

Ensemble Learning

Introducción + Voting Classifiers

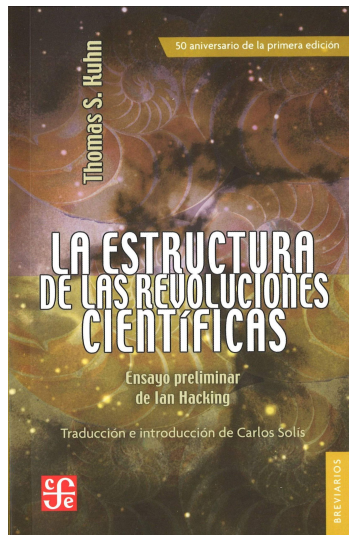
Luis Norberto Zúñiga Morales

8 de febrero de 2022

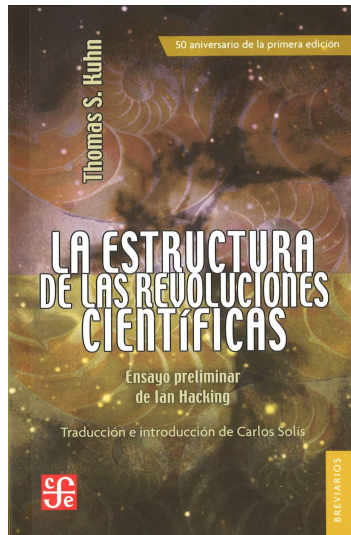
Pregunta

¿Alguien sabe qué es un *paradigma* en la ciencia?

- *Paradigma* tiende a ser sinónimo de *ejemplo* o, en algunas instancias, *modelo*.

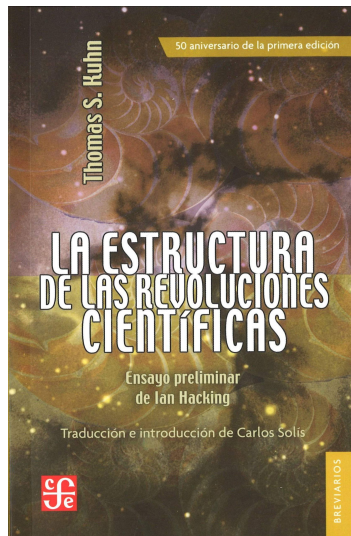


- *Paradigma* tiende a ser sinónimo de *ejemplo* o, en algunas instancias, *modelo*.
- En 1960, Thomas S. Kuhn redefine el concepto en el libro "La Estructura de las Revoluciones Científicas".

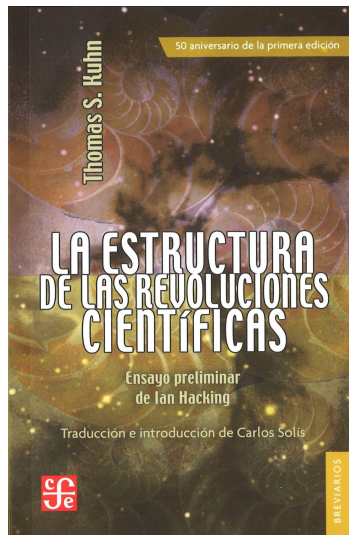


Motivación

- «... 'ciencia normal' significa investigación basada firmemente en una o más realizaciones científicas pasadas, realizaciones que alguna comunidad científica particular reconoce, durante cierto tiempo, como fundamento para su práctica posterior. [...] Voy a llamar, de ahora en adelante, a las realizaciones que comparten esas dos características, 'paradigmas', término que se relaciona estrechamente con 'ciencia normal'.»

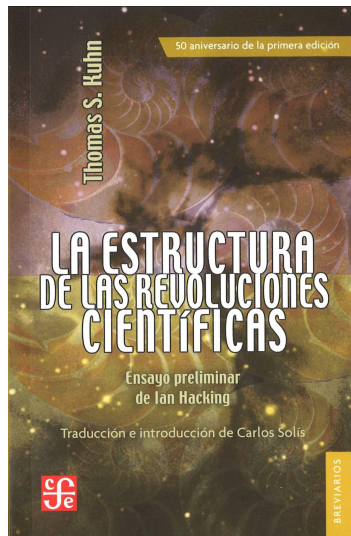


- *Paradigma* se adoptó para referirse al conjunto de prácticas y saberes que definen una disciplina científica durante un período específico.



Motivación

- *Paradigma* se adoptó para referirse al conjunto de prácticas y saberes que definen una disciplina científica durante un período específico.
- Se manifiesta a través de los libros de texto propios de una ciencia o disciplina y las teorías aceptadas por las comunidades científicas.



Ejercicio

Mencionen brevemente un paradigma en su campo de estudio (probabilidad, ciencia de datos, actuaría, matemáticas, etc.).

Motivación

- ¿Alguna vez han consultado opiniones o comentarios sobre un producto antes de adquirirlo?



Motivación

- ¿Alguna vez han consultado opiniones o comentarios sobre un producto antes de adquirirlo?
- ¿Qué clase de productos tienden a ser analizados cuidadosamente antes de su compra?



Motivación

- ¿Alguna vez han consultado opiniones o comentarios sobre un producto antes de adquirirlo?
- ¿Qué clase de productos tienden a ser analizados cuidadosamente antes de su compra?
- ¿Por qué creen que de esta manera llegan a una decisión «correcta»?



Ensemble Learning

- El *Ensemble Learning* (EL) es un paradigma de aprendizaje automático (*Machine Learning*) donde **múltiples modelos de aprendizaje** se entrenan para resolver un problema.

Ensemble Learning

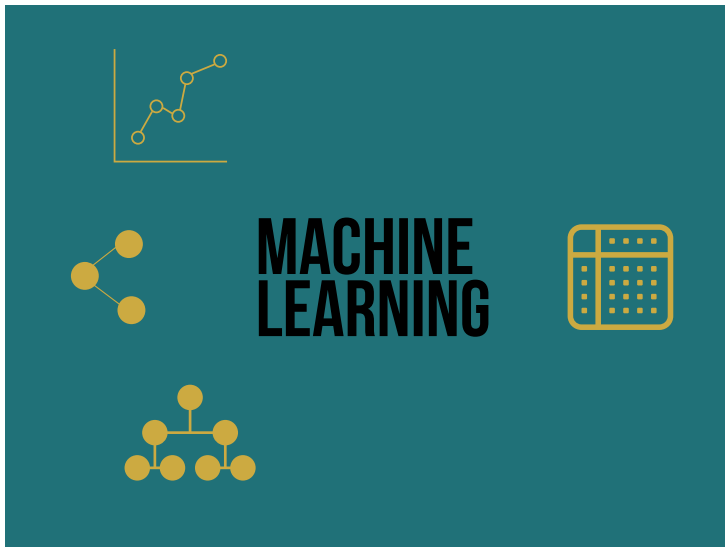
- El *Ensemble Learning* (EL) es un paradigma de aprendizaje automático (*Machine Learning*) donde **múltiples modelos de aprendizaje** se entrenan para resolver un problema.
- En conjunto, los modelos intentan construir **múltiples hipótesis** para resolver el problema de aprendizaje, en lugar de una única hipótesis si se trabajara con un modelo en específico [3].

- La idea básica del EL surge al **imitar el comportamiento de aprendizaje social humano** donde se busca el consejo u opinión de otros miembros de nuestro entorno social para realizar una decisión crucial o importante.

Ensemble Learning

- La idea básica del EL surge al **imitar el comportamiento de aprendizaje social humano** donde se busca el consejo u opinión de otros miembros de nuestro entorno social para realizar una decisión crucial o importante.
- De forma similar, el EL **combina distintos modelos de aprendizaje** (modelos base) que combinan sus formas de aprendizaje, la forma en la que manejan los datos u otras características particulares para obtener mejores resultados [1].

Ensemble Learning



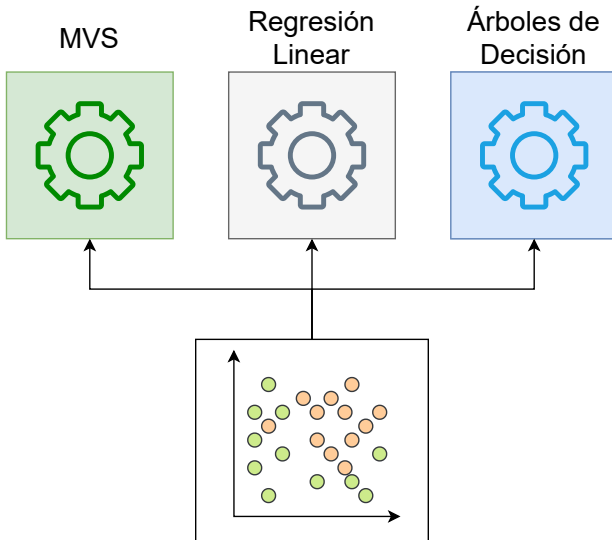
- **Métodos que promedian** (*Averaging Methods*), los cuales construyen distintos modelos de clasificación independientes entre ellos para promediar las predicciones que realizan.
- En esta clase entran los métodos de Bagging, Árboles y Bosques Aleatorios, entre otros.

- **Métodos que aumentan** (*Boosting Methods*), los cuales combinan diferentes modelos débiles para que, en conjunto, puedan formar un modelo robusto reduciendo el error de entrenamiento.
- En esta clase entran modelos como AdaBoost, Gradient Boosting, XGBoosting, entre otros.

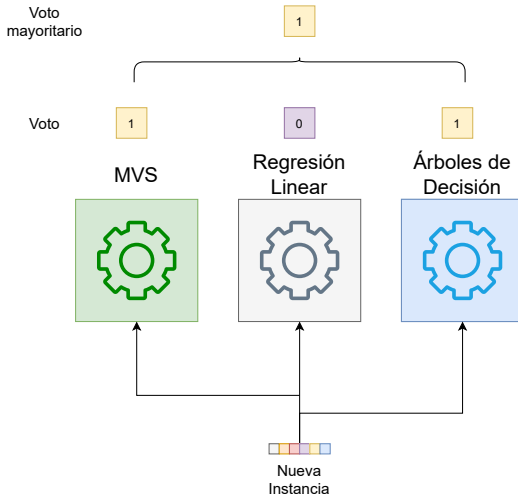
Existen otras formas de clasificar:

- Una de las más famosas es Bagging vs Boosting.
- Re y Valentini [2] consideran dos grandes clases: métodos generativos y no generativos
 - La principal diferencia consiste en si generan o no nuevos modelos base como parte del proceso del ensamble.

Voting Classifiers



Voting Classifiers



Voting Classifiers

- Supongamos que se tienen M distintos modelos de aprendizaje $\{\phi_m \mid m = 1, \dots, M\}$ todos utilizando el conjunto de datos L .
- El ensamble, representado como $\psi_{\phi_1, \dots, \phi_m}$, busca reducir el error de generalización al considerar el error esperado de cada modelo individual.

Voting Classifiers

Voto Duro

La votación dura se expresa de la siguiente manera:

$$\psi_{\phi_1, \dots, \phi_m}(\mathbf{x}) = \operatorname{argmax}_{c \in y} \sum_{m=1}^M 1(\phi_m(\mathbf{x}) = c) \quad (1)$$

Voto Suave

Cuando algún modelo de aprendizaje individualmente estima una probabilidad $\hat{p}_L(Y = c|X = \mathbf{x})$ para cada clase, es posible promediar la probabilidad estimada para cada una de ellas y después asignar aquella que sea la más probable:

$$\psi_{\phi_1, \dots, \phi_m}(\mathbf{x}) = \operatorname{argmax}_{c \in y} \frac{1}{M} \sum_{m=1}^M \hat{p}_L(Y = c|X = \mathbf{x}) \quad (2)$$

Principio del Jurado de Condorcet

Suponga que se tiene un jurado compuesto por M personas las cuales desean llegar a un veredicto por medio de una mayoría en una votación. Si cada jurado tiene una probabilidad independiente $p > \frac{1}{2}$ de elegir la decisión correcta, añadir más votantes incrementa la probabilidad de que la decisión colectiva sea la correcta. Cuando $M \rightarrow \infty$, la probabilidad de que se tome la decisión correcta tiende a 1. Por otro lado, si $p < \frac{1}{2}$, cada votante es más propenso a votar de forma incorrecta e incrementar M empeora las cosas.

Voting Classifiers

- El teorema anterior recalca un punto importante.
- Los modelos de aprendizaje deben ser débiles (*weak learners*), i.e., su desempeño debe ser ligeramente mejor que el baseline aleatorio.
- En conjunto, se transforman en un modelo de aprendizaje fuerte (*strong learner*).
- Eso lleva al resultado que un modelo débil es equivalente a uno fuerte.

- Demostrar el Principio del Jurado de Condorcet.
 - Pista: usen la computadora.

- [1] Thomas G. Dietterich. Ensemble methods in machine learning. In *Proceedings of the First International Workshop on Multiple Classifier Systems*, MCS '00, page 1–15, Berlin, Heidelberg, 2000. Springer-Verlag.
- [2] Matteo Re and Giorgio Valentini. *Ensemble methods: A review*, pages 563–594. 01 2012.
- [3] Zhi-Hua Zhou. *Ensemble Learning*, pages 270–273. Springer US, Boston, MA, 2009.