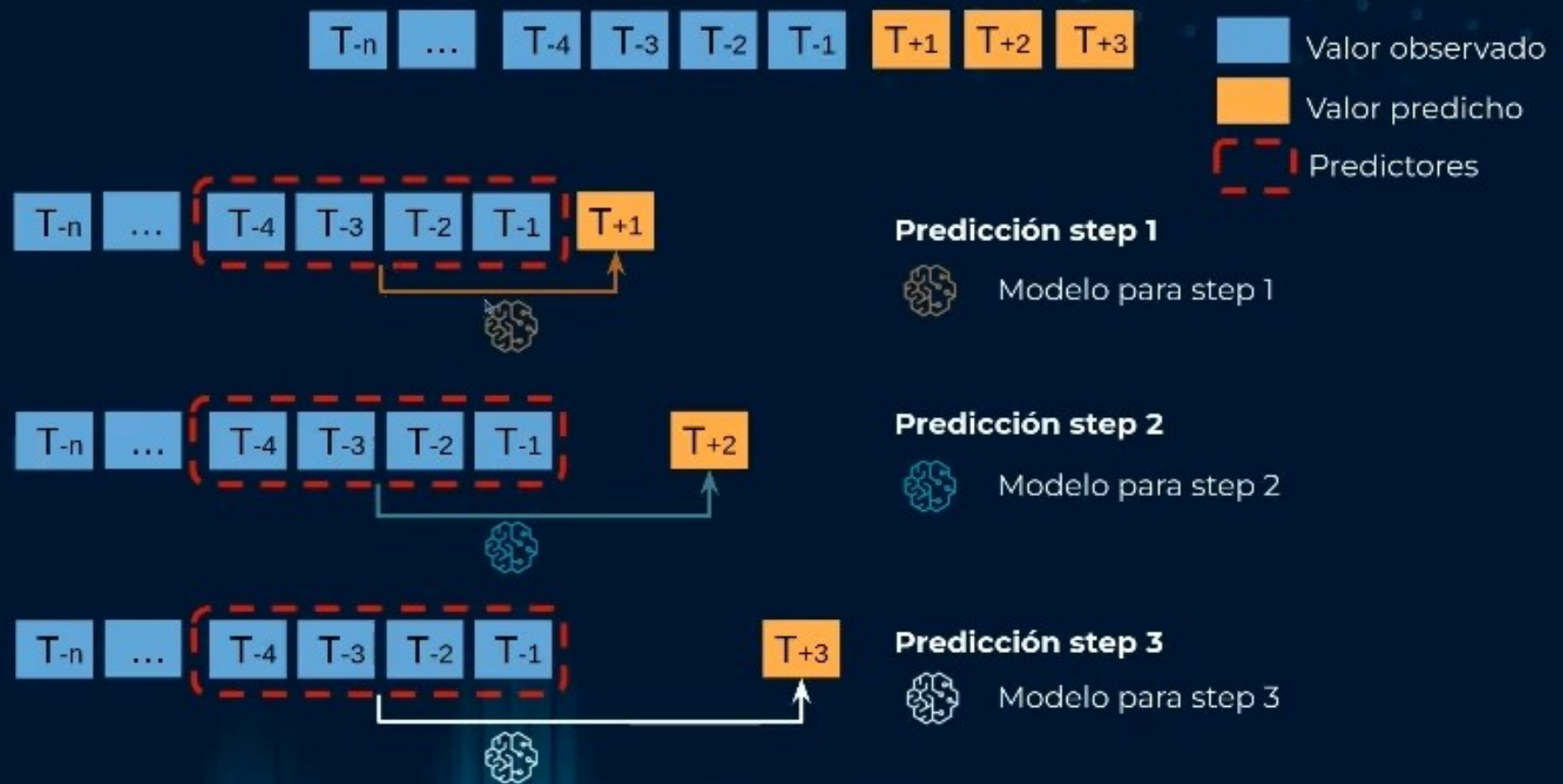
El objetivo es predecir los n siguientes valores de la serie.



Multi-Step Recursiva



Multi-Step Directa



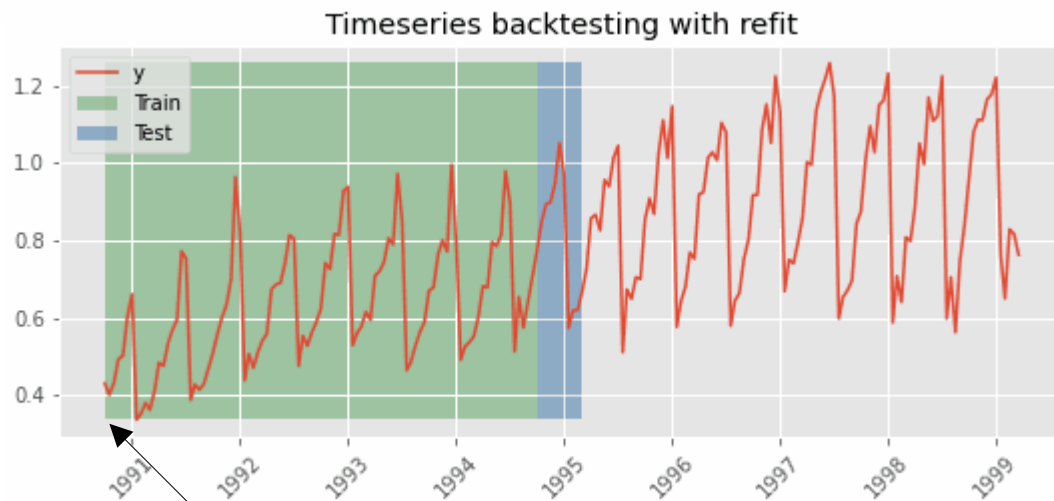
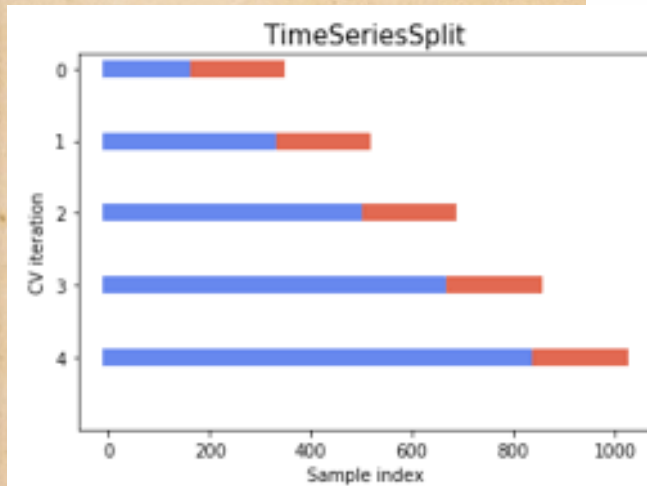
Validación del modelo

- NO PUEDO BARAJAR ALEATORIAMENTE TRAIN Y TEST
- El orden tiene un sentido importante en las Series Temporales



Time Series Cross Validation

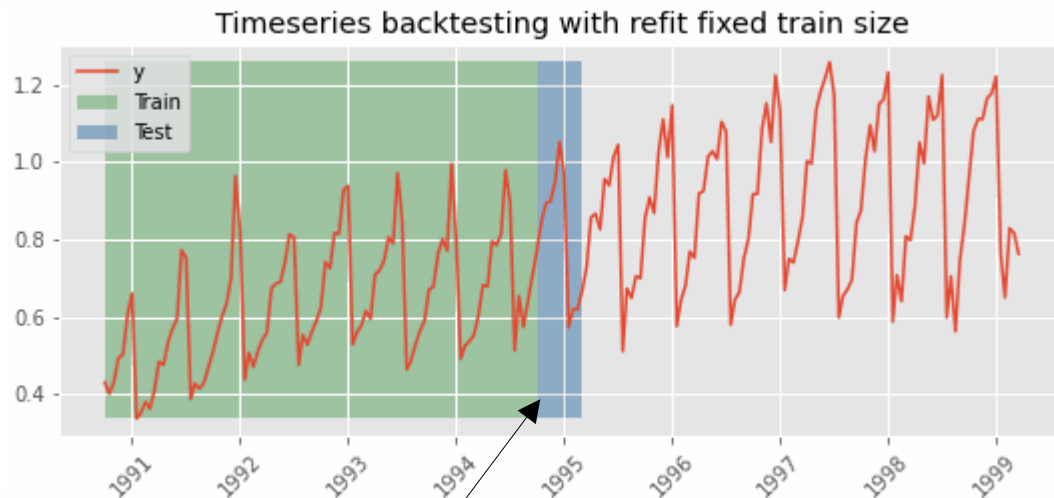
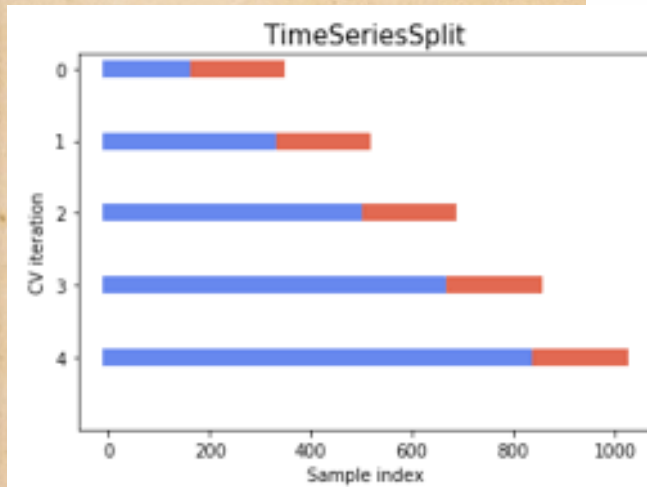
- Back Testing con reentrenamiento
- Con origen fijo



Fixed origin

Time Series Cross Validation

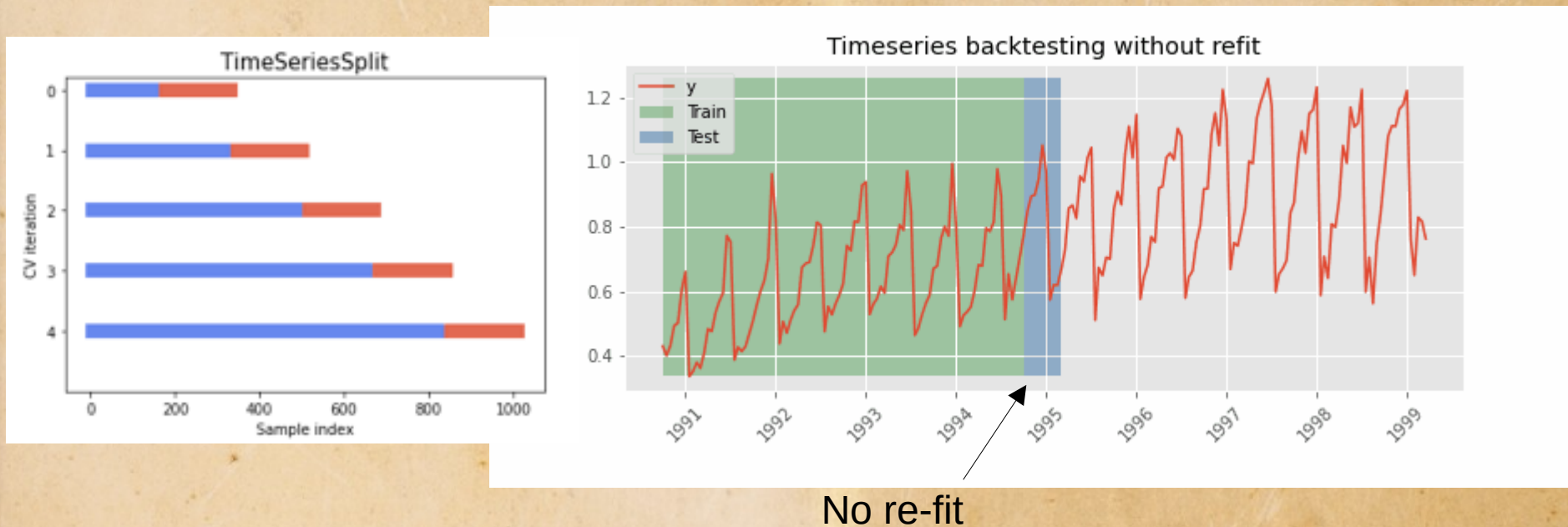
- Back Testing con reentrenamiento
- Amplitud de entrenamiento fija



Fixed train width

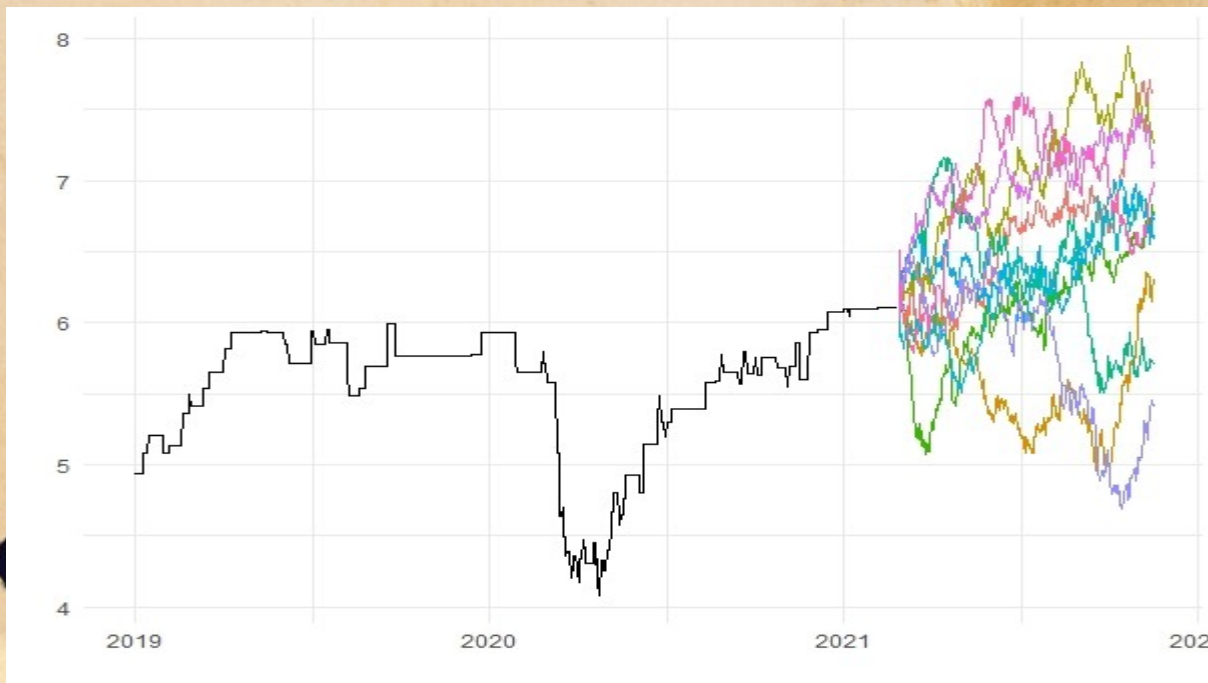
Time Series Cross Validation

- Back Testing SIN reentrenamiento



Intervalos de predicción

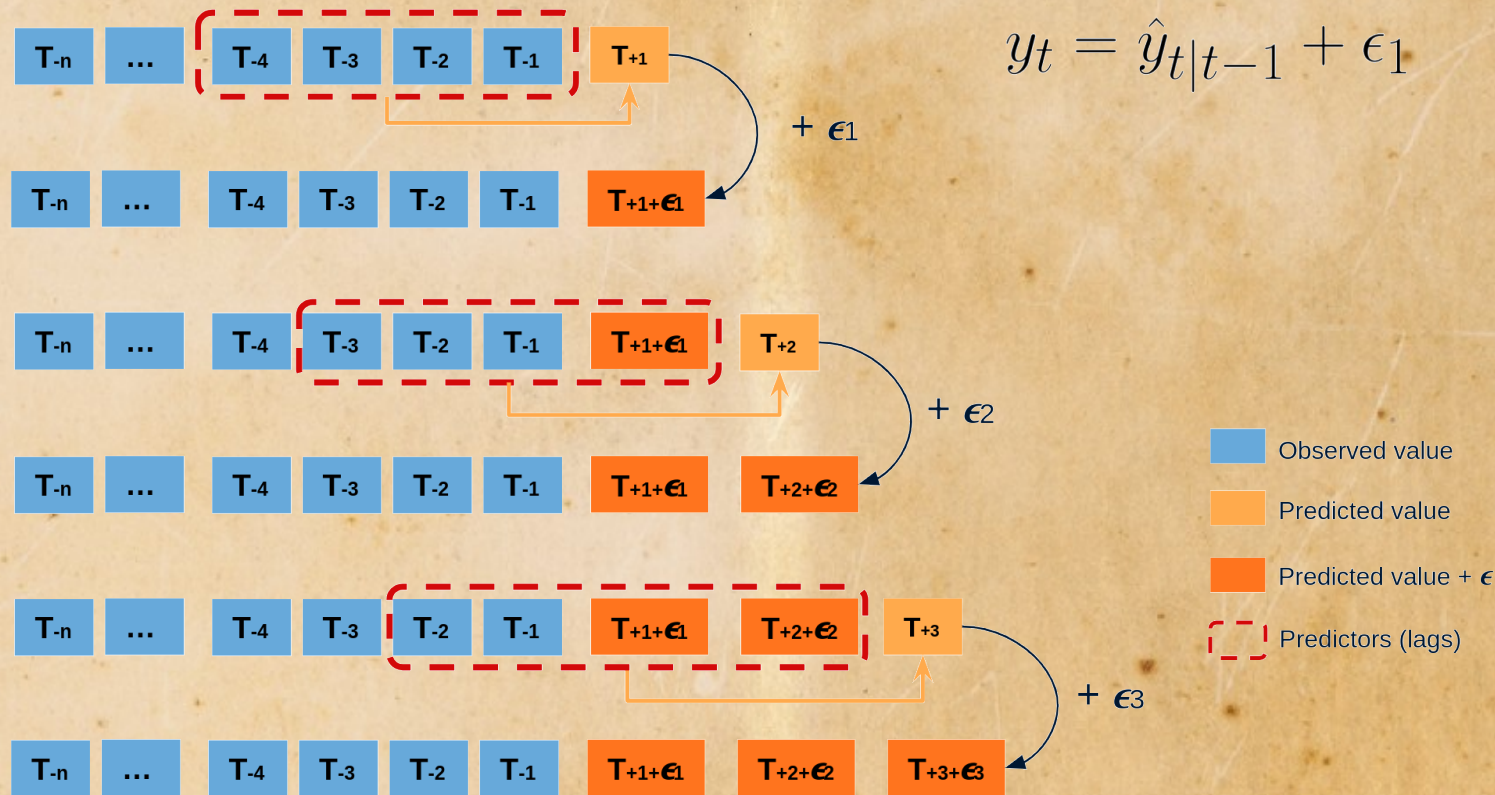
- BOOTSTRAPPED RESIDUALS
- Un intervalo de predicción define el intervalo dentro del cual es de esperar que se encuentre el verdadero valor de y con una determinada probabilidad.
- Por ejemplo, es de esperar que el intervalo de predicción del 98% contenga el verdadero valor de la predicción con un 98% de probabilidad.



El valor de la predicción suele residir más en lo “fino” que sea el intervalo de confianza de la predicción que en el valor predicho.

Intervalos de predicción

- El error de una predicción se acumula en las predicciones siguientes.

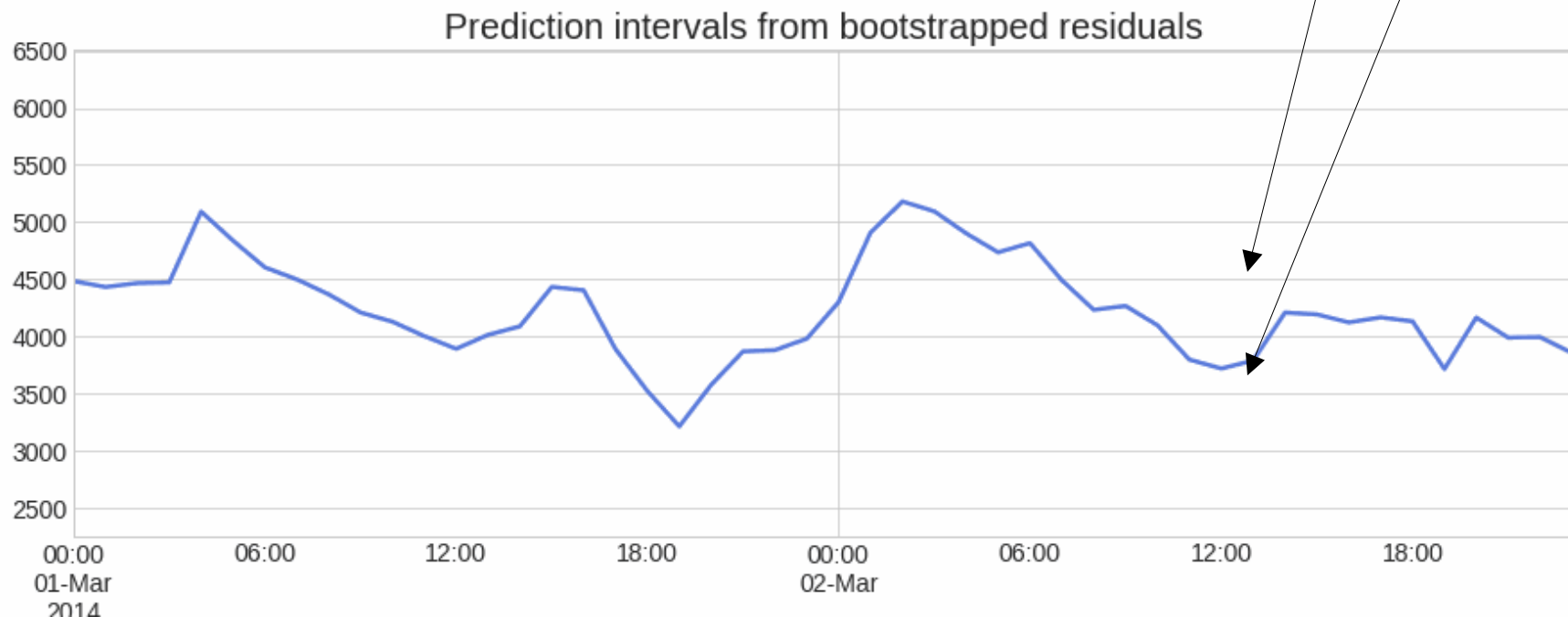


Intervalos de predicción



No hago una única predicción sino cientos de ellas.

Finalmente me quedo con el intervalo entre percentiles (10 y 90, por ejemplo)



Variables exógenas

- Además de utilizar predictores autorregresivos generados a partir del pasado de la propia variable respuesta, es posible Incorporar en el modelo variables adicionales, siempre y cuando **su valor a futuro se conozca**.
- Variables poco problemáticas
 - Variables creadas a partir del calendario (día de la semana, festivos, mes del año..)
- Variables potencialmente problemáticas
 - Variables meteorológicas (temperatura, viento, precipitación..). Cuando el modelo se ponga en producción, no se conoce a ciencia cierta qué condiciones habrá en el futuro, sino que se dispondrá de predicciones generadas por los servicios meteorológicos. Dado que se trata de predicciones, contendrán errores que se introducirán en el modelo. Como consecuencia, es muy probable que las predicciones del modelo empeoren

Modelos de Forecasting en producción

Opciones

- Reentrenar el modelo antes de ejecutar el proceso de predicción
 - ✓ El modelo utilizará en el entrenamiento todos los datos disponibles hasta el momento
 - ✗ Se tiene que almacenar un histórico incremental
 - ✗ Mayor requerimiento computacional.
 - ✗ El tiempo entre ejecución de predicciones debe ser suficiente para entrenar el modelo
- Ejecutar las predicciones sin necesidad de reentrenamiento
 - ✓ Menor requerimiento computacional
 - ✓ No hay restricciones de tiempo entre predicciones
 - ✓ No es necesario almacenar el histórico
 - ✗ No todas las implementaciones permiten este comportamiento
 - ✗ El modelo no utiliza en el entrenamiento todos los datos disponibles hasta el momento, puede quedar desactualizado



This work is licensed under
a Creative Commons Attribution-ShareAlike 3.0 Unported License.
It makes use of the works of
Kelly Loves Whales and Nick Merritt.