

Introducción a la Ciencia de Datos

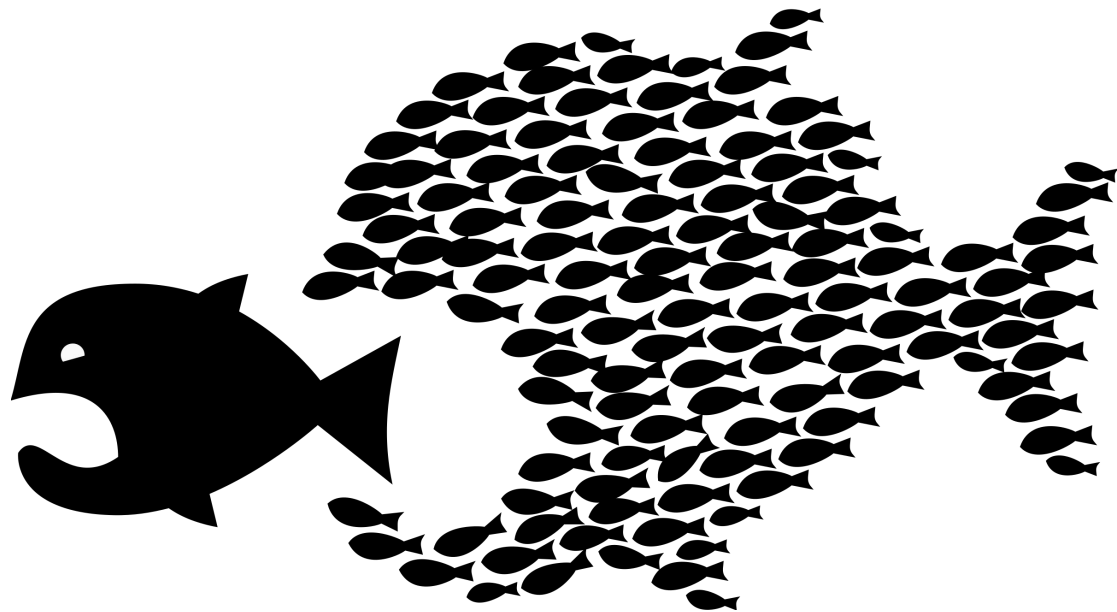
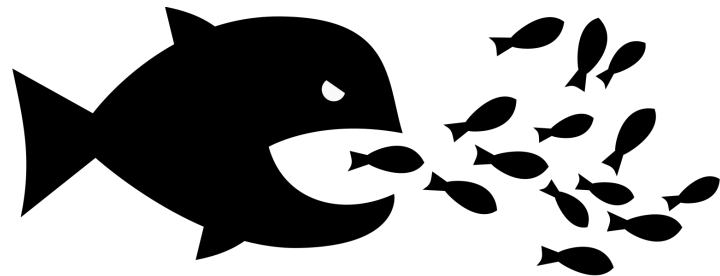
Maestría en Ciencias
de la Computación

Dr. Irvin Hussein López Nava

The slide features several decorative squares in various shades of blue and white, scattered across the background. A large blue square with the number '05' is positioned in the upper center. Other squares of different sizes and colors are located in the corners and along the right side of the slide.

05

Meta-aprendizaje



5.1 Ensembles

A woman with dark hair, wearing an orange sweater, is shown in profile, looking upwards and to the right. She is holding a tablet in her left hand and reaching out with her right hand towards a glowing digital globe. The background is a complex, futuristic interface with various data visualizations, including line graphs, bar charts, and circular progress indicators, all in shades of blue and white. The text "¿Qué es el Aprendizaje basado en Ensamblajes?" is overlaid on the left side of the image.

¿Qué es el Aprendizaje basado en Ensamblajes?

Definiciones

ChatGPT

Es una técnica en el campo del aprendizaje automático en la que se combinan múltiples modelos para mejorar el rendimiento y la precisión de las predicciones en comparación con un solo modelo. Se basa en la idea de que la combinación es más robusta y generalizable que un solo modelo.

<https://chat.openai.com/>

Wikipedia

En estadística y aprendizaje automático, los métodos de ensamble utilizan múltiples algoritmos de aprendizaje para obtener un rendimiento predictivo mejor que el que podría obtenerse con cualquiera de los algoritmos de aprendizaje constituyentes por sí solos.

https://en.wikipedia.org/wiki/Ensemble_learning

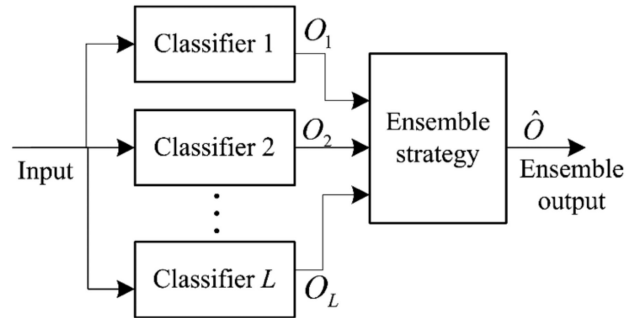
Gemini

Es un enfoque de aprendizaje automático que consiste en combinar los resultados de varios modelos de aprendizaje automático para mejorar el rendimiento general. Al combinar los resultados de los modelos, se puede reducir el sesgo y mejorar la robustez.

<https://gemini.google.com/>

¿Qué es?

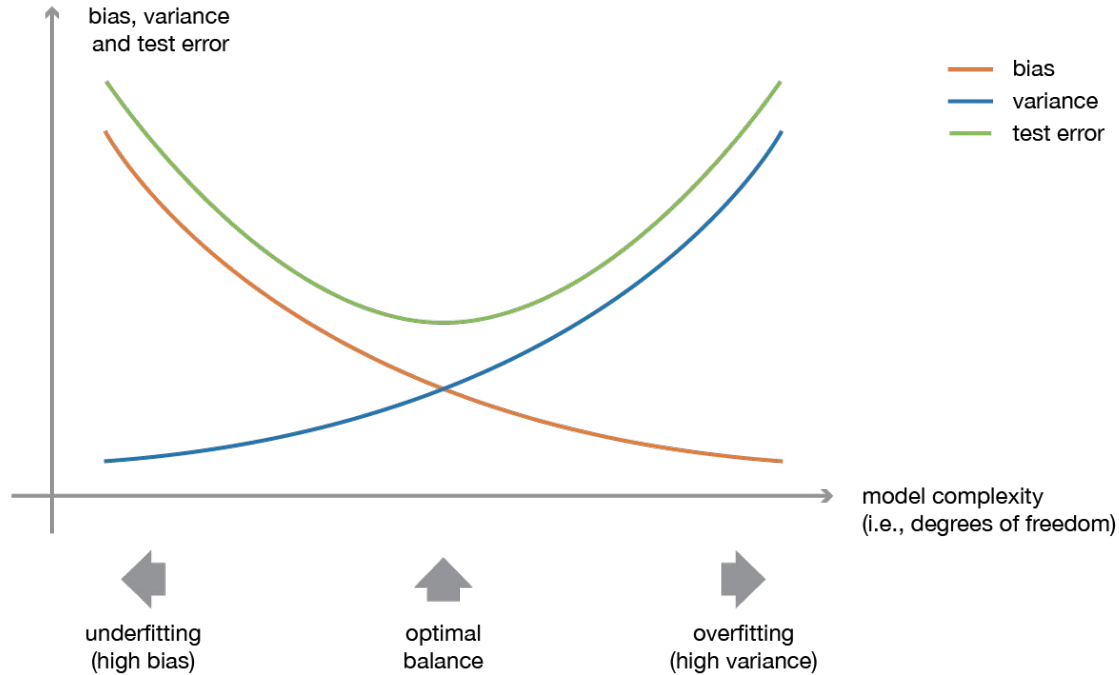
- Es un paradigma de aprendizaje automático que utiliza múltiples algoritmos de aprendizaje (a menudo denominados "*weak learners*" o "aprendices débiles") para obtener un rendimiento predictivo mejor que el que podría obtenerse con cualquiera de los constituyentes por sí solos.



Single weak learner

- **Un dato**: la elección del algoritmo de inferencia es extremadamente importante para tener alguna posibilidad de obtener buenos resultados.
- Esta elección puede depender de muchas variables: cantidad de datos, dimensionalidad del espacio, hipótesis de distribución...
- Un **sesgo** y una **varianza** bajos, aunque la mayoría de las veces varían en direcciones opuestas, son las dos características fundamentales que se esperan de un modelo.
- Se trata del conocido equilibrio (*tradeoff*) entre sesgo y varianza.

Bias-variance tradeoff



Weak learners

- En el aprendizaje por ensambles, los modelos de aprendizaje "débiles" (o base) pueden utilizarse como bloques de construcción para diseñar modelos más complejos combinando varios de ellos.
- La mayoría de las veces, estos modelos no funcionan bien por sí solos, ya sea porque tienen un sesgo elevado o demasiada varianza para ser robustos.
- Entonces, la idea de los métodos de ensamble es intentar reducir el sesgo y/o la varianza de esos aprendices débiles combinando varios de ellos para crear un aprendiz fuerte (ensamble) que consiga mejores resultados.

Ensemble theory

- Los algoritmos de aprendizaje buscan a través de un espacio de hipótesis para encontrar una hipótesis adecuada para un problema concreto; aunque exista, es muy **difícil encontrarla**.
- Los **ensambles** combinan múltiples hipótesis para formar una hipótesis (con suerte) mejor.
- Evaluar la predicción de un **ensamble** requiere más cálculo que evaluar modelos individuales.
- En este aprendizaje se suelen utilizar algoritmos rápidos, como los árboles de decisión, para compensar la demanda computacional.

Ensemble theory

- Un **ensamble** es en sí mismo un algoritmo de aprendizaje supervisado, porque puede entrenarse y luego utilizarse para hacer predicciones.
- Por lo tanto, el **ensamble** entrenado representa una única hipótesis.
- Esta hipótesis no está necesariamente **contenida** en el espacio de hipótesis de los modelos a partir de los cuales se construye.
- Así, se puede demostrar que los **ensambles** tienen más flexibilidad en las funciones que pueden representar.
- Sin embargo, pueden **sobreajustar** los datos **más** que un modelo único.

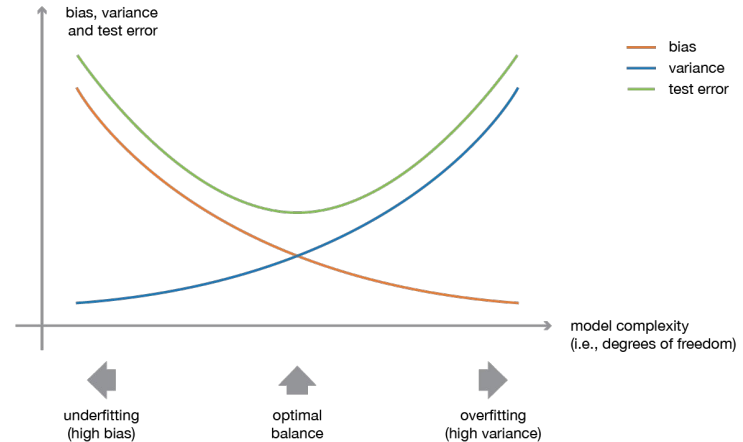
Combine weak learners

- Un método basado en aprendizaje de ensambles necesita **modelos base** para ser agregado.
 - Por lo general, se utiliza un **único algoritmo** de aprendizaje de base, de modo que los aprendices débiles homogéneos se entrenan de distintas formas, por lo que el modelo de **ensamble** es "homogéneo".
 - Sin embargo, también existen algunos métodos que utilizan **diferentes** tipos de **algoritmos** de aprendizaje de base, construyendo un modelo de **ensamble** "heterogéneo".

Combine weak learners

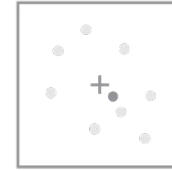
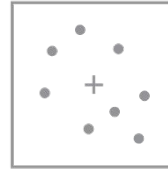
Para los **modelos base** con **bajo sesgo** pero **alta varianza**, se utiliza un **método de agregación** que tiende a reducir la varianza.

Mientras que para los **modelos base** con **baja varianza** pero **alto sesgo**, se utiliza un **método de agregación** para reducir el sesgo.





predictions
visualisation

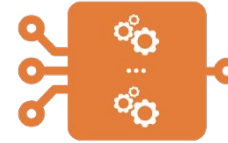


models
representation



several single weak learners
with **low bias but high variance**

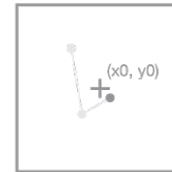
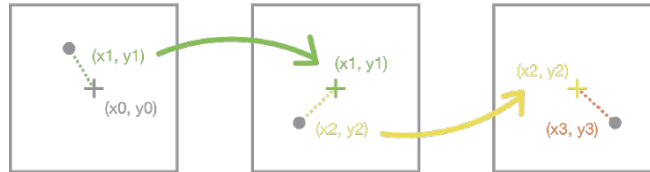
"average"
weak learners



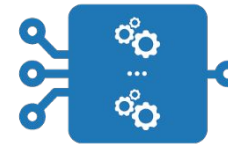
ensemble model with a **lower variance**
than its components

LOW BIAS HIGH VARIANCE WEAK LEARNERS

LOW VARIANCE HIGH BIAS WEAK LEARNERS



"compose"
weak learners



several single weak learners with **high bias but low variance**:
each model target the error of the previous one

ensemble model with a **lower bias**
than its components

Tipos comunes de ensambles

- Bayesian model combination.
- Boosting:
 - homogeneous weak learners
- Bootstrap aggregating (bagging):
 - homogeneous weak learners
- Stacking:
 - heterogeneous weak learners

Bayesian model combination

Basado en el modelo Bayes de promedio y el clasificador óptimo de Bayes.

Clasificador óptimo de Bayes:

- Ensamble (mediante la regla de Bayes) de todas las hipótesis en el espacio de hipótesis.

$$y = \arg \max_{v_j \in V} \sum_{h_i \in H} P(v_j|h_i)P(D|h_i)P(h_i)$$

donde y es la clase predicha, V es el conjunto de todos los valores objetivo, H es el espacio de hipótesis, P se refiere a una probabilidad y D son los datos de entrenamiento.

- Como ensamble, el clasificador óptimo de Bayes representa una hipótesis que no está necesariamente en H .

Bayesian model combination

- En promedio, se obtiene la hipótesis óptima en el espacio de ensamble.
 - el espacio de todos los ensambles posibles formados sólo por hipótesis en H .
- Esta fórmula puede reescribirse utilizando el teorema de Bayes, que dice que la **probabilidad a posteriori** es proporcional a la probabilidad condicional por la probabilidad *a priori*:

$$P(h_i|D) \equiv P(D|h_i)P(h_i)$$

por lo tanto

$$y = \arg \max_{v_j \in V} \sum_{h_i \in H} P(v_j|h_i)P(h_i|D)$$

Bayesian model combination

- Modelo Bayes promedio:
 - Se aproxima al clasificador óptimo de Bayes.
 - Muestreo del espacio de hipótesis.
 - Muestreo Monte Carlo.
 - Tiende a promover el **sobreajuste**.
 - En la práctica funciona peor que otras técnicas (boosting/bagging).

Boosting

- El *boosting* consiste en construir un conjunto de forma incremental entrenando a cada nuevo modelo para enfatizar los casos que los modelos anteriores clasificaron mal.
- En algunos casos, se ha demostrado que *boosting* ofrece mayor precisión que *bagging*, pero también tiende a **sobreajustar** los datos de entrenamiento.
- La implementación más común de *boosting* es **Adaboost**.
- Los resultados se combinan siguiendo una estrategia determinista.

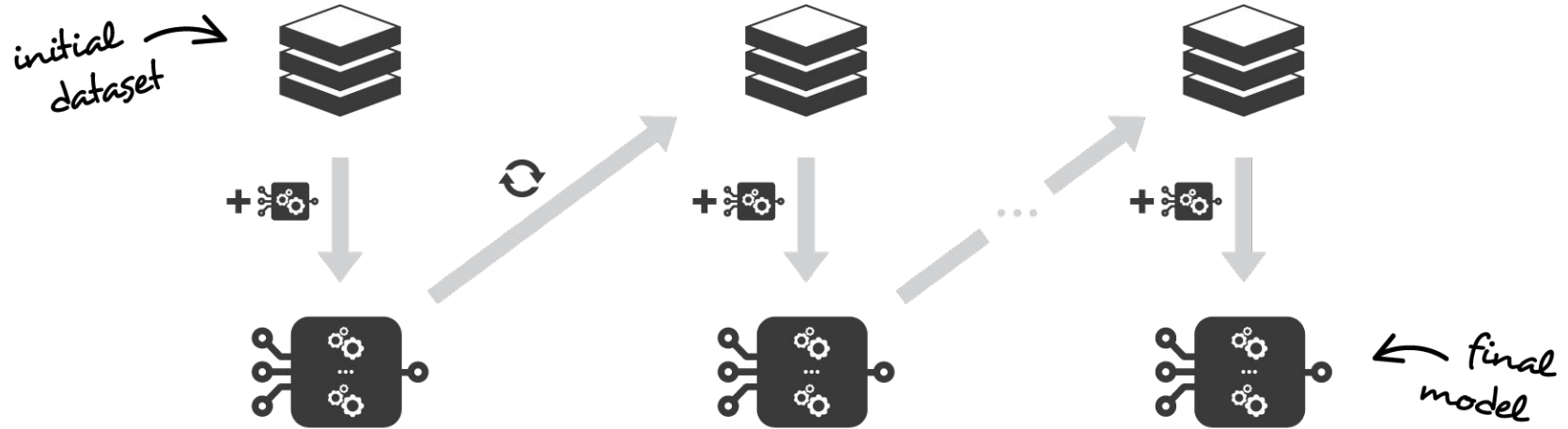
Boosting



train a weak model
and aggregate it to
the ensemble model



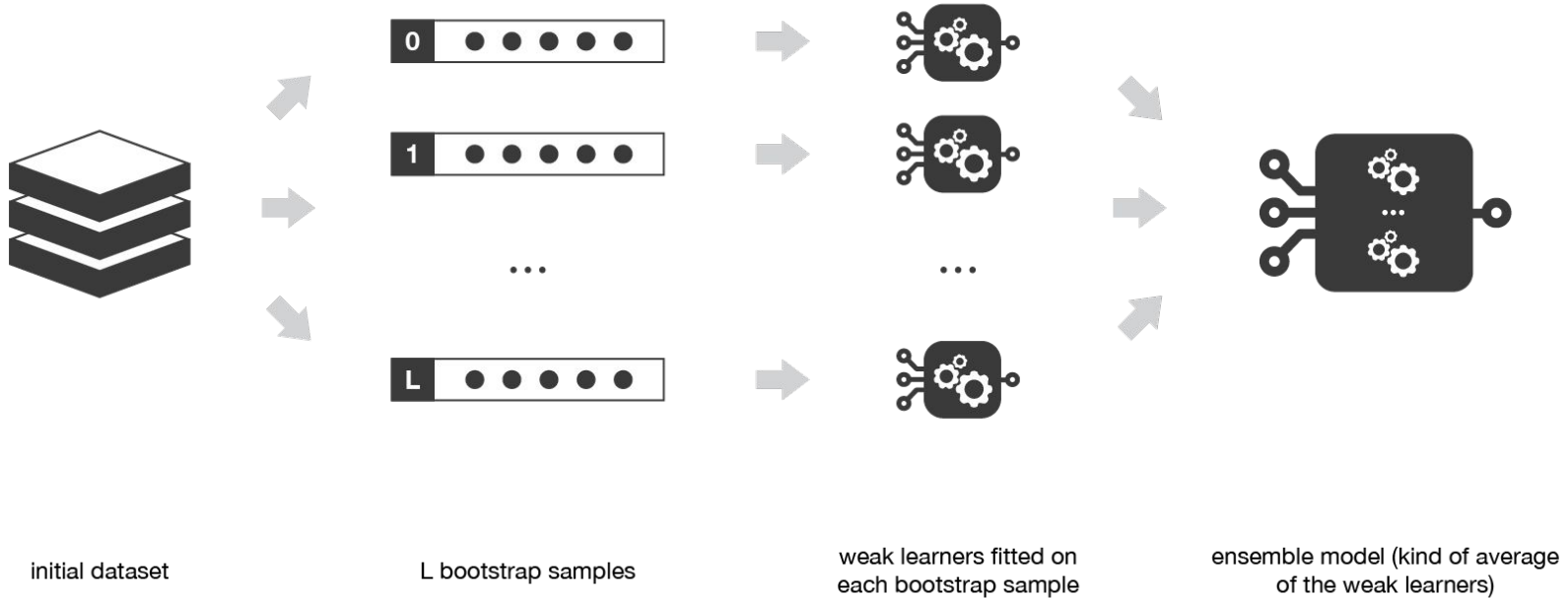
update the training dataset
(values or weights) based on the
current ensemble model results



Bagging

- *Bootstrap aggregating (bagging)* implica que cada **modelo débil** del conjunto vote con el mismo peso.
- Para promover la varianza de los modelos, se entrena a cada modelo del conjunto utilizando un **subconjunto** del conjunto **de entrenamiento** extraído aleatoriamente.
- Las muestras se generan de forma que sean diferentes entre sí, aunque se permite el reemplazo.
- Estas muestras se entregan a múltiples modelos y los resultados de cada uno se combinan en forma de votación.

Bagging



Stacking

- *Stacking* difiere del *bagging* y del *boosting* principalmente en dos puntos.
 - En primer lugar, el apilamiento suele tener en cuenta aprendices débiles **heterogéneos** (se combinan distintos algoritmos de aprendizaje), mientras que *bagging* y *boosting* tienen en cuenta aprendices débiles **homogéneos**.
 - En segundo lugar, el stacking aprende a combinar los modelos base utilizando un **metamodelo**, mientras que el bagging y el boosting combinan los aprendices débiles siguiendo algoritmos deterministas.

Stacking

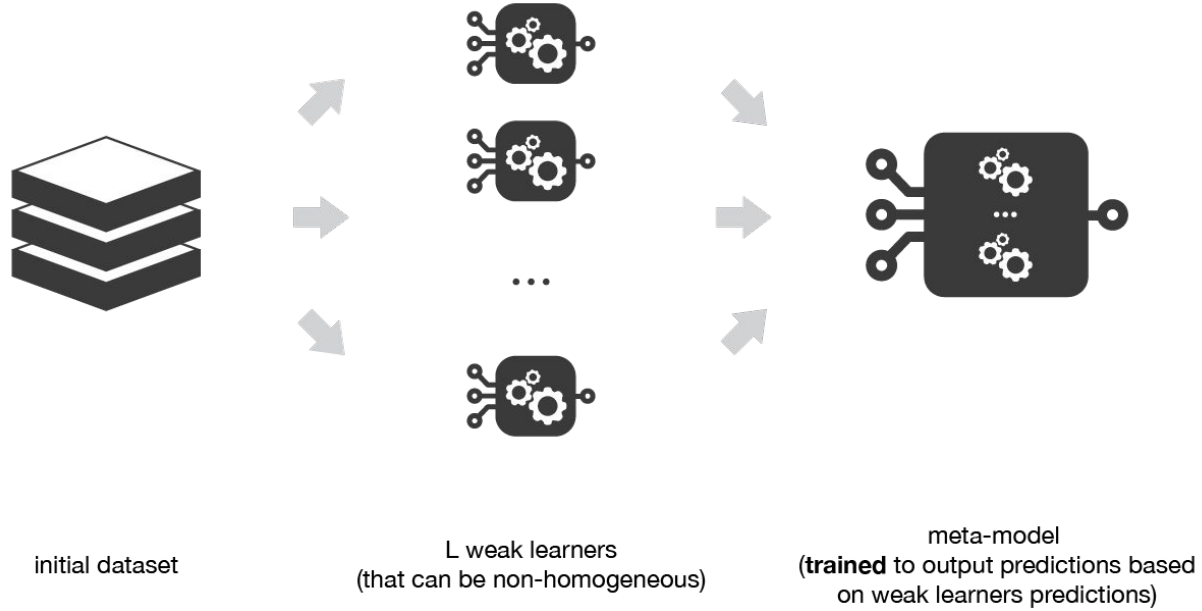
Algorithm 1 - Stacking

Input : $D = \{(x_i, y_i) | x_i \in \mathcal{X}, y_i \in Y\}$

Output : An ensemble classifier H

1. **Step 1 :** Learn first-level classifiers
2. For $t \leftarrow 1$ to T do
3. Learn a base classifier h_t based on D
4. **Step 2 :** Construct new data set from D
5. For $i \leftarrow 1$ to m do
6. Construct a new data set that contains $\{x_i^{new}, y_i\}$, where
 $x_i^{new} = \{h_j(x_i) \text{ for } j = 1 \text{ to } T\}$
7. **Step 3 :** Learn a second-level classifier
8. Learn a new classifier h^{new} based on the newly constructed data set
9. **Return** $H(x) = h^{new}(h_1(x), h_2(x), \dots, h_T(x))$

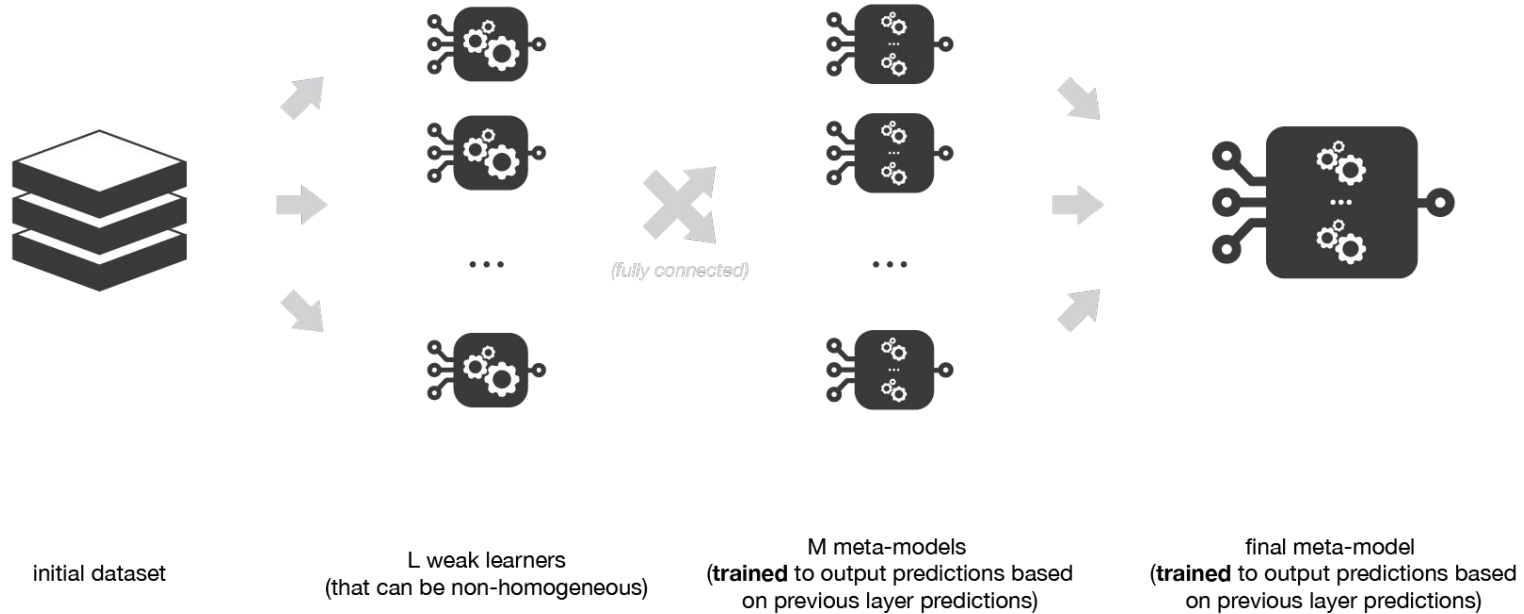
Stacking



Multi-levels Stacking

- Consiste en hacer apilamientos con múltiples capas.
- Considere un *stacking* de 3 niveles:
 - En el primer nivel (capa), ajustar los L aprendices débiles.
 - A continuación, en lugar de ajustar un único metamodelo sobre las predicciones de los modelos débiles, ajustar M metamodelos de este tipo.
 - Por último, ajustar un último metamodelo que toma como entradas las predicciones devueltas por los M metamodelos.
- La adición de niveles puede ser costosa en datos (si no se utiliza la técnica k -folds y, por tanto, se necesitan más datos) o en tiempo (si se utiliza la técnica k -folds y, por tanto, es necesario ajustar muchos modelos).

Multi-levels Stacking



Tamaño de los ensambles

- Aunque el **número** de clasificadores componentes de un **ensamble** tiene un gran impacto en el rendimiento, hay pocos estudios que aborden este problema.
 - La determinación a priori del tamaño del conjunto y la velocidad para grandes flujos de datos hacen que este aspecto sea aún más crucial.
- Para determinar el número adecuado de componentes se han utilizado principalmente pruebas estadísticas.
- La "ley de los rendimientos decrecientes en la construcción de ensambles" sugiere utilizar el mismo **número de clasificadores** independientes que de **etiquetas de clase**.

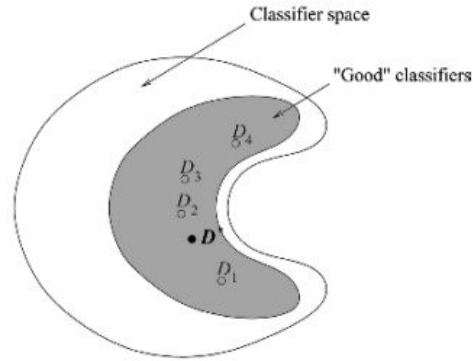
Aplicaciones

- Land cover mapping
- Computer security: malware and intrusion detection
- Face recognition
- Emotion recognition
- Fraud detection
- Financial decision-making
- Medicine, in neuroscience, proteomics and medical diagnosis

Resumen

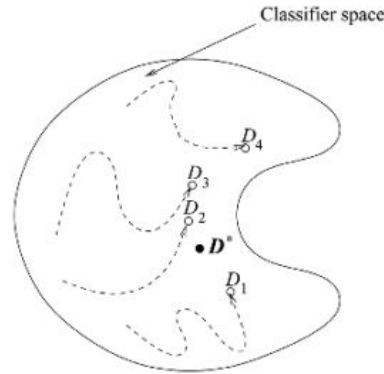
- Combinar clasificadores pretende una decisión de clasificación más precisa a costa de una mayor complejidad.
 - La cuestión es si la combinación de clasificadores está justificada.
- La combinación de clasificadores parece un paso adelante natural cuando se ha acumulado una masa crítica de conocimientos de modelos de clasificación individuales.
 - La advertencia aquí es que deberíamos ser capaces de aprovechar al máximo las herramientas y metodologías, antes de lanzarnos a nuevos diseños complicados.

The statistical reason



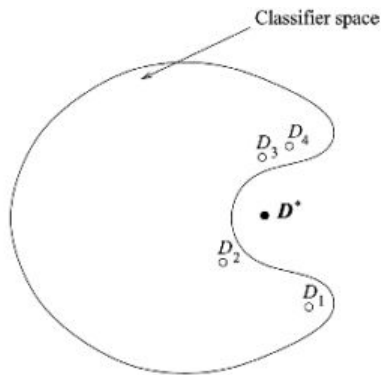
- D^* es el mejor clasificador para el problema, la curva exterior muestra el espacio de todos los clasificadores.
- El área sombreada es el espacio de los clasificadores con buenos resultados en los datos de entrenamiento.

The computational reason



- D^* es el mejor clasificador para el problema, la curva exterior muestra el espacio de todos los clasificadores.
- Las líneas sombreadas son las trayectorias hipotéticas de los clasificadores durante el entrenamiento.

The representational reason



- D^* es el mejor clasificador para el problema, la curva exterior muestra el espacio de todos los clasificadores.
- La forma cerrada muestra el espacio elegido por los clasificadores.

Referencias

- Russell, S., & Norvig, P. (2002). Artificial intelligence: a modern approach
- Kuncheva, L. I. (2014). Combining pattern classifiers: methods and algorithms. John Wiley & Sons.
- Kittler, J., Hatef, M., Duin, R. P., & Matas, J. (1998). On combining classifiers. IEEE transactions on pattern analysis and machine intelligence, 20(3), 226–239.
- Joseph Rocca (2019). Ensemble methods: bagging, boosting and stacking, Understanding the key concepts of ensemble learning.



Lectura 4: Ensamblas

Programado para el Mañana,...

Fecha de entrega: 27 oct

Preparar un reporte de máximo dos cuartillas con un análisis del artículo adjunto. El reporte debe finalizar con una crítica al artículo, el cual será abordado en las siguientes sesiones.

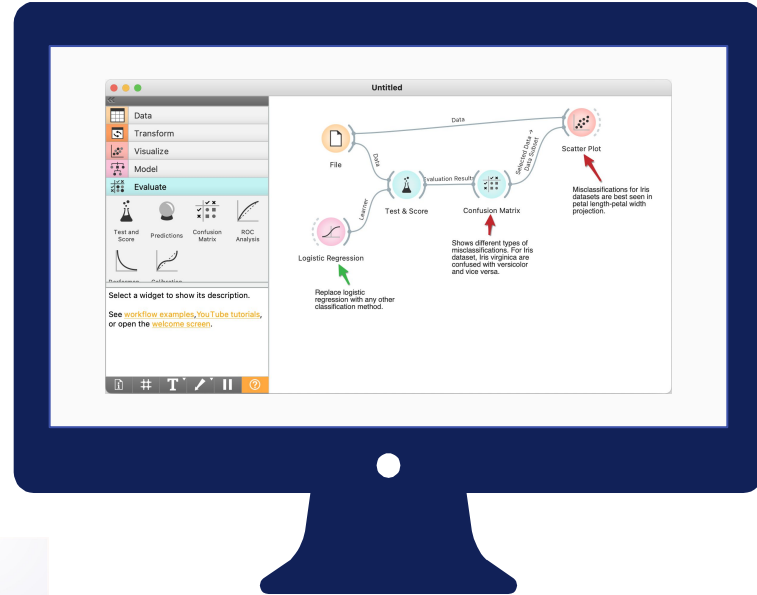
Ma, Y., Zhao, S., Wang, W., Li, Y., & King, I. (2022). Multimodality in meta-learning: A comprehensive survey. Knowledge-Based Systems, 250, 108976.



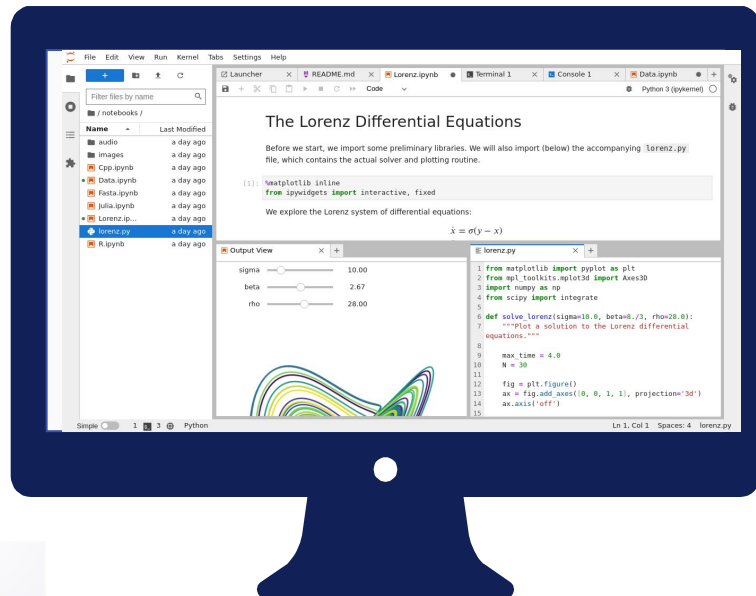
1-s2.0-S0950705122004...
PDF

[Editar tarea](#)

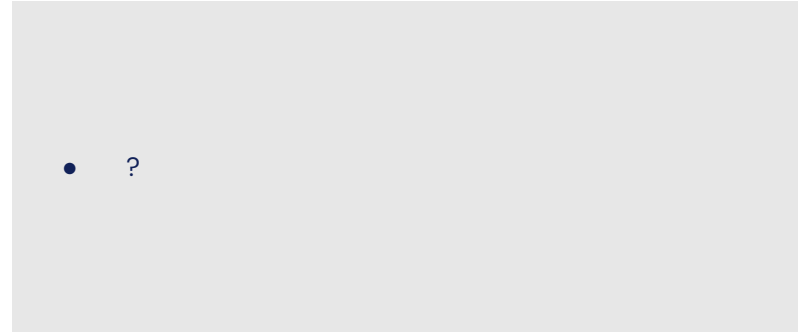
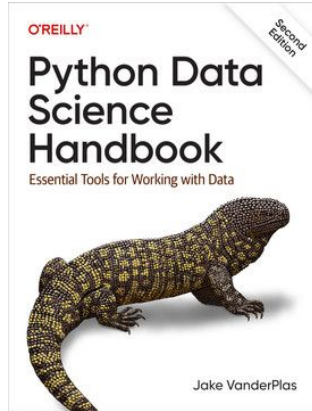
(Go to live)



(Go to live notebook)



Extra Libro



Gracias!

¿Alguna pregunta?

hussein@cicese.mx

<https://sites.google.com/view/husseinlopeznava>



CREDITS: This presentation was based on a template by [Slidesgo](#), and includes icons by [Flaticon](#).