

# AJUSTE DE HIPERPARÁMETROS



# ¿Qué son los hiperparámetros?

Los hiperparámetros (HPs) son aquellos parámetros que caracterizan un modelo de aprendizaje automático / aprendizaje profundo, cuyos valores no se ajustan de forma automática durante el entrenamiento.

Árbol de decisión	Random forest	Máquina de soporte vectorial	Red neuronal (MLP)
<ul style="list-style-type: none"><li>• Métrica por la cual se van a realizar las divisiones en los nodos.</li><li>• Profundidad del árbol.</li><li>• Cantidad de características a evaluar en cada nodo.</li></ul>	<ul style="list-style-type: none"><li>• Cantidad de estimadores.</li><li>• Taza de aprendizaje para GBM.</li></ul>	<ul style="list-style-type: none"><li>• Kernel.</li><li>• Grado del kernel para polinómicos.</li><li>• Parámetro de regularización C.</li></ul>	<ul style="list-style-type: none"><li>• Cantidad de capas.</li><li>• Cantidad de neuronas por capa.</li><li>• Taza de aprendizaje.</li></ul>

# ■ ¿Qué son los hiperparámetros?

- El ajuste de los HPs tiene gran **impacto en el desempeño** del modelo.
- Los HPs óptimos **difieren para distintos conjuntos de datos** → tienen que ser optimizados para cada conjunto de datos.

# ¿Cómo seleccionamos los hiperparámetros óptimos?

- El proceso de selección de hiperparámetros se conoce como ajuste de hiperparámetros u optimización de hiperparámetros → método para seleccionar los HPs que minimizan el error de generalización.

*Hiperparámetros  $\rightarrow$  min(métricas de performance del modelo)*

Los HPs son un conjunto de valores que minimizan las métricas de performance del modelo.

Si estamos intentando minimizar/maximizar una función...

# ¿Por qué no utilizamos algún método conocido para encontrar el mínimo?

*Hiperparámetros  $\rightarrow$  min(métricas de performance del modelo)*

La métrica que se intenta minimizar depende de:

- La arquitectura del modelo
- Los datos
- Los propios HPs
- Métrica seleccionada



*Superficie de  
respuesta de los  
HPs*

NO ES POSIBLE ASOCIAR UNA FÓRMULA  
MATEMÁTICA QUE REPRESENTA LA SUPERFICIE  
DE RESPUESTA DE LOS HPS  $\rightarrow$  NO ES  
DIFERENCIABLE



## ■ ¿Qué se necesita para poder optimizar los HPs?

- Un espacio de hiperparámetros.
- Un método para seleccionar los candidatos dentro del grupo de HPs a evaluar.
- Un esquema de validación cruzada.
- Una métrica para evaluar la performance.

# Métodos para seleccionar HPs

- Manual Search
- Grid Search
- Random Search
- Sequential Search
- Otros

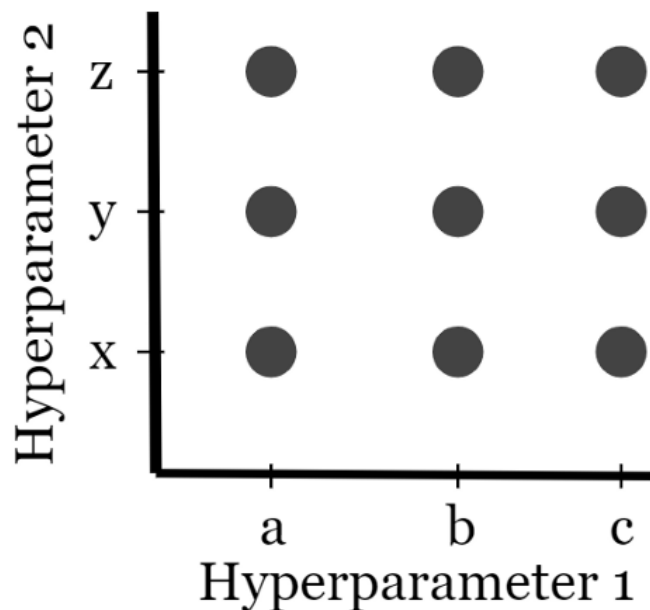
# Métodos clásicos

## Grid Search

Pseudocode

```
Hyperparameter_One = [a, b, c]
```

```
Hyperparameter_Two = [x, y, z]
```

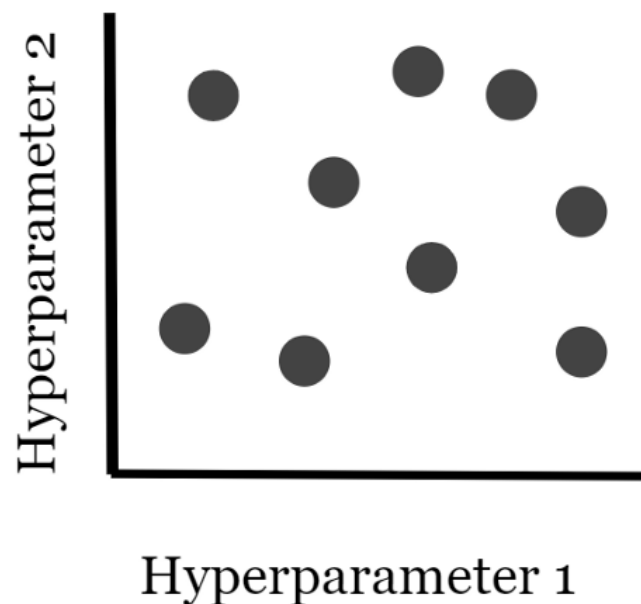


## Random Search

Pseudocode

```
Hyperparameter_One = random.num(range)
```

```
Hyperparameter_Two = random.num(range)
```



Esquema de búsqueda de hiperparámetros por los métodos clásicos Grid Search y Random Search



# Búsqueda secuencial

Son técnicas que seleccionan un conjunto de HPs, evalúan la calidad/performance y en base a ello deciden hacia donde orientar el siguiente muestreo de HPs.

- Es un proceso iterativo y secuencial.
- No es paralelizable en su totalidad.
- Tiene como objetivo probar una menor cantidad de HPs pero hacerlo de manera inteligente sobre los más prometedores.

## ¿Por qué y cuando utilizar búsqueda secuencial para ajustar los HPs?

Cuando la complejidad del modelo que se está entrenando es muy elevada, los métodos clásicos se vuelven muy costosos.

Existe una relación de compromiso entre:

- Tiempo de entrenamiento
- Tiempo de estimación del próximo conjunto de HPs

Tiempo de entrenamiento > Tiempo de estimación del próximo conjunto de HPs



# ¿Cómo podemos implementar una búsqueda secuencial de HPs?



# Optimización bayesiana

Es una estrategia secuencial para **optimización** de funciones de **caja negra** las cuales no asumen ninguna fórmula/función.

Es utilizada comúnmente para optimizar funciones que son **complejas/costosas de evaluar**.

La función objetivo debe poder ser evaluada en puntos arbitrarios.

$$x^* = \arg \max f(x) \quad ; \quad \text{donde } f = \text{superficie de respuesta de los HPs}$$

$x = \text{HPs}$

# Optimización bayesiana

Se trata a  $f$  con una función aleatoria y se realiza una estimación a priori de la distribución sobre ella.



Se evalúan algunos puntos de  $f$ .

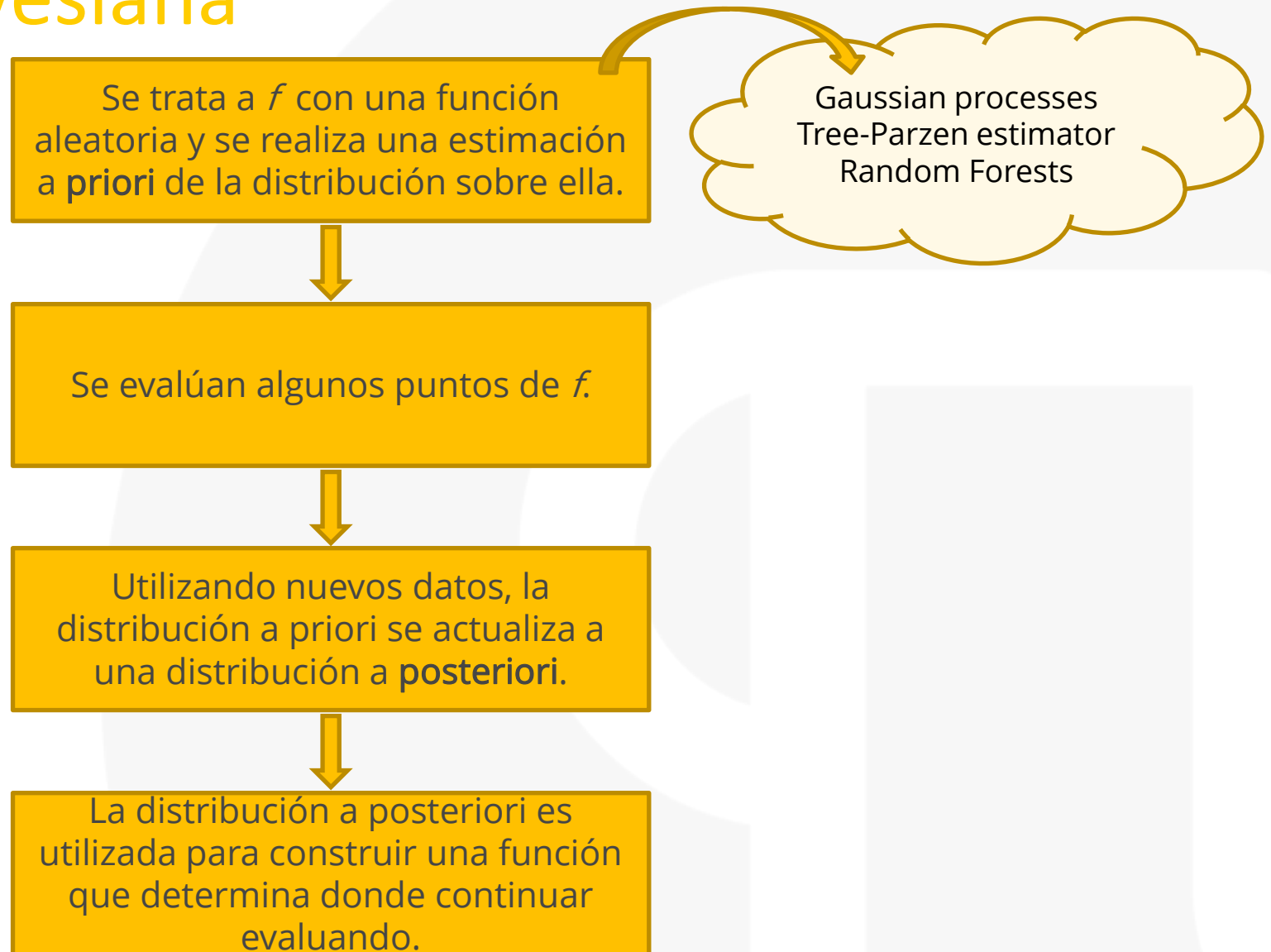


Utilizando nuevos datos, la distribución a priori se actualiza a una distribución a posteriori.

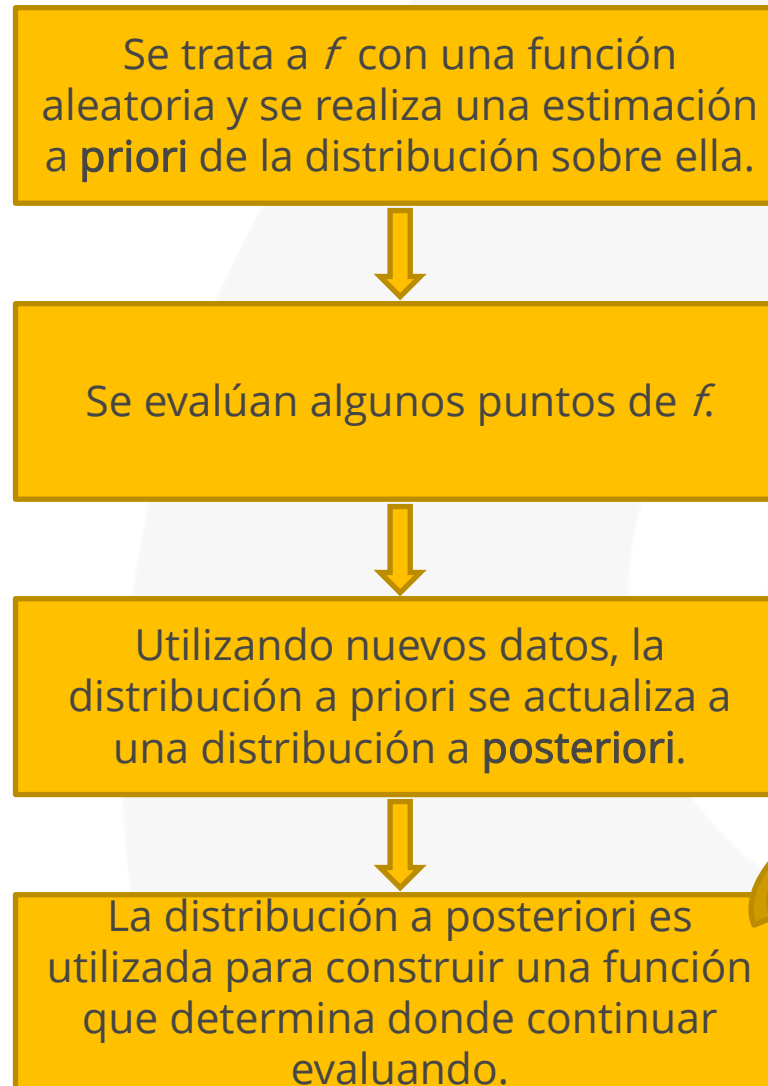


La distribución a posteriori es utilizada para construir una función que determina donde continuar evaluando.

# Optimización bayesiana

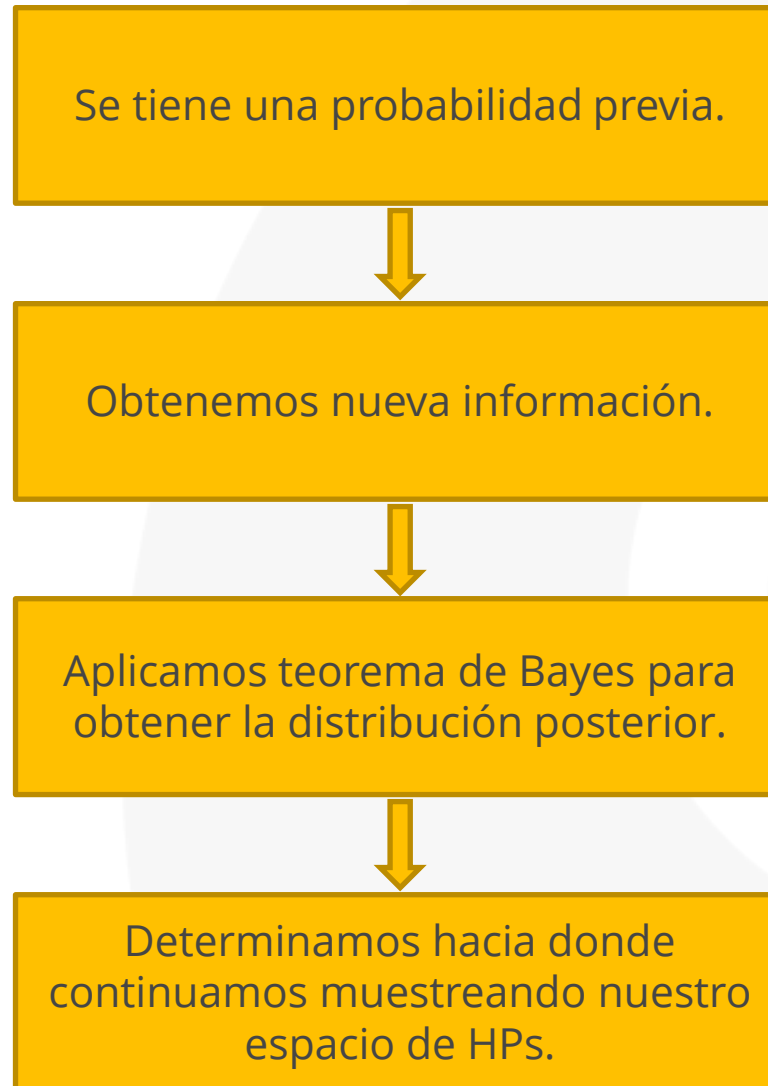


# Optimización bayesiana



Expected Improvement  
Probability of Improvement  
Gaussian process upper  
confidence bound

# Optimización bayesiana





# Teorema de Bayes

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

- A y B son eventos.
- $P(A|B)$  es la probabilidad a posterior, o condicional, de A dada la evidencia B.
- $P(A)$  y  $P(B)$  son las probabilidades marginales de los eventos dados en forma independiente.



# Teorema de Bayes

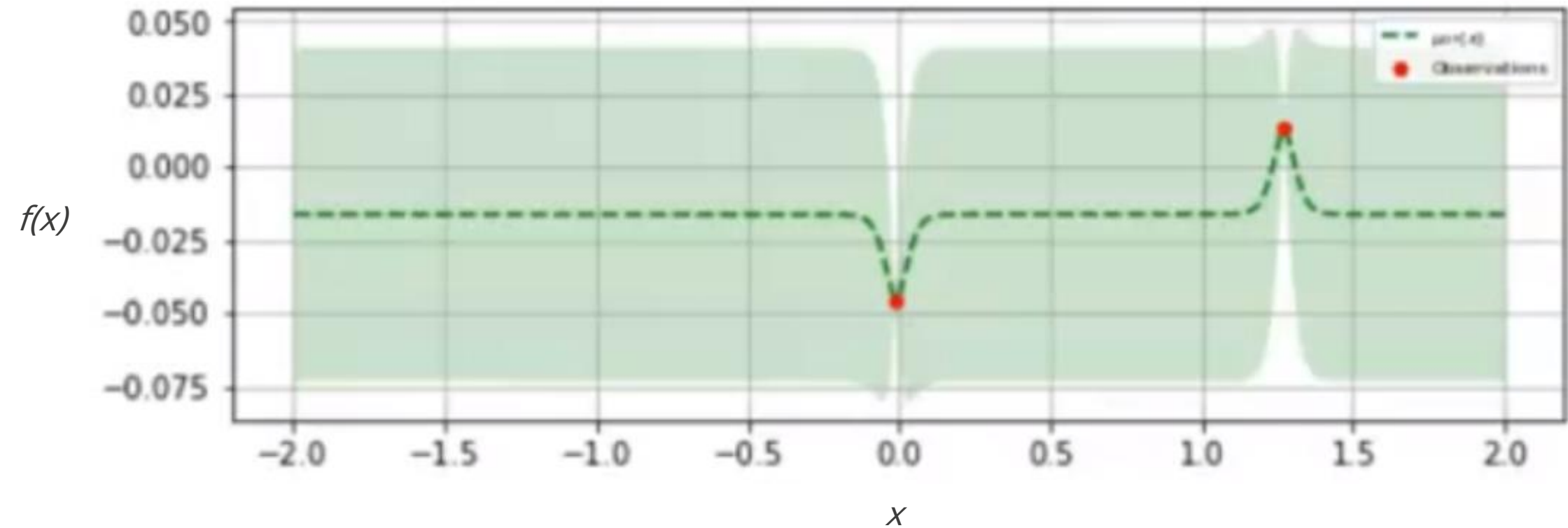
$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

$$P(w|D) = \frac{P(D|w) \times P(w)}{P(D)}$$

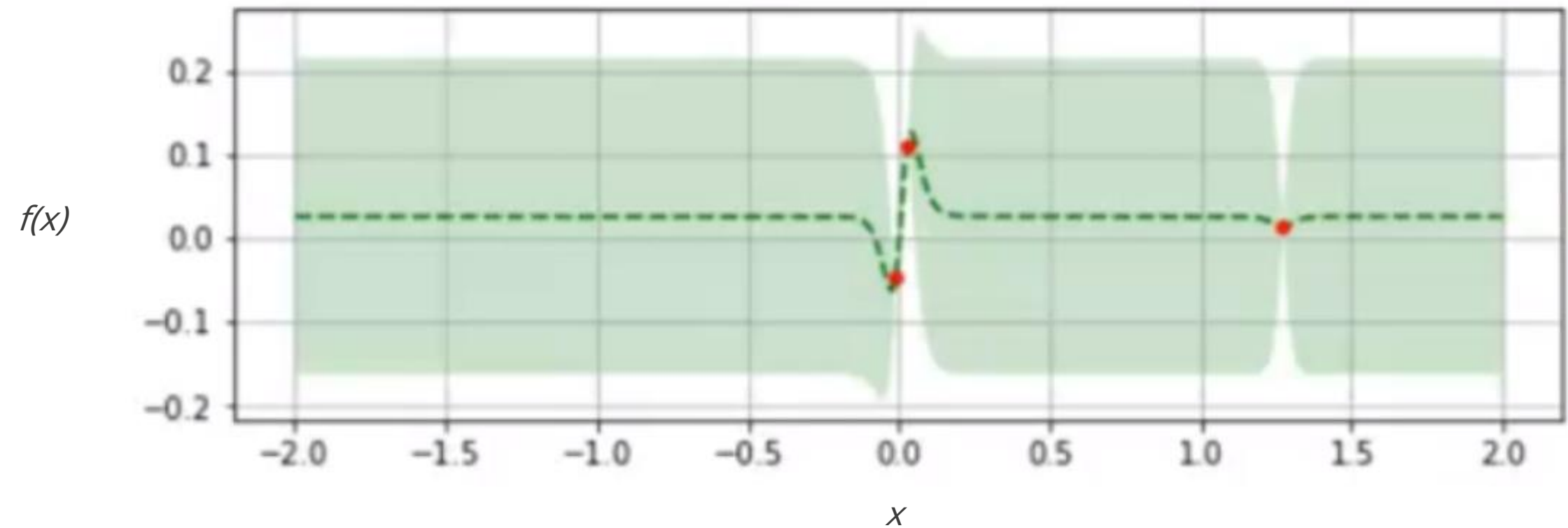
$w: f(x)$

$D: \text{datos}$

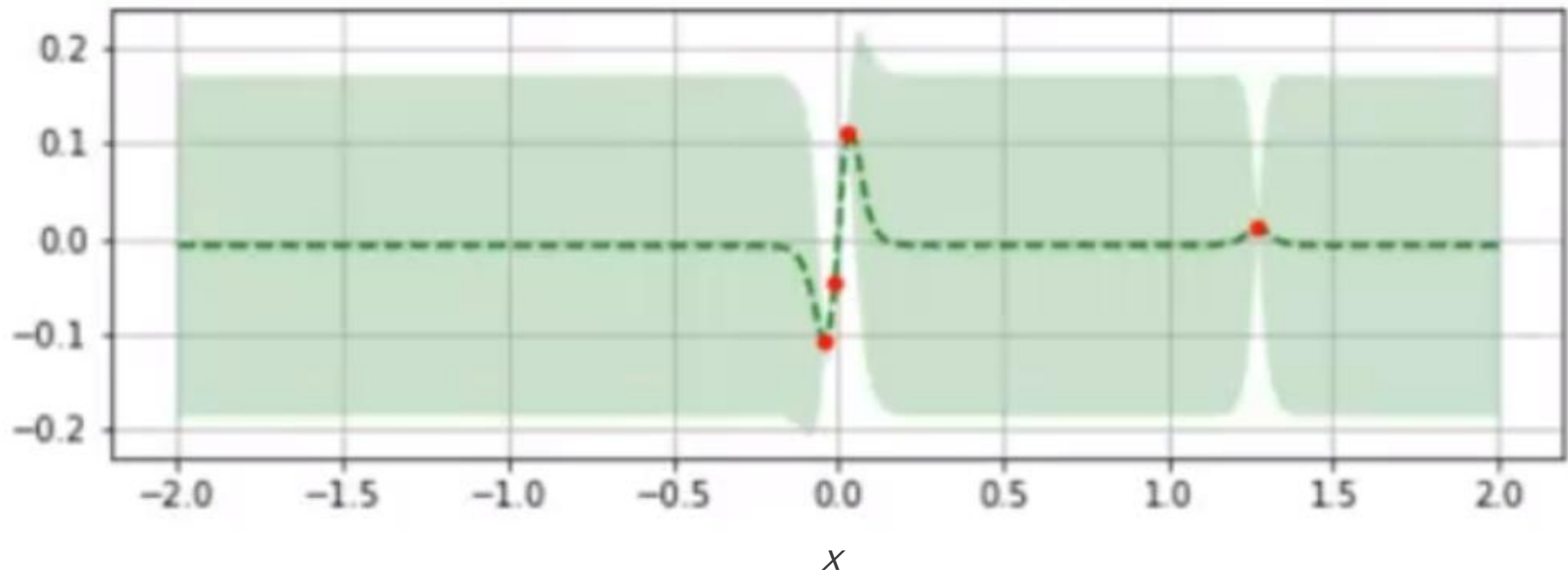
# Optimización de hiperparámetros



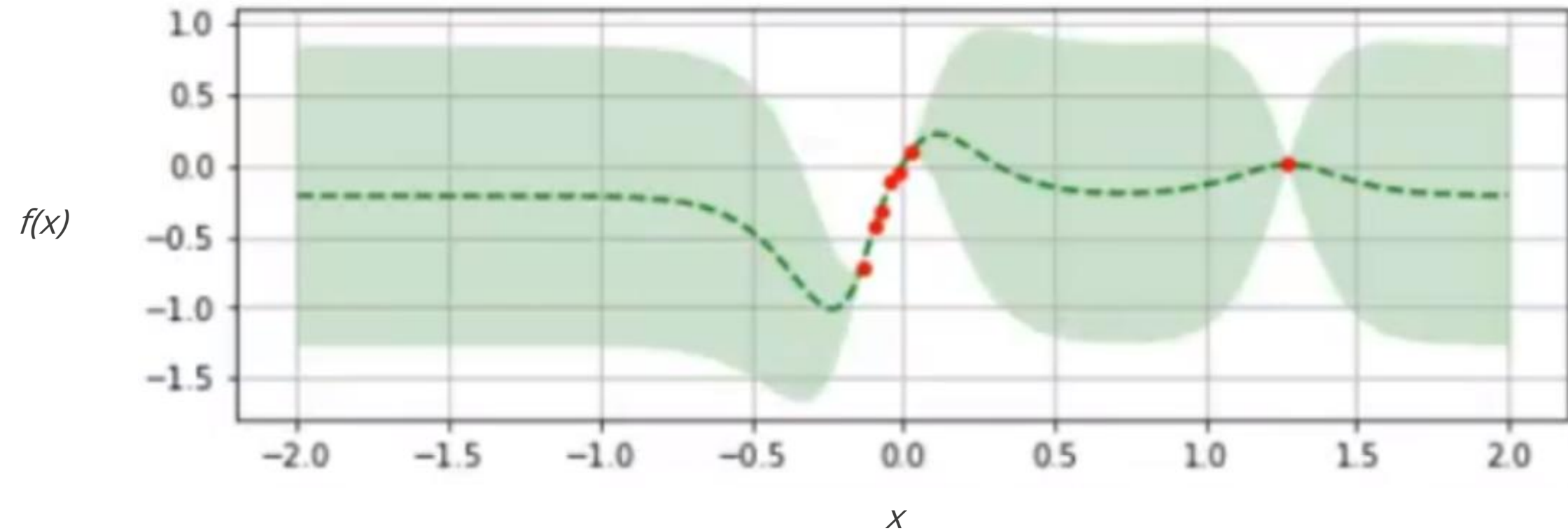
# Optimización de hiperparámetros



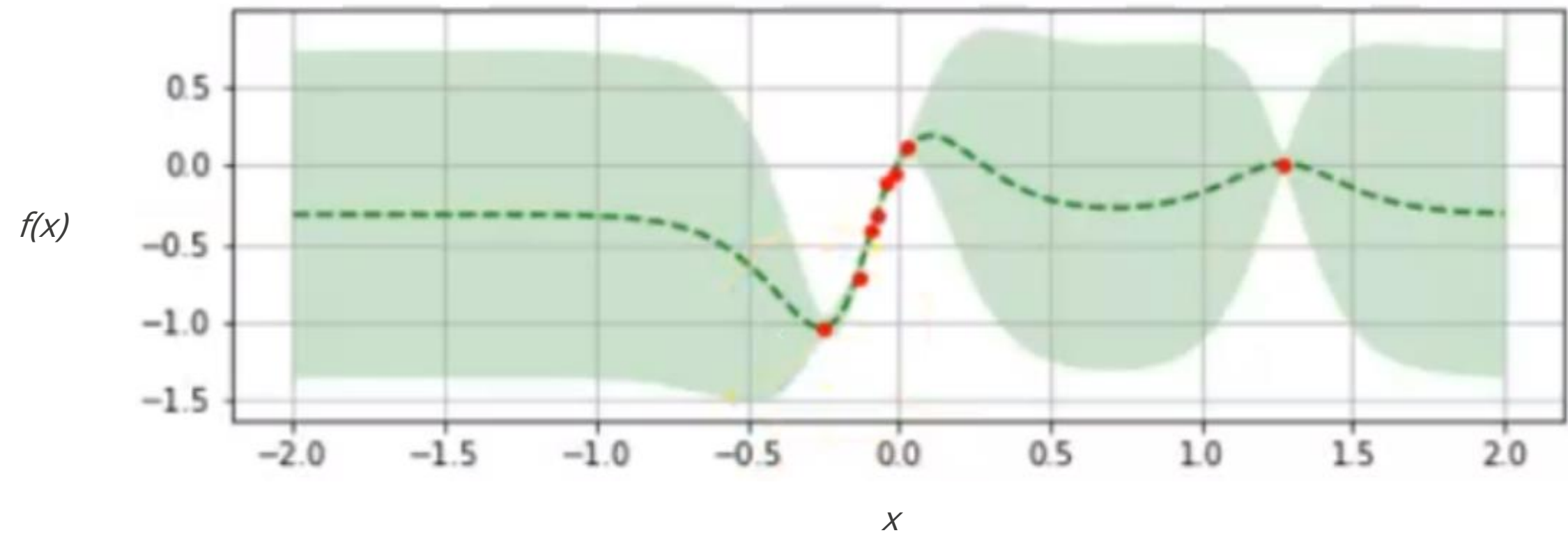
# Optimización de hiperparámetros



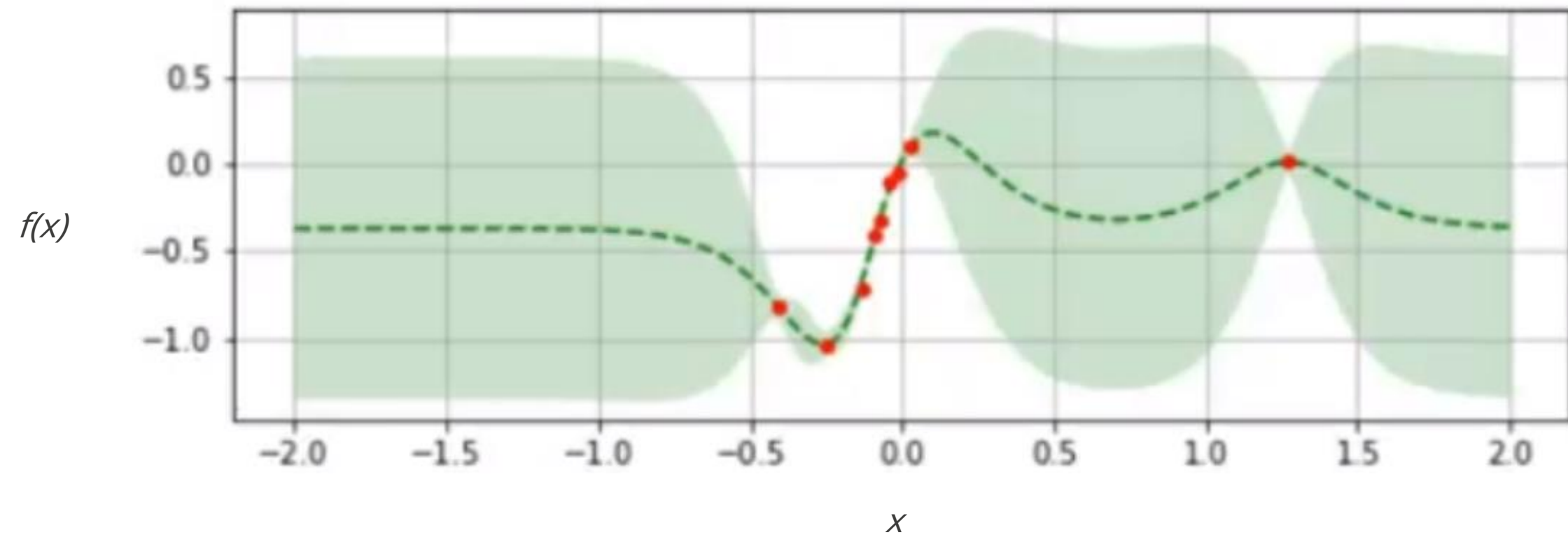
# Optimización de hiperparámetros



# Optimización de hiperparámetros



# Optimización de hiperparámetros





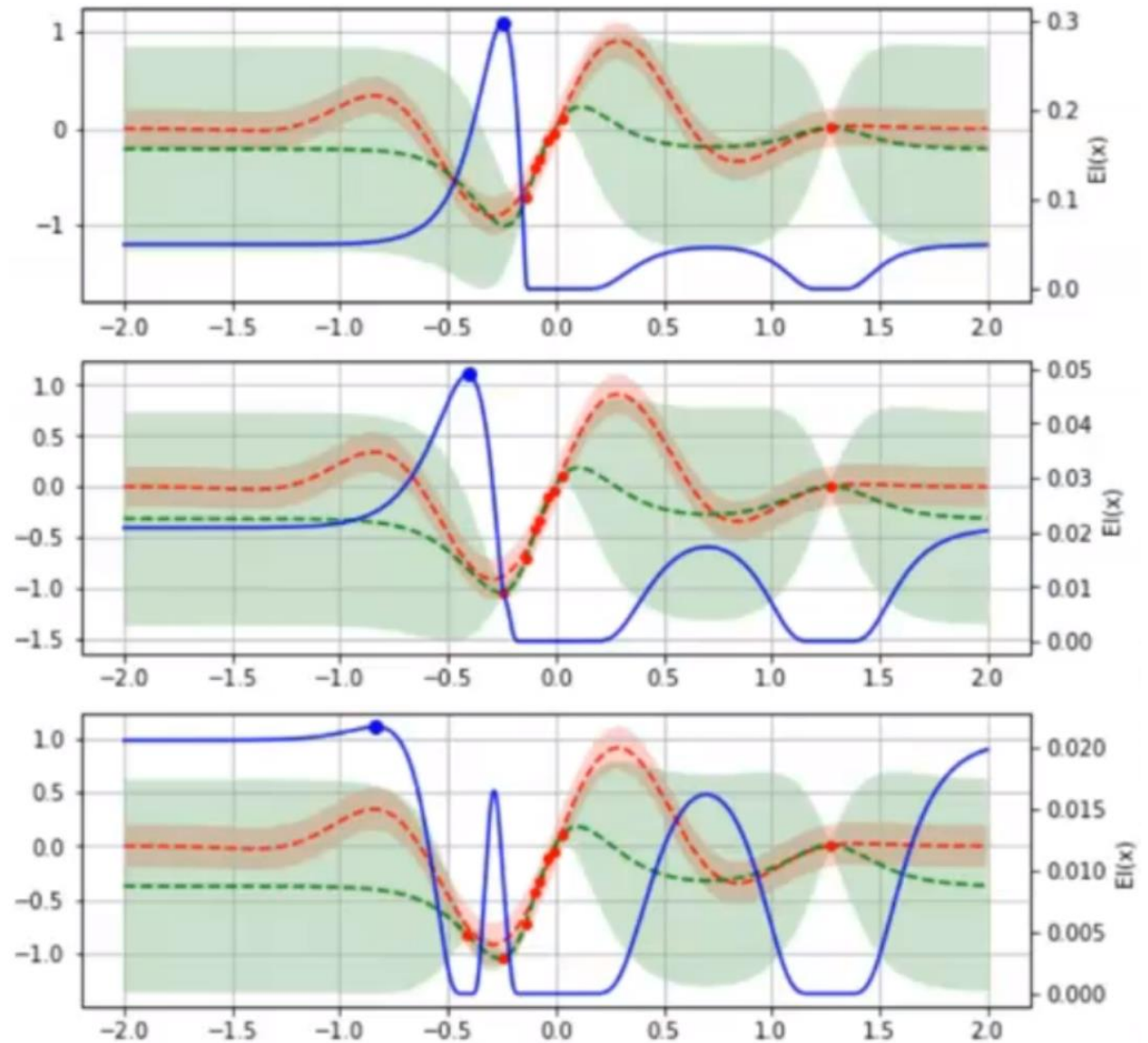
# Optimización de hiperparámetros

- **Modelo sustituto (surrogate model)**

Necesitamos inferir  $f(x)$  de alguna manera y estimar donde se pueden encontrar sus posibles valores

- **Función de adquisición**

Necesitamos una manera de estimar donde continuar muestreando nuestro espacio de HPs



*La clave para este método de optimización es que tanto la función de adquisición como el modelo sustituto, sean mas sencillos de evaluar que la propia  $f(x)$ .*







# TREE PARZEN ESTIMATORS



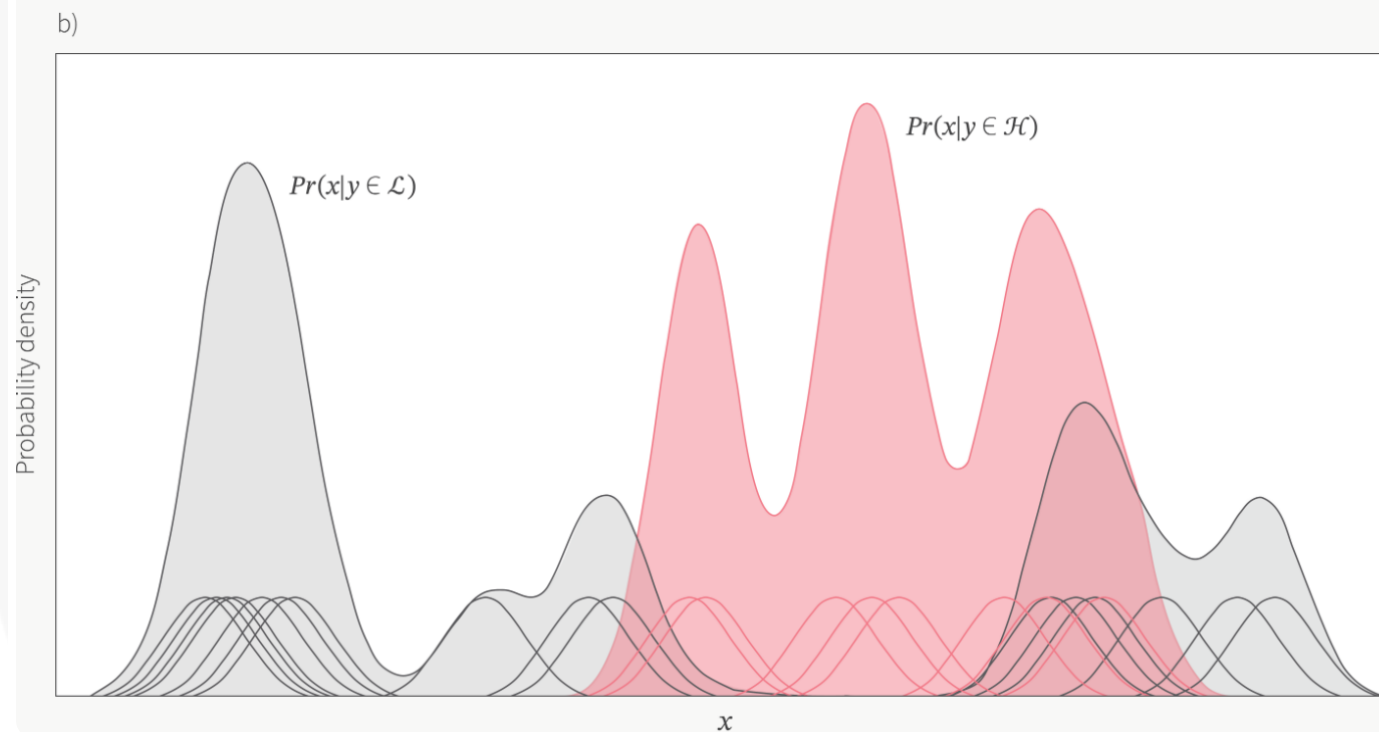
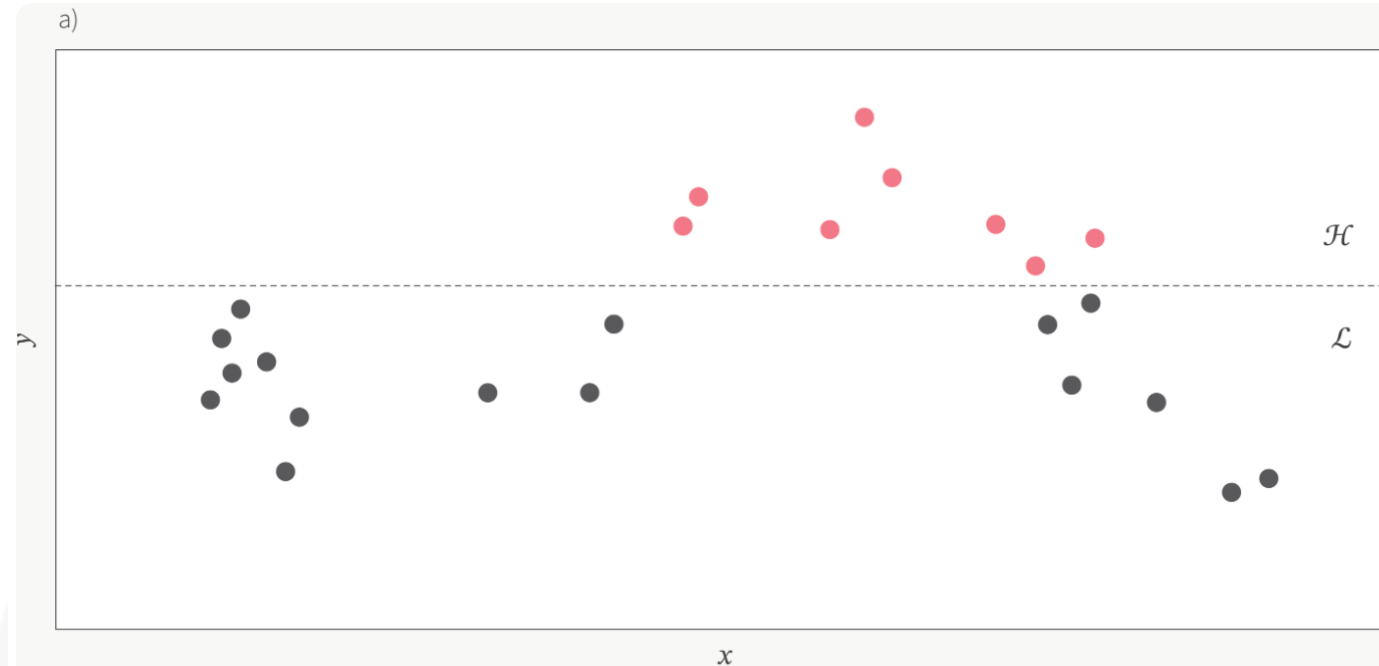
Muestrear  $f(x)$  en varios valores de  $x$ .

Ordenar los valores muestreados en dos grupos, valores pequeños y valores grandes de  $f(x)$ .

Estimar la distribución de los HPs dada por  $f(x)$ . Utilizar Parzen Windows para realizar una estimación de densidad de la distribución.

Seleccionar algunas muestras de  $L(x)$ .

(...)

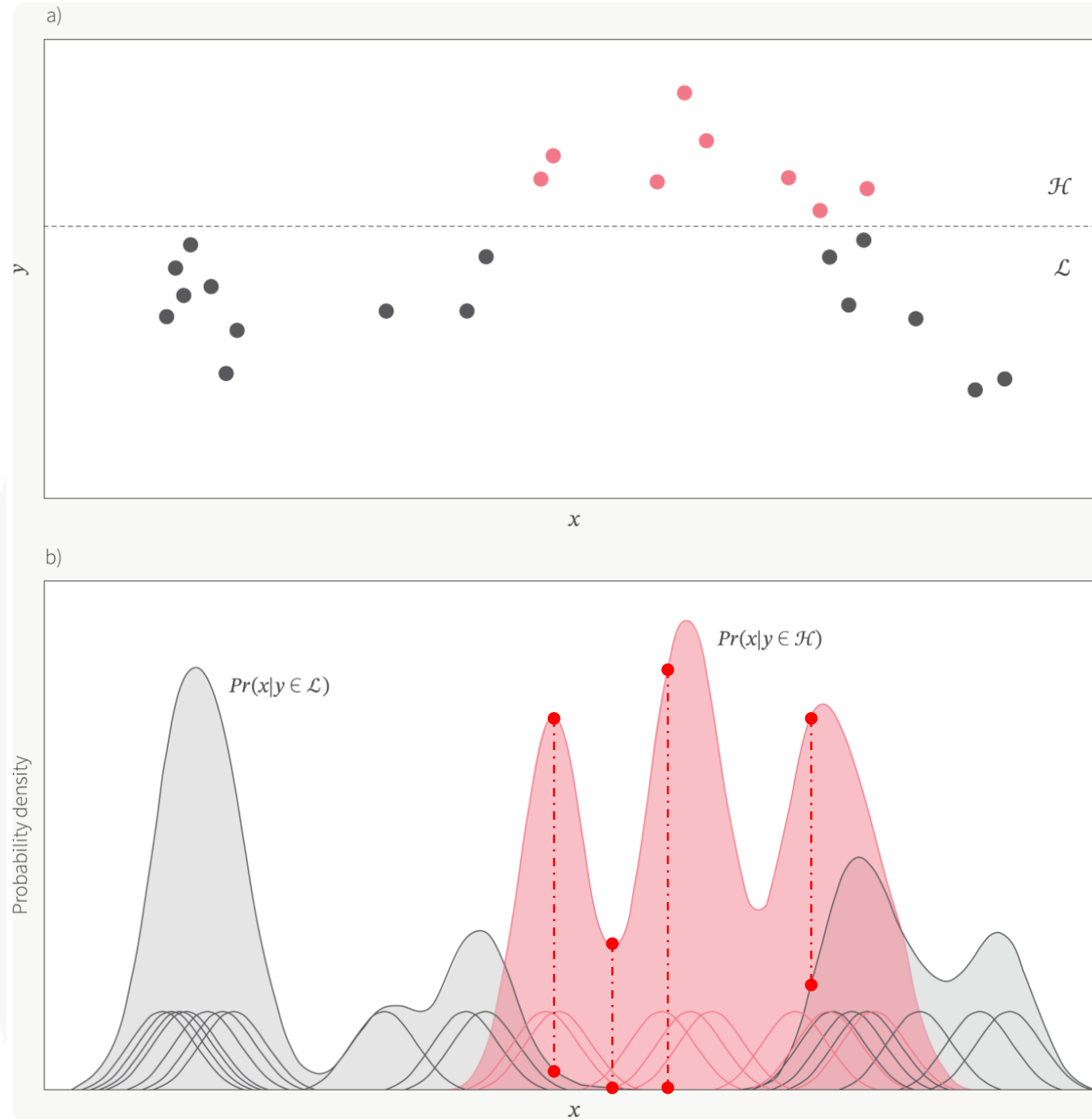


(...)

Calcular la mejora esperada (EI).

Encontrar la máxima EI.

Evaluar  $f(x)$  con nuevos hiperparámetros.





# VEAMOS UNA DEMOSTRACIÓN

Bibliografía/fuentes:

BOREALIS AI [<https://www.borealisai.com/research-blogs/tutorial-8-bayesian-optimization/>]

Curso: Hyperparameter Optimization for Machine Learning de Udemy