

Ajuste de

hiperparámetros

M.Sc. Angelo Jonathan Diaz Soto

2025

(+51) 976 760 803 Data Science – Business Intelligence – Big Data – Machine Learning – Artificial Intelligence – Innovation and
www.datayanalytics.com info@datayanalytics.com

Contenido



1 Introducción



Enfoques de optimización

2

Grid search

3

4

Random search

5

Optimización bayesiana

6

Referencias

67

Introducción



Los hiperparámetros son parámetros de aprendizaje automático que se eligen antes de que el proceso de aprendizaje automático comience. Son **ajustables** y pueden afectar directamente a la forma en la que se entrena un modelo de aprendizaje automático.

Algunos de estos



son parámetros de aprendizaje automático que se eligen antes de que el proceso de aprendizaje automático comience. Pueden afectar directamente a la forma en la que se entrena un modelo de aprendizaje automático.

hiperparámetros son:

- Número de épocas.
- Impulso
- Constante de regularización.
- Número de ramas en un árbol de decisión.
- Número de clústeres en un algoritmo de agrupamiento. Descenso de gradientes. ■
- Tamaño del kernel.
- Tamaño del lote. Etc.

■

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Introducción



- Muchos
lineal

modelos, entre ellos la regresión con regularización Ridge, contienen parámetros que no pueden “aprenderse” a partir de

los datos de entrenamiento y, por lo tanto, deben de ser establecidos por el analista.

- A estos se les conoce como hiperparámetros.
- Los resultados de un modelo pueden depender en gran medida del valor que tomen sus hiperparámetros, sin embargo, **no se puede conocer de antemano cuál es el adecuado.**

En la práctica los especialistas en machine learning ganan intuición sobre qué valores pueden funcionar mejor en cada problema, **no hay reglas fijas.**

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Introducción



- La diferencia entre un parámetro y un

hiperparámetro es que el segundo lo utilizan los algoritmos de aprendizaje cuando este apenas está aprendiendo, pero no son partícipes en el modelo que resulta de este aprendizaje.

- **Los hiperparámetros** son flexibles, ya que se refieren a cualquier elemento en el machine learning.
- **Los parámetros**, por su lado, son características inherentes al modelo. Esto quiere decir que se aprenden o se estiman a partir de los datos brindados durante el entrenamiento, ya que el algoritmo que se use intentará aprender el mapeo entre las características de entrada y los objetivos.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com



Algunos ejemplos de parámetros

Coeficientes de modelos de regresión lineal y logística.

1 Pesos y
sesgos.

2



Centroides del clúster en un agrupamiento.

3

- Los parámetros, por tanto, son los valores que el algoritmo de aprendizaje puede cambiar de forma deliberada a medida que aprende y estos valores se van viendo afectados por la elección de los hiperparámetros que se proporcionen.
- En resumen, **los hiperparámetros se establecen antes de que comience**

el entrenamiento y el algoritmo de aprendizaje los usa para aprender los parámetros.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Introducción



La forma más común de encontrar los valores óptimos es probando diferentes posibilidades.

☐ 1 Escoger un
o los

☐ **grid search:**
exhaustiva

usuario.

☐ **random**

conjunto de valores para el
hiperparámetros.

se hace una búsqueda
sobre un conjunto de valores
previamente definidos por el

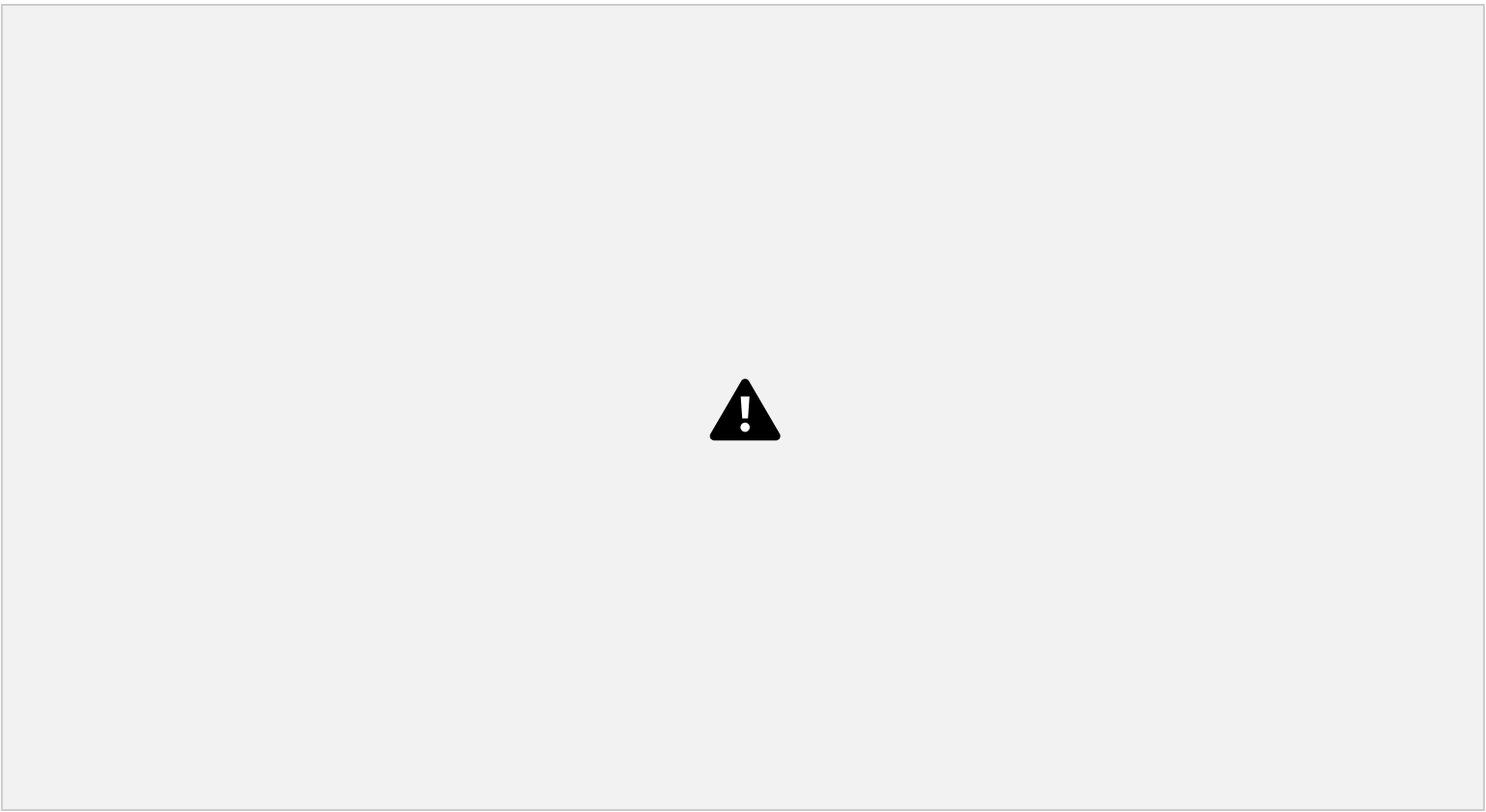
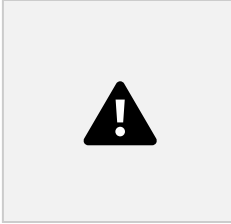
search: se evalúan valores aleatorios



dentro de unos límites definidos por el usuario.

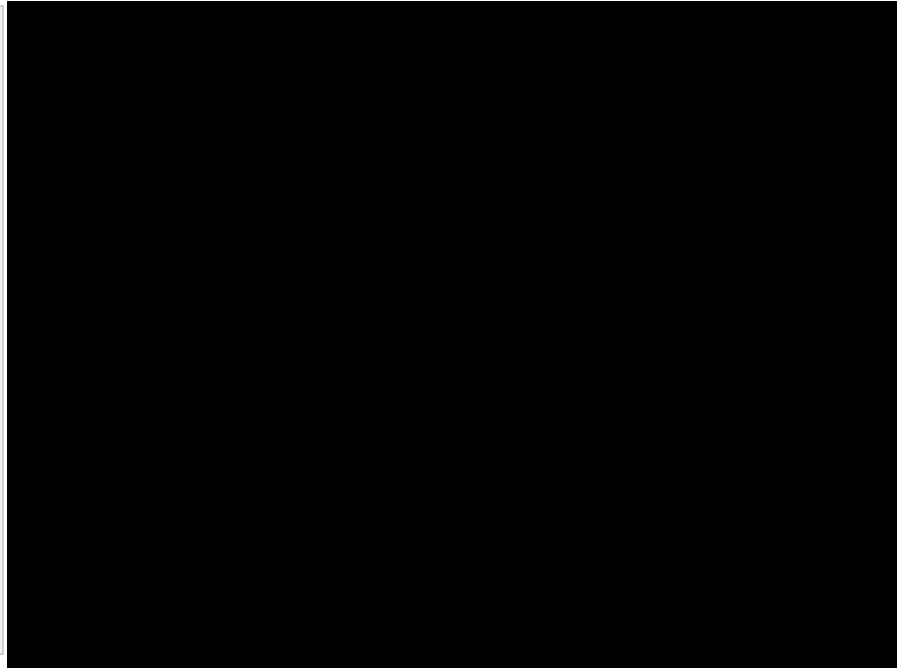
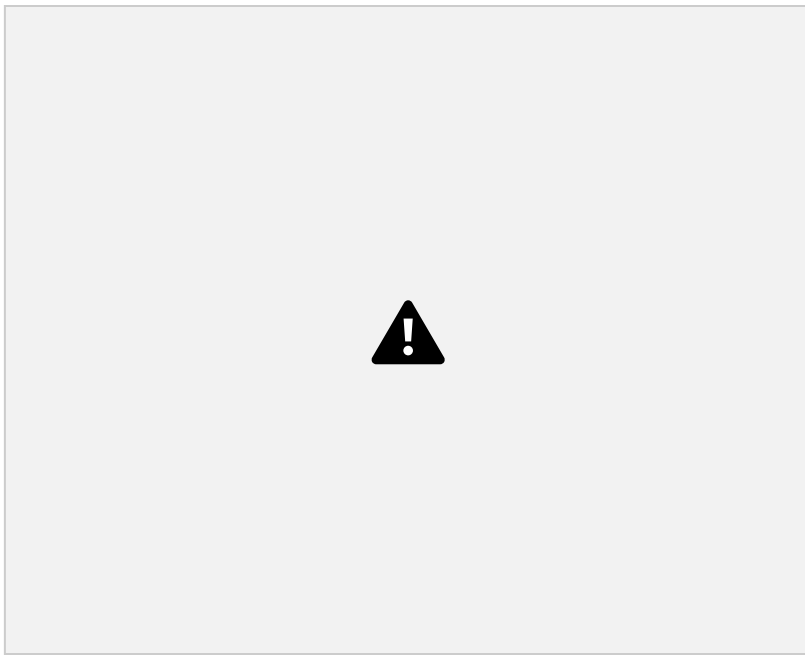
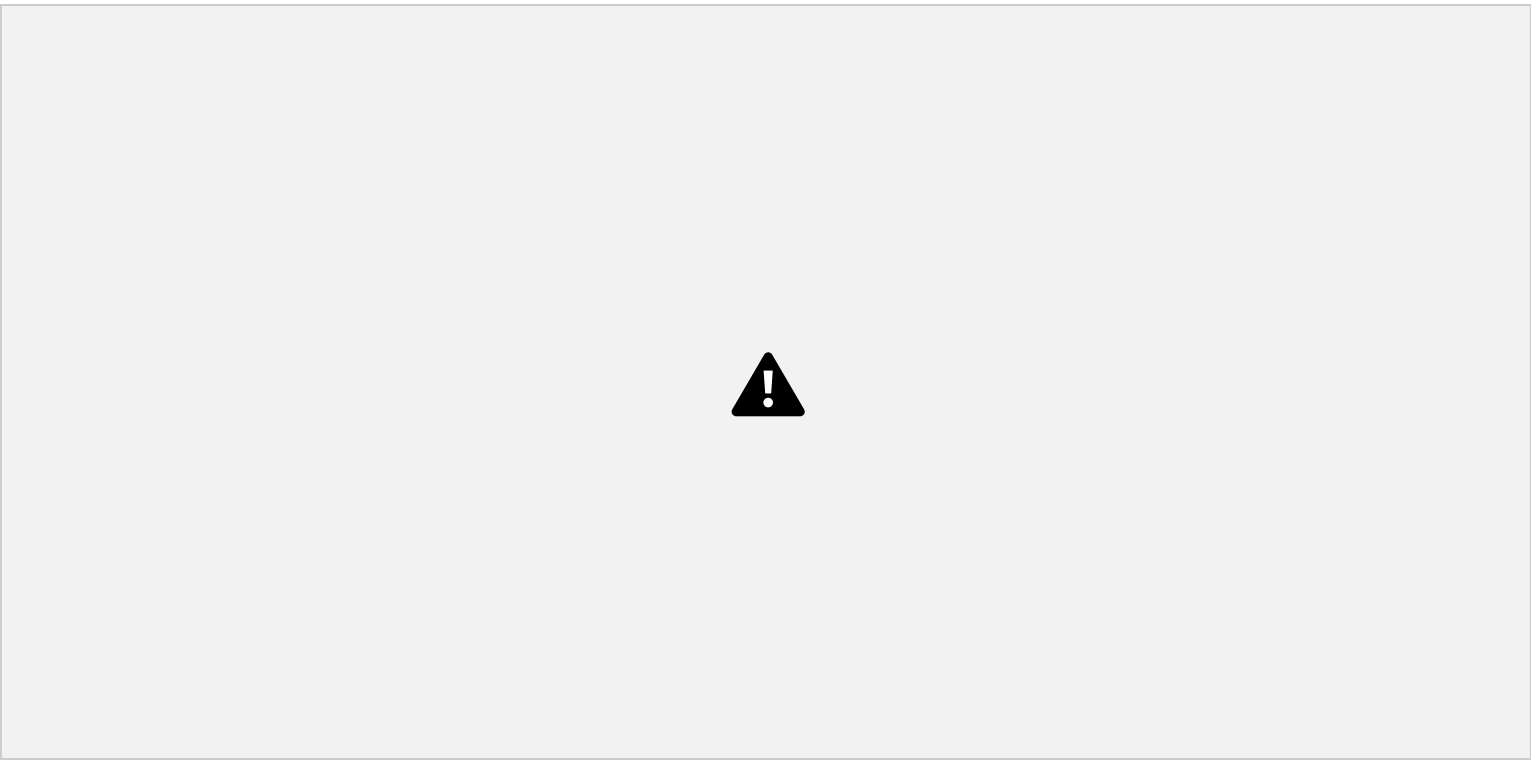
- 2 Para cada valor (combinación de valores si hay más de un hiperparámetro), entrenar el modelo y estimar su error mediante un método de validación.

- 3 Finalmente, ajustar de nuevo el modelo, esta vez con todos los datos de entrenamiento y con los mejores hiperparámetros encontrados.





(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com



Grid search



- El valor
- Esta en



valores para dos o más parámetros,

método consiste en probar diferentes valores y luego elegir el que da la mejor puntuación. técnica se conoce como búsqueda cuadrícula (Grid search). Si tuviéramos que seleccionar los

- evaluaríamos todas las

combinaciones de los conjuntos de valores formando así una cuadrícula de valores.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

¿Cómo funciona el grid search en Python?

search en Python funciona por

medio
la

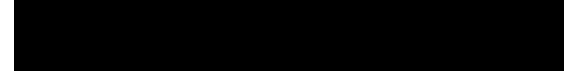
de

 El grid



comparación: se toman dos valores de un conjunto pequeño de valores para dos o más parámetros, se evalúan todas las combinaciones posibles y con ellas se forma una cuadrícula de valores.

El grid search en Python ayuda a la
■ **optimización de hiperparámetros.**



■ El

el
es

■ Este

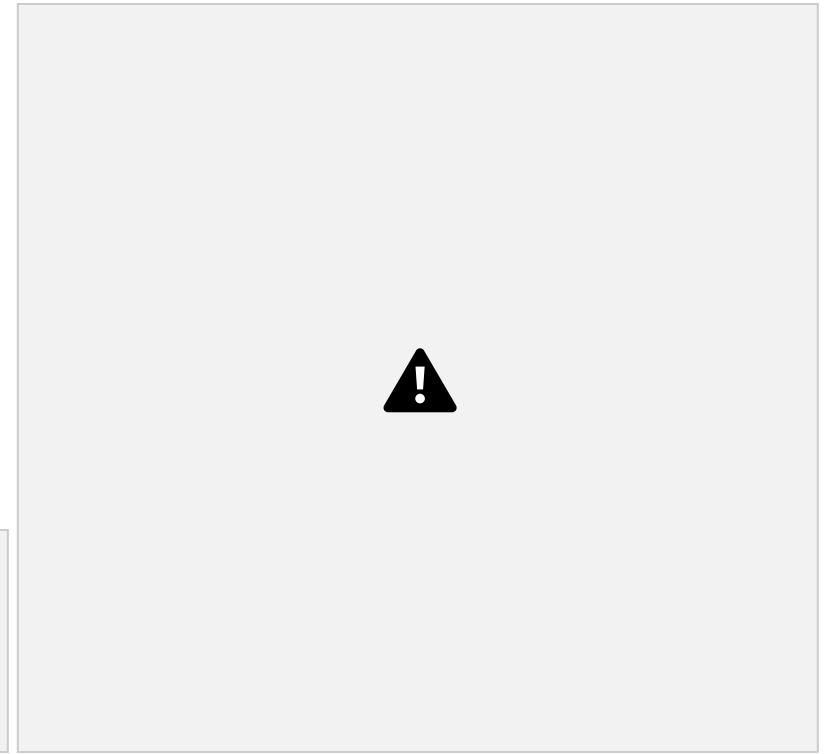
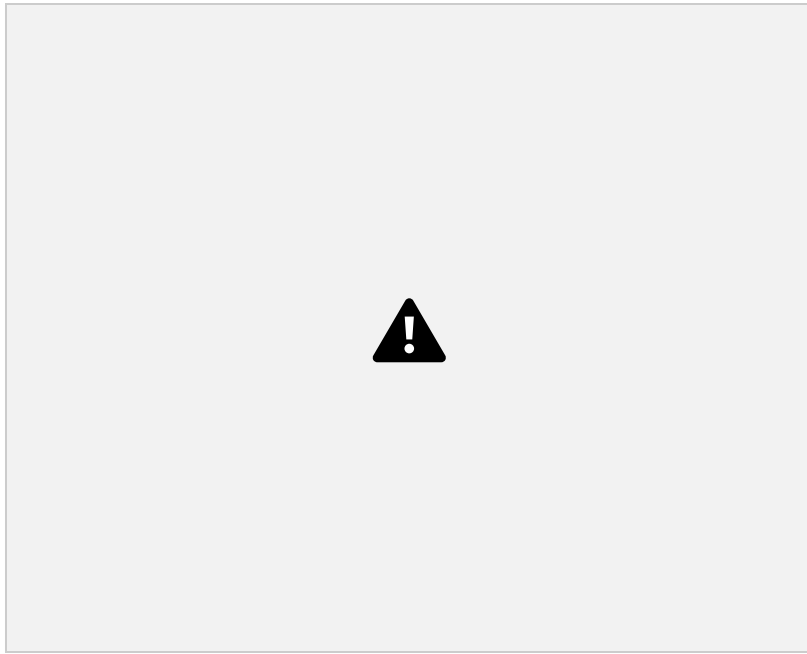


RandomizedSearch es útil cuando se tienen muchos parámetros para probar y tiempo de entrenamiento muy largo.

método realiza una búsqueda aleatoria sobre parámetros,

donde cada configuración se muestrea a partir de una distribución sobre posibles valores de parámetros.

- Intuitivamente, parece que usar un grid search en Python sea lo más lógico; no obstante, los resultados del RandomizedSearch pueden funcionar mejor.



Optimización bayesiana



hiperparámetros consiste en crear un modelo probabilístico en el que la función de validación etc).

- Con esta búsqueda iteración interés.

- El de

hiperparámetros con las que se evalúa el modelo, eligiendo únicamente los mejores candidatos.

- Esto significa que, la ventaja frente a las otras estrategias mencionadas, se maximiza cuando el espacio de búsqueda es muy

- La optimización bayesiana de objetivo es la métrica de del modelo (rmse, auc, precisión,

estrategia, se consigue que la se vaya redirigiendo en cada hacia las regiones de mayor

objetivo final es reducir el número de combinaciones de

amplio o la evaluación del modelo es muy lenta.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com



Optimización bayesiana

- La
a una



optimización bayesiana pertenece a una clase de algoritmos de optimización basada en modelos secuenciales (SMBO) que permiten utilizar los resultados de nuestra iteración anterior para mejorar nuestro método de muestreo del siguiente experimento.

- La optimización bayesiana funciona construyendo una distribución

posterior de funciones (proceso gaussiano) que describe mejor la función que se quiere optimizar.

- A medida que aumenta el número de observaciones, la distribución posterior mejora, y el algoritmo se vuelve más seguro de qué regiones del espacio de parámetros vale la pena explorar y cuáles no.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com







Referencias Bibliográficas

■ An

(2013).

■ Applied
Kuhn



Introduction to Statistical
Learning: with Applications in R
James G., Witten D., Hastie T.,
Tibshirani R.

Predictive Modeling by Max
and Kjell Johnson. ■

Introduction to Statistical

Learning, Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani.

- Introduction to Machine Learning with Python: A Guide for Data Scientists.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com





(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com