

Aprendizaje Profundo

Maestría en Modelación y Optimización de Procesos

CIMAT-Aguascalientes

Dra. Lilí Guadarrama Bustos¹
Dr. Isidro Gómez-Vargas²

¹Investigadora tiempo completo en CIMAT-Aguascalientes

²Investigador posdoctoral en ICF-UNAM

Clase del semestre enero-julio 2022

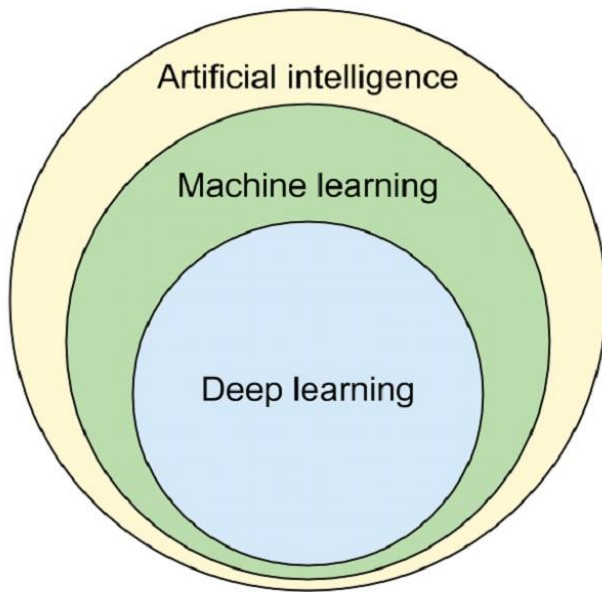
Contenido

1 Clase 1: Bases de Aprendizaje Automático

- Algoritmos de aprendizaje.
- Capacidad, subajuste y sobreajuste
- Hiperparámetros y conjuntos de validación.
- Estimadores, sesgo y varianza.
- Estimación de Máxima Verosimilitud(MLE).
- Estadística Bayesiana.
- Algoritmos de aprendizaje supervisado.
- Descenso del gradiente estocástico.
- Desafíos actuales Aprendizaje Profundo.

2 Clase 2: Deep Feedforward Networks

Aprendizaje profundo



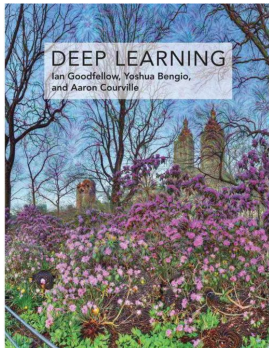
Bibliografía

Bibliografía principal

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Diapositivas:

- https://www.deeplearningbook.org/lecture_slides.html
- <https://github.com/InfolabAI/DeepLearning>



Yoshua Bengio in 2019

Born March 5, 1964 (age 57)
Paris, France

Citizenship Canada

Alma mater McGill University

Known for Deep learning, neural machine translation, generative adversarial networks, "attention model", word embeddings, denoising auto-encoders, neural language models, learning to learn

Awards Marie-Victorin Prize (2017)
Turing Award (2018)
AAAI Fellow (2019)

Scientific career

Born 1985/1986 (age 35–36)

Nationality American

Alma mater Stanford University
Université de Montréal

Known for Generative adversarial networks, **Adversarial Examples**

Scientific career

Fields Computer science

Institutions Apple Inc.
Google Brain
OpenAI

Thesis *Deep Learning of Representations and its Application to Computer Vision* (2014)

Doctoral advisor Yoshua Bengio
Aaron Courville

Website www.lsgoodfellow.com



Aaron Courville

Université de Montréal

Dirección de correo verificada de umontreal.ca - [Página personal](#)

Machine learning Artificial Intelligence

TÍTULO

CITADO POR

Generative adversarial nets

40148

I Goodfellow, J Pouget-Abadie, M Mirza, B Xu, D Warde-Farley, S Ozair, ...
Advances in neural information processing systems 27

Deep learning

36121

IJ Goodfellow, Y Bengio, A Courville
Nature

Representation learning: A review and new perspectives

10333

Y Bengio, A Courville, P Vincent
IEEE transactions on pattern analysis and machine intelligence 35 (8): 1708-1725

Temario

Se abordarán partes de los siguientes capítulos:

- Capítulo 5: Bases del Aprendizaje Automático.
- Capítulo 6: Redes neuronales profundas de propagación hacia adelante.
- Capítulo 7: Regularización.
- Capítulo 8: Optimización para el entrenamiento de modelos profundos.

Temario

- Capítulo 9: Redes neuronales convolucionales.
- Capítulo 10: Modelación secuencial (redes recurrentes y recursivas).
- Capítulo 11: Metodología práctica.
- Capítulo 12: Aplicaciones.
- Temas elegidos por el grupo de la parte III del libro (capítulos 13-20).

Algoritmos de aprendizaje.

“

Un programa de computadora se dice que aprende de la experiencia E con respecto a cierta clase de tareas T y con medida de rendimiento P , si su rendimiento en las tareas en T , medido por P , mejora con la experiencia E .

”

Tom Mitchell, 1997

Glosario

- **Ejemplo.** $x \in \mathbf{R}^n$, colección de n características.
- **Conjunto de datos.** Colección de ejemplos.
- **Características.** Atributos.
- **Matriz de datos.**

Tareas T

- Clasificación
- Regresión
- Transcripción
- Traducción
- Salidas estructuradas
- Detección de anomalías
- Síntesis y muestreo
- Imputing valores eprdidos
- Quitar ruido
- Estimación de densidad o probabilidad

La medida de rendimiento P

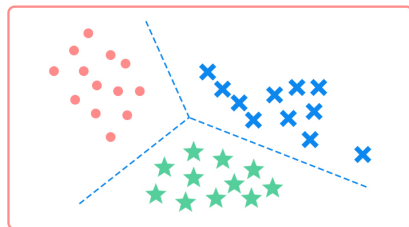
- Exactitud $\frac{TP+TN}{TP+TN+FP+FN}$
- Error cuadrático medio $MSE = \frac{1}{n} \sum_i^n ||\hat{y}_i - y_i||$
- Densidad de probabilidad
- etc

La experiencia, E



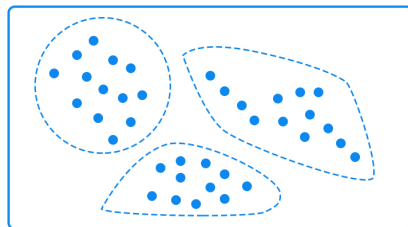
Supervised vs. Unsupervised Learning

Classification



Supervised learning

Clustering

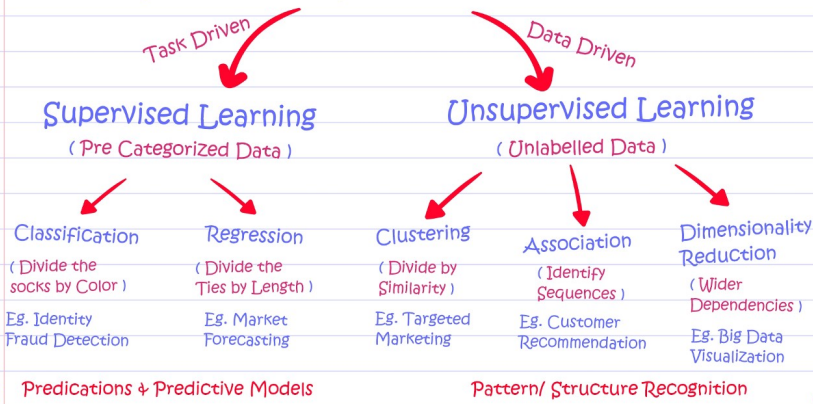


Unsupervised learning

Fuente:

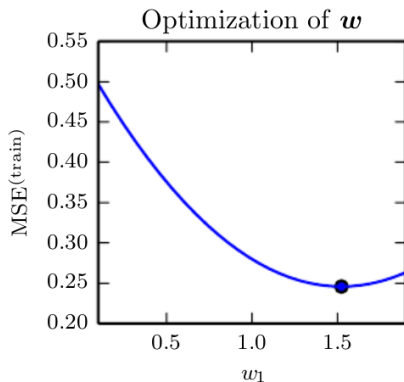
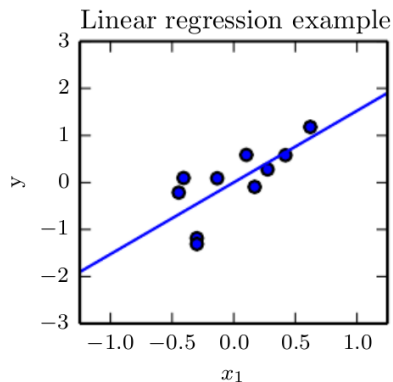
https://analystprep.com/study-notes/cfa-level-2/quantitative-method/supervised-machine-learning-unsupervised-machine-learning-deep-learning/attachment/img_12-4/

Classical Machine Learning



Fuente: <https://medium.com/@recrosoft.io/supervised-vs-unsupervised-learning-key-differences-cdd46206cdcb>

Ejemplo: regresión lineal

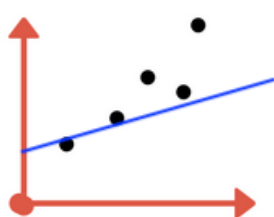


Ver páginas 107-108 del libro.

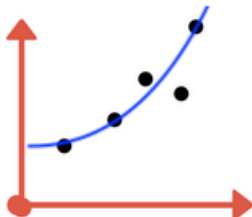
Capacidad, subajuste y sobreajuste

- Algo que separa al aprendizaje automático de la optimización es el uso de un error de generalización o error de prueba (*test error*).
- Hay que minimizar tanto el error de entrenamiento como el error de generalización.
- La brecha entre ambos errores debe ser pequeña.

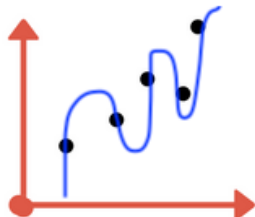
Capacidad, subajuste y sobreajuste



underfitting

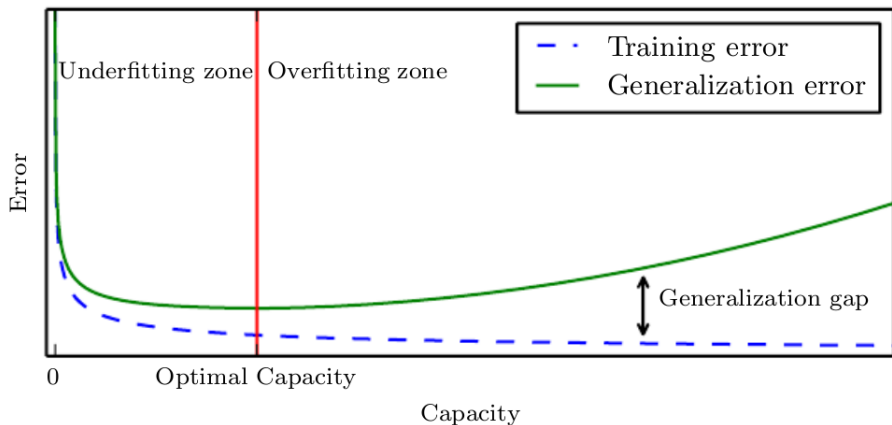


correcto



overfitting

Capacidad, subajuste y sobreajuste



Capacidad, subajuste y sobreajuste: No free lunch theorem



"No Free Lunch" :(

D. H. Wolpert. The supervised learning no-free-lunch theorems. In Soft Computing and Industry, pages 25–42. Springer, 2002.

Our model is a simplification of reality



Simplification is based on assumptions (model bias)



Assumptions fail in certain situations

Roughly speaking:

"No one model works best for all possible situations."

Fuente: <https://analyticsindiamag.com/what-are-the-no-free-lunch-theorems-in-data-science/>

Capacidad, subajuste y sobreajuste: Regularización

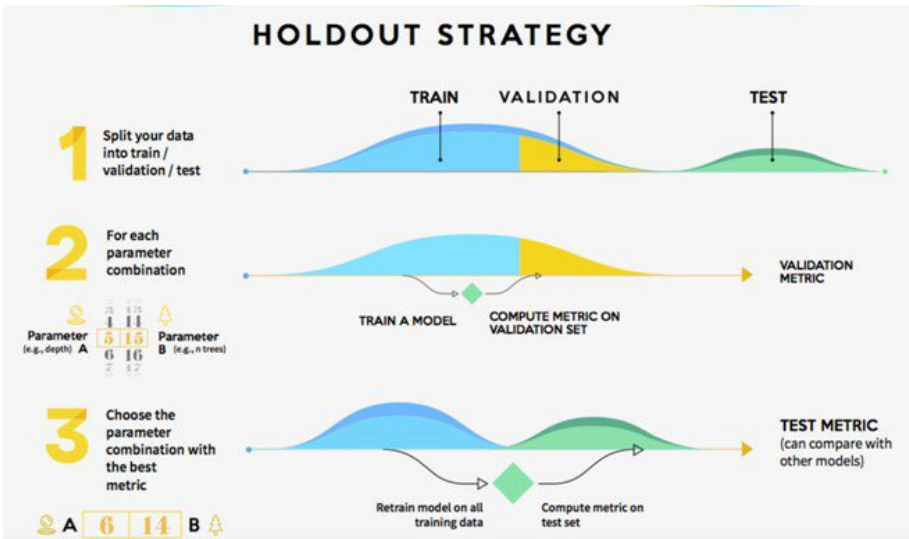
Por ejemplo, decaimiento del peso:

$$J(w) = MSE + \lambda \omega^T \omega$$

Def.

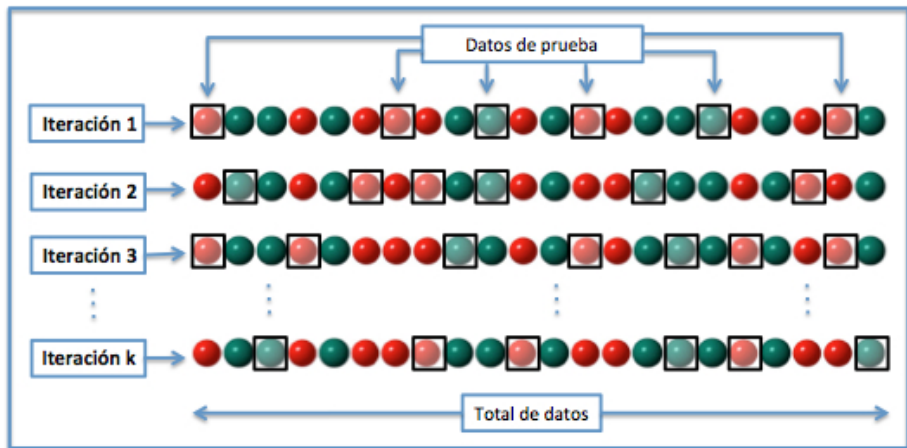
Regularización es cualquier mecanismo que ayuda al algoritmo de aprendizaje a reducir su error de generalización.

Hiperparámetros y conjuntos de validación.



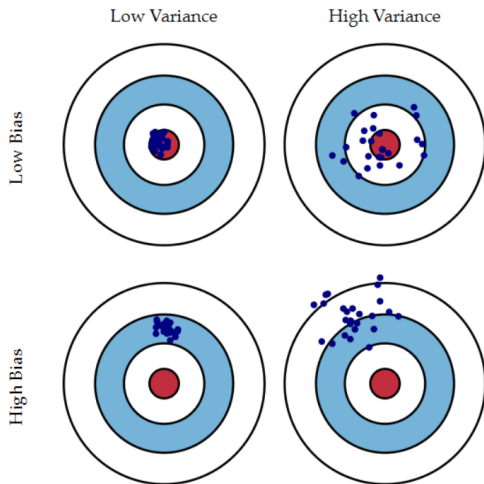
Fuente: <https://www.kdnuggets.com/2017/08/dataiku-predictive-model-holdout-cross-validation.html>

Hiperparámetros y conjuntos de validación: validación cruzada

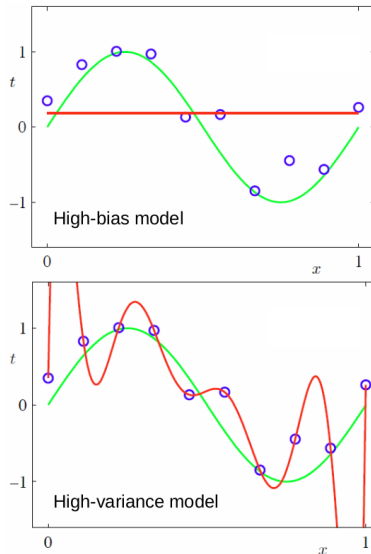


Fuente: https://es.wikipedia.org/wiki/Validaci%C3%B3n_cruzada

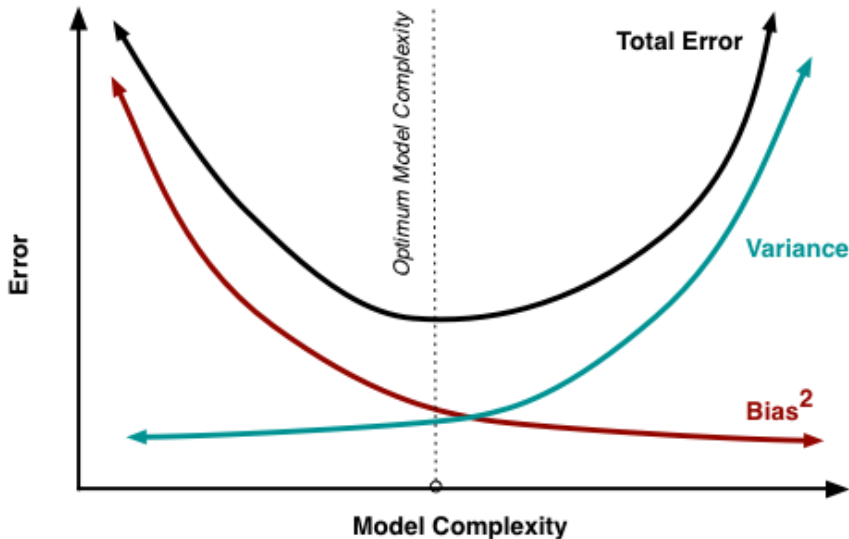
Estimadores, sesgo y varianza.



Scott Fortmann-Roe, Understanding the Bias-Variance Tradeoff, 2012



Estimadores, sesgo y varianza.



Estimadores, sesgo y varianza.

Visitar los siguientes enlaces:

- <http://scott.fortmann-roe.com/docs/BiasVariance.html>
- <https://ml.berkeley.edu/blog/posts/crash-course/part-4/>
- [https://machinelearningmastery.com/
learning-curves-for-diagnosing-machine-learning-model-per](https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-per)

Estimación de Máxima Verosimilitud(MLE).

- Estimación del Máximo Likelihood (MLE):

$$\ln \mathcal{L}(D, \theta) = \sum_{i=1}^n \ln f(x_i; \theta),$$

$$\theta_{MLE} = \arg \max(\mathcal{L}(\theta, D))$$

Teorema de Bayes

Considerando funciones de densidad de probabilidad:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}, \quad (1)$$

donde:

$$P(D) = \int_{\mathbb{R}^N} P(D|\theta)P(\theta)d\theta, \quad (2)$$

Estadística Bayesiana.

- Estimación del A Posteriori (MAP) ó estimación de parámetros ó inferencia Bayesiana:

$$\theta_{MAP} = \arg \max(\mathcal{L}(\theta, D)P(\theta))$$

Estadística Bayesiana.

- Estimación del A Posteriori (MAP) ó estimación de parámetros ó inferencia Bayesiana:

$$\theta_{MAP} = \arg \max(\mathcal{L}(\theta, D)P(\theta))$$

- Comparación de modelos (puede ser parte de la inferencia Bayesiana).

Estadística frecuentista vs Bayesiana.

You are no good when sample is small



You give a different answer for different priors

Fuente: <https://laptrinhx.com/maximum-likelihood-estimation-vs-maximum-a-posteriori-2539680111/>

Estadística frecuentista vs Bayesiana.

Suppose we have data $\mathcal{D} = \{x^{(i)}\}_{i=1}^N$

$$\theta^{\text{MLE}} = \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^N p(\mathbf{x}^{(i)} | \theta)$$

Maximum Likelihood Estimate (MLE)

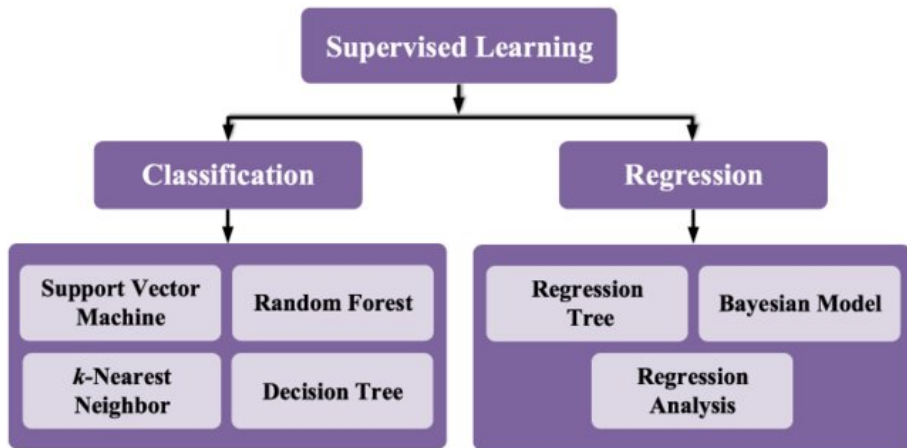
$$\theta^{\text{MAP}} = \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^N p(\mathbf{x}^{(i)} | \theta) \underbrace{p(\theta)}_{\text{Prior}}$$

Maximum a posteriori (MAP) estimate

Fuente: <https://medium.com/@tzjy/>

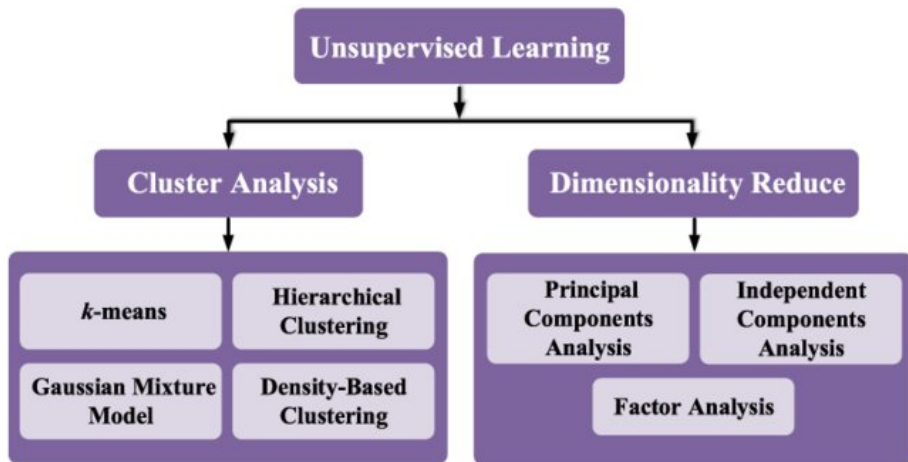
whats-the-difference-between-maximum-likelihood-estimation-mle-and-maximum

Algoritmos de aprendizaje supervisado.



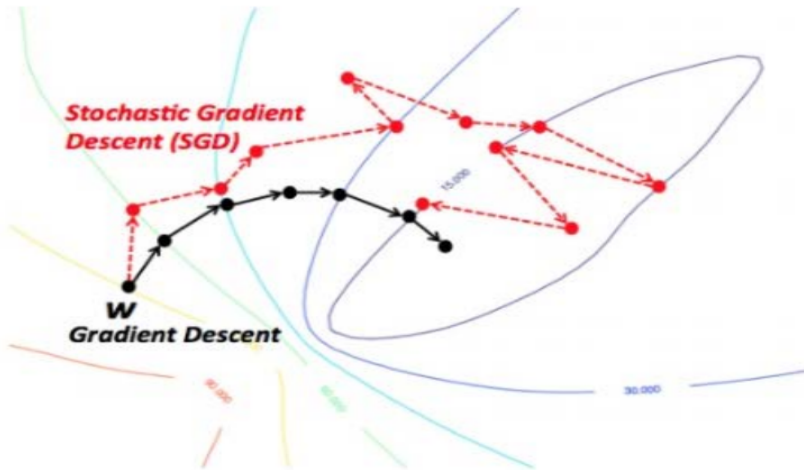
Fuente: arXiv:2003.10146

Algoritmos de aprendizaje no supervisado.



Fuente: arXiv:2003.10146

Descenso del gradiente estocástico.



Fuente: <https://www.slideshare.net/microlife/from-neural-networks-to-deep-learning>

Descenso del gradiente estocástico.

Tarea propuesta

Programar un descenso del gradiente estocástico y minimizar una función con él.

Retos que motivan al Aprendizaje Profundo.

- Curso de la dimensionalidad.
- Local Constancy and Smoothness Regularization.
- Manifold Learning.

Clase 2: Introducción al Aprendizaje Profundo

- Inicia la parte II del libro de referencia.
- Nos centraremos en el capítulo 6.

Ver notebooks: <https://github.com/igomezv/DLCIMATAGS/tree/main/notebooks/clase%202>