Enfoque Estadístico del Aprendizaje

Maestría en Explotación de Datos y el Descubrimiento de Conocimiento - UBA

Predicción de series temporales de producción de petróleo y gas en Argentina

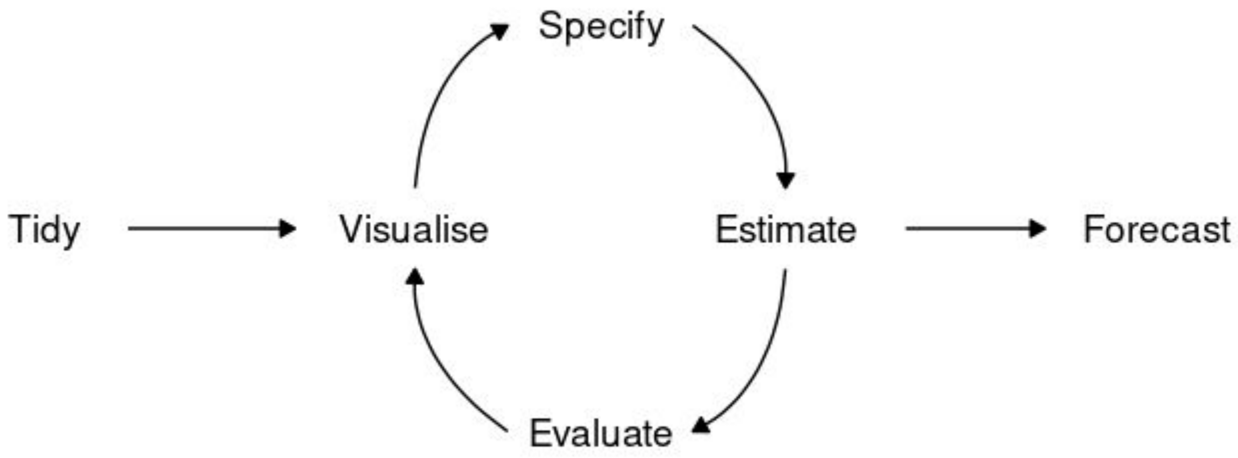
Autores:

- Baez, Lucas
- Pedersen, Sebastián
- Suster, Mateo

Objetivos

- Modelar la evolución temporal de la producción petróleo y gas en Argentina
- Entender el resultado de las distintas predicciones
- Comparar las predicciones entre modelos

Metodología



Fuente: Hyndman and Athanasopoulos (2021)

Datos

- La **Secretaría de Energía de la Nación** publica información sobre los volúmenes de producción de hidrocarburos (gas y petróleo) por pozo de todo el país de manera mensual, entre otras variables (precios, inversiones, etc.)
- Tareas de preprocesamiento (concatenación, ajustes, etc.)
- Período contemplado: 2013 a 2020
  - Separación entre train y test

Modelos

- Benchmarks
- STL
- Prophet
- Procesos gaussianos

Referencias (I)

- Forecasting
  - Hyndman, R.J., & Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia
- STL
  - Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. J. (1990). STL: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics, 6(1).
  - Theodosiou, M. (2011). Forecasting monthly and quarterly time series using STL decomposition. International Journal of Forecasting, 27(4), 1178–1195
  - O'Hara-Wild (2019) Introducing feasts
  - Paquete R: fabletols
- Prophet
  - Harvey & Peters (1990). Estimation Procedures for Structural Time Series Models.
  - Barriola y Perini (2021). GAM en series de tiempo: Prophet. Nootebook de EEA
  - Librería Prophet para R

Referencias (II)

- Procesos Gaussianos
  - Robert B. Gramacy (2021). Surrogates: gaussian process modeling, design and optimization for the applied sciences.
  - Felipe Tobar (2021). Aprendizaje de máquinas. Cap. 8: Procesos gaussianos.
  - Carl E. Rasmussen, Christopher K. I. Williams (2006). Gaussian Processes for Machine Learning. MIT Press.
  - Kevin P. Murphy (2012). Machine Learning: A Probabilistic Perspective. MIT Press. Ch. 15: Gaussian processes.
  - Paquete de R, Kernlab (Kernel-Based Machine Learning Lab).

Modelos benchmark (I)

- **Mean method.** Predicción futura por medio de la media histórica de los datos

$$\hat{y}_{T+h|T} = \bar{y} = (y_1 + \cdots + y_T)/T.$$

- **Naïve method**. Predicción por última observación

$$\hat{y}_{T+h|T} = y_T.$$

Modelos benchmark (II)

- **Seasonal naïve method**. Predicción por última observación de acuerdo a la estacionalidad del tiempo

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)},$$

- **Drift method.** Variación del método naïve, pero con cambios en el tiempo iguales a la variación promedio en los datos históricos

$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1}) = y_T + h\left(\frac{y_T - y_1}{T-1}\right).$$

Descomposición clásica de una serie temporal

- Componente de tendencia-ciclo (T_t)
  - Moving average
- Componente estacional (S_t)
  - Promedio ajustado de la serie sin tendencia para cada estación
- Componente remanente (R_t)

- Aditiva $y_t = S_t + T_t + R_t,$
- Multiplicativa $y_t = S_t \times T_t \times R_t.$

➔ Desventajas
- Patrón de estacionalidad constante en todos los períodos
- Imposibilidad de estimar una tendencia para las primeras y últimas observaciones
- No es robusta a datos atípicos

STL

- Nombre de Seasonal and Trend decomposition using Loess
  - Loess (locally weighted smoothing): método no paramétrico
- Ventajas
  - Puede manejar cualquier tipo de estacionalidad
  - Componente estacional controlable y cambiante a lo largo del tiempo
  - Mayor flexibilidad y control del componente de tendencia-ciclo
  - Robustez ante outliers
  - Buen desempeño en series largas
- Desventajas
  - Dificultad de obtener descomposiciones multiplicativas (necesidad de aplicar transformaciones, e.g. Box - Cox)

STL. Predicción

- **Idea general**: *dividir el problema de predicción en partes más pequeñas*

- Los **componentes de tendencia y estacionalidad** se predicen de manera separada
- Para estimar el **componente estacional** con cualquier método de forecasting no estacional
- A diferencia de la predicción por descomposición clásica, se modela el **componente remanente**

$$y_t = \hat{S}_t + \hat{A}_t,$$

Donde:

$$\hat{A}_t = \hat{T}_t + \hat{R}_t$$

Prophet (Facebook)

- Librería desarrollada por Facebook para facilitar la tarea de creación de modelos predictivos en series temporales y detección de anomalías.

- En el fondo, es parte de la familia de los llamados modelos aditivos generalizados (GAM).

Prophet (Facebook)

- Propone un modelo de la forma

$$y(t) = trend(t) + seasonality(t) + holliday\_effects(t)$$

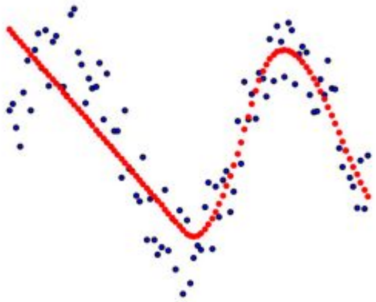
Donde

- trend(t): Se compone de funciones lineales (o logisticas) a trozos que capturan las tendencias de la serie.
- seasonality(t): Captura las tendencias periódicas.
- holliday_effects(t): Permite al usuario introducir momentos particulares o anómalos a priori para ajustar los modelos.

¿Qué es un Proceso Gaussiano (PG)?

Problema de regresión: dado un conjunto de observaciones (x,y) busca encontrar un función f(x)=y.

$$\{(x_i, y_i)\}_{i=1}^N \longrightarrow y = f(x)$$



Regresión clásica: busca un único candidato de f.

Proceso Gaussiano: busca una distribución sobre la f.

$$\mathbb{P}(f) \qquad f : \mathcal{X} \to \mathbb{R}.$$

¿Qué es un Proceso Gaussiano (PG)?

Problema de regresión $\{(x_i, y_i)\}_{i=1}^{N} \longrightarrow y = f(x)$

Proceso Gaussiano: busca una distribución sobre la f.

$\mathbb{P}(f) \qquad f : \mathcal{X} \to \mathbb{R}.$

Para cualquier subconjunto $\{x_i\}_{i=1}^{n} \subset \mathcal{X}$ la distribución de $\mathbb{P}(f(\mathbf{x}))$ es una gaussiana multivariada $f(\mathbf{x}) = (f(x_1), \ldots, f(x_n))$

¿Qué es un Proceso Gaussiano (PG)?

Problema de regresión $\{(x_i, y_i)\}_{i=1}^{N} \longrightarrow y = f(x)$

Proceso Gaussiano: busca una distribución sobre la f.

$$\mathbb{P}(f) \qquad f : \mathcal{X} \to \mathbb{R}.$$

Para cualquier subconjunto $\{x_i\}_{i=1}^{n} \subset \mathcal{X}$ la distribución de $\mathbb{P}(f(\mathbf{x}))$ es una gaussiana multivariada $f(\mathbf{x}) = (f(x_1), \ldots, f(x_n))$

La gaussiana multivariada queda caracterizada por media y covarianza, las cuales suponemos dependen de $\{x_i\}_{i=1}^{n} \subset \mathcal{X}$

$$f(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), K(\mathbf{x}, \mathbf{x}))$$

¿Qué es un Proceso Gaussiano (PG)?

Es normal asumir media cero en la normal multivariada.

Para la covarianza K, también llamada Kernel, hay que pedir que sea semi definida positiva (el análogo a pedir varianza positiva o cero en el caso univariado). Una elección popular es el kernel RBF (radial bases function):

$$K_{SE}(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2\ell^2}\right)$$

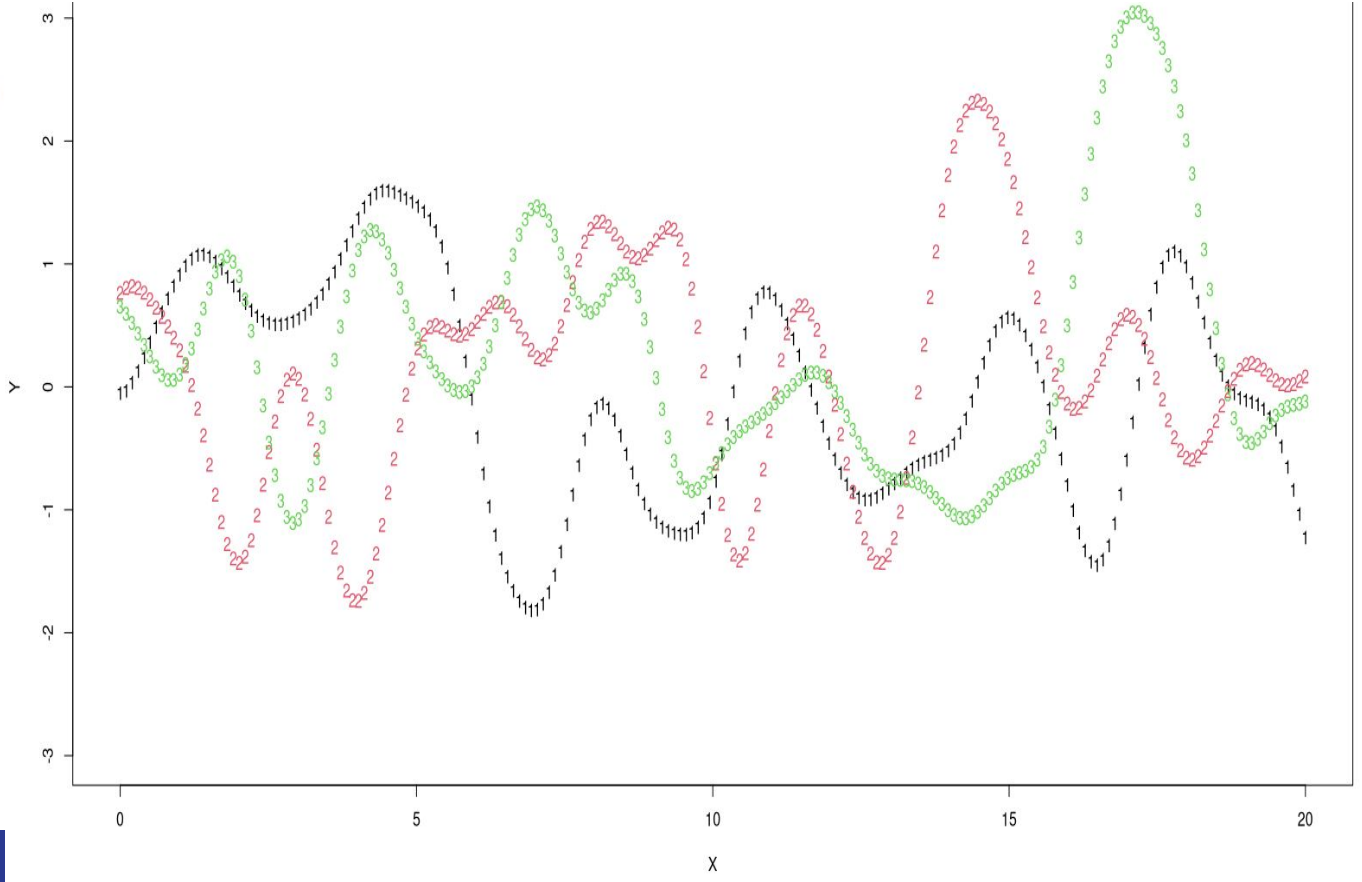
(si las obs. x son vectores es la norma de la resta).

Sampleo de un PG (PG “prior”)

$$f(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), K(\mathbf{x}, \mathbf{x}))$$

Con media cero y kernel RBF.

Sampleamos tres veces para 200 valores de x equiespaciados entre 0 y 20:



Incorporando información: posterior predicción sin ruido

Tenemos observaciones $\{(x_i, f(x_i))\}_{i=1}^{n}$ y queremos realizar un predicción sobre un conjunto distinto X_* de m valores. Vale que:

$$\begin{bmatrix} f(X) \\ f(X_*) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(X) \\ m(X_*) \end{bmatrix}, \begin{bmatrix} K(X,X) & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right)$$

Y además la distrib. cond. nueva cumple:

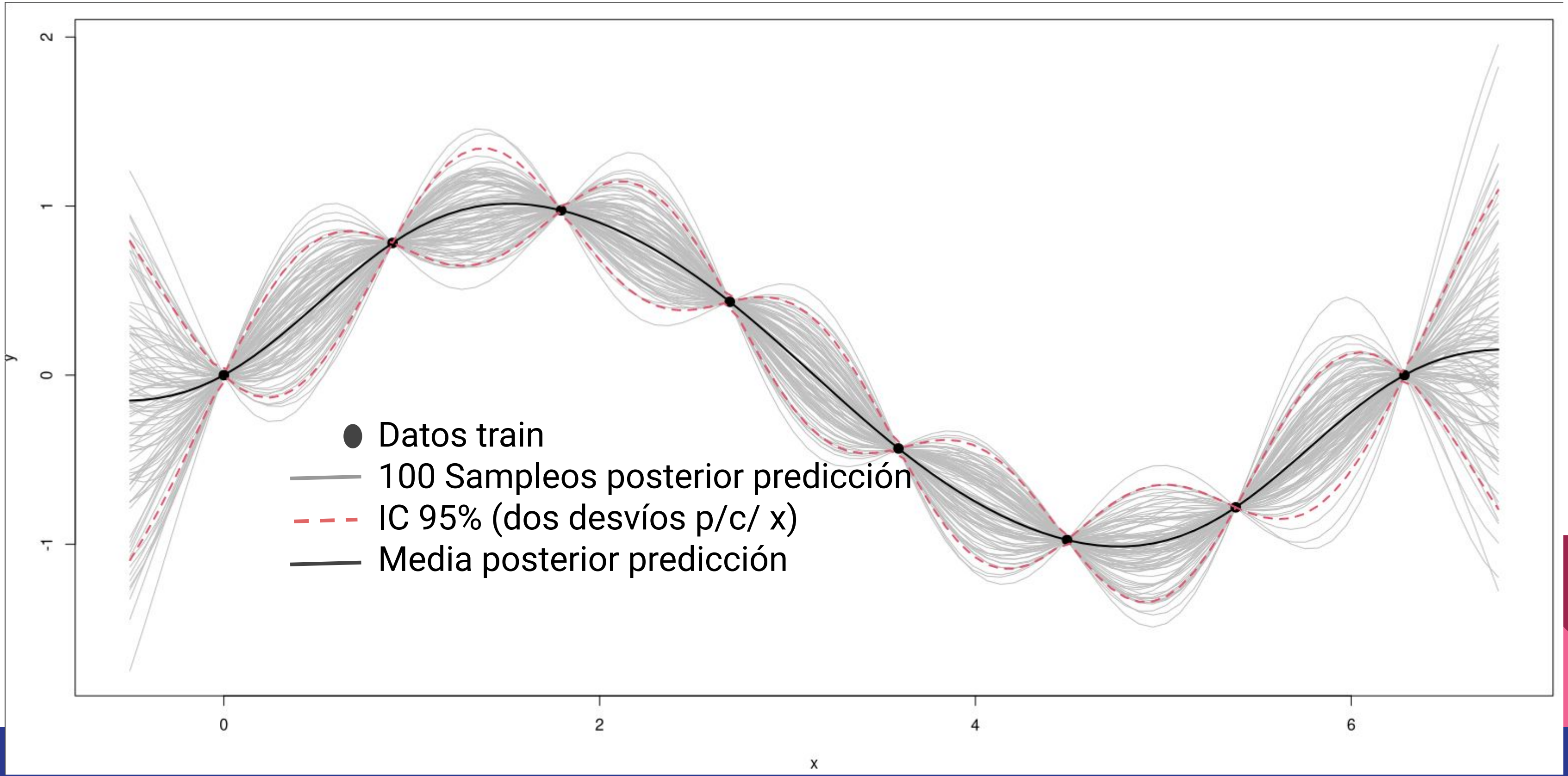
$$f(X_*)|f(X), X \sim \mathcal{N}(m_{X_*|X}, \Sigma_{X_*|X})$$

$$m_{X_*|X} = m(X_*) + K(X_*,X)K^{-1}(X,X)(f(X) - m(X))$$
$$\Sigma_{X_*|X} = K(X_*,X_*) - K(X_*,X)K^{-1}(X,X)K(X,X_*)$$

Es decir tenemos estas observaciones dato, ponemos un PG prior, damos un rango donde queremos predecir, y obtenemos la distrib. posterior para predecir.

Posterior predicción sin ruido: Ejemplo

$f(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), K(\mathbf{x}, \mathbf{x}))$, con media cero y kernel RBF.



Incorporando información: posterior predicción con ruido

Ahora tenemos datos con ruido $y_i = f(x_i) + \eta$, $\eta \sim \mathcal{N}(0, \sigma_n^2)$. Es fácil ver que esto equivale a agregar un término a la diagonal de la covarianza: $cov(Y) = K(X,X) + \sigma_n^2 \mathbb{I}$

Análogo al caso anterior, vale que:

$$\begin{bmatrix} Y \\ f(X_*) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(X) \\ m(X_*) \end{bmatrix}, \begin{bmatrix} K(X,X) + \sigma_n^2 \mathbb{I} & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix} \right)$$

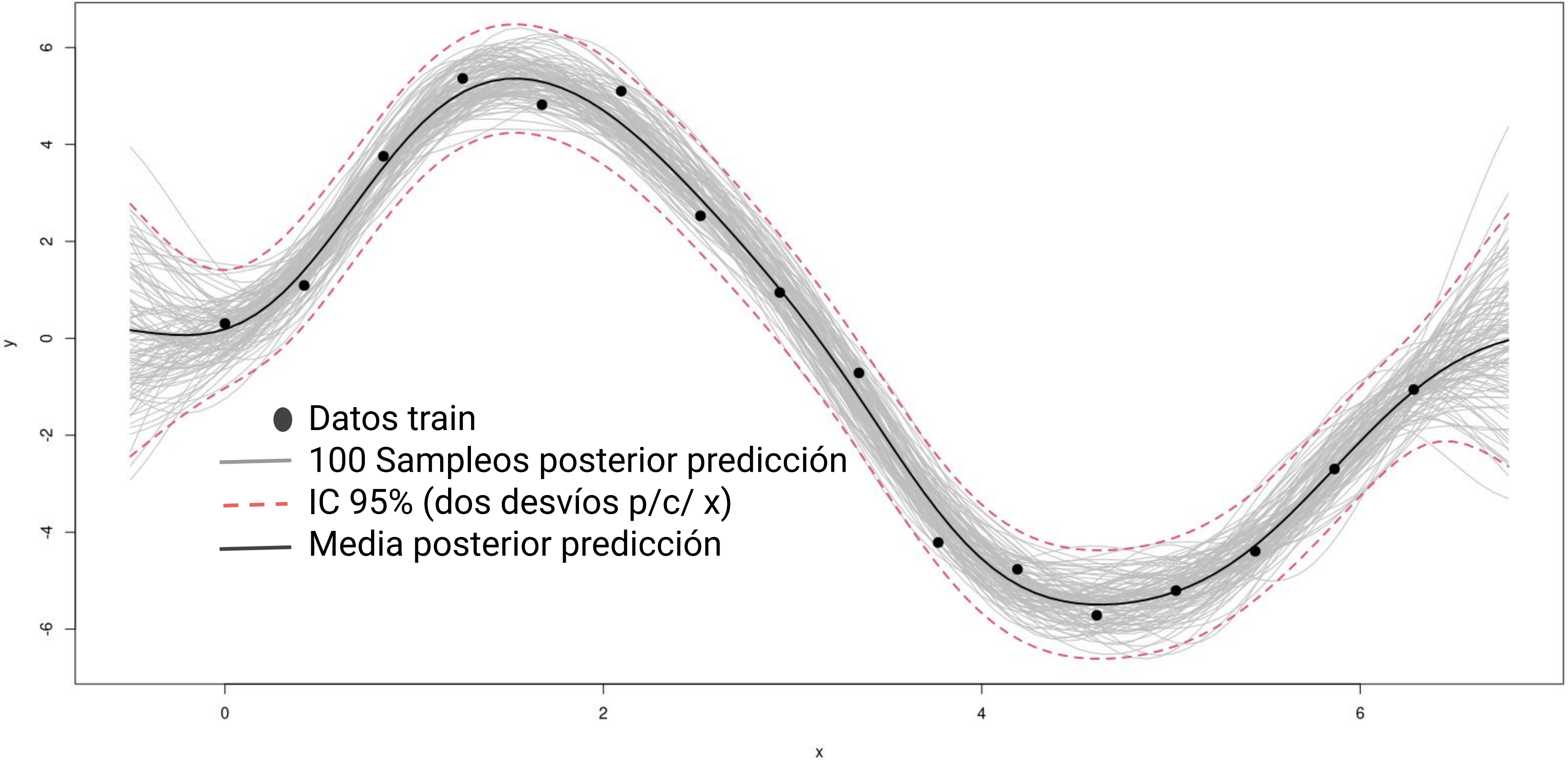
Y la distrib. cond. nueva cumple:

$$f(X_*)|Y,X \sim \mathcal{N}(m_{X_*|X}, \Sigma_{X_*|X}) \quad \begin{aligned} m_{X_*|X} &= m(X_*) + K(X_*,X)[K(X,X) + \sigma_n^2 \mathbb{I}]^{-1}(Y - m(X)) \\ \Sigma_{X_*|X} &= K(X_*,X_*) - K(X_*,X)[K(X,X) + \sigma_n^2 \mathbb{I}]^{-1} K(X,X_*) \end{aligned}$$

Nuevamente: datos train, prior,
y obtenemos la posterior para predecir

Posterior predicción con ruido: Ejemplo

$f(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), K(\mathbf{x}, \mathbf{x}))$, con media cero y kernel RBF.



Estimación de parámetros en PG.

Por supuesto es muy importante, antes de calcular y usar la posterior predicción, estimar los parámetros del modelo:

- Parámetros del Kernel
- Parámetro del ruido de las obs.
- Otros.

Popularmente se realiza mediante Máxima Verosimilitud, ya sea MLE o por inferencia bayesiana la distribución.

Ejemplo: estimar el parámetro de varianza del RBF es muy importante pues sino la posterior predicción puede estar subestimando el error.

Algunas cosas más en PG.

- Más dimensiones: no mucho más que trabajar con normas en vez de módulos.

- Otros Kernels.

$$K_{RQ}(x, x') = \sigma^2 \left(1 + \frac{(x - x')^2}{2\alpha\ell^2}\right)^{-\alpha}$$

$$K_P(x, x') = \sigma^2 \exp\left(-\frac{2\sin^2\left(\pi|x - x'|/p\right)}{\ell^2}\right)$$

Aplicación datos reales hidrocarburos