Aprendizaje NO Supervisado: Componentes Principales y Redes Neuronales

11 de mayo de 2022

Componentes Principales - Aprendizaje no supervisado

- Las componentes principales son utilizadas para extraer características destacadas o importantes de un conjunto de datos bajando la dimensión del mismo.
- Por ejemplo la componente Y_1 nos permite agrupar los registros que tengan valores de Y_1 parecidos.
- O también ordenar registros de acuerdo a esta característica.

Componentes Principales - Redes Neuronales

- Algunos modelos de redes neuronales permiten calcular las componentes principales en forma iterativa, ofreciendo ventajas computacionales.

Perceptron Lineal Simple

Dados m datos de entrada, $x^1, \ldots, x^m$, $x^i \in \Re^N$, $\forall i$, una red neuronal calcula la salida

$$O^n = \sum_{j=1}^{N} w_j^n * x_j^i$$

donde $w^n \in \Re^N$ $\forall n$.

Actualizar los pesos, ζ la verdadera

$\Delta w_j = \eta(\zeta - y^n)x_j^i$, η es la tasa de aprendizaje.
$w_j^{n+1} = w_j^n + \eta(\zeta - y^n)x_j^i$

Regla de Aprendizaje Hebbiano para Aprendizaje No supervisado

Dados m datos de entrada, $x^1, \ldots, x^m$, $x^i \in \Re^N$, $\forall i$, una red neuronal calcula la salida

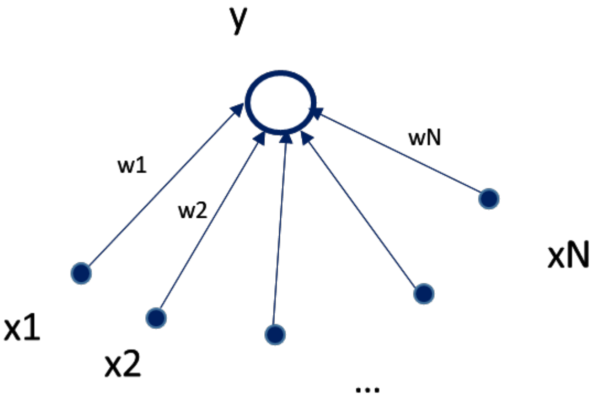
$$y^n = \sum_{j=1}^{N} w_j^n * x_j^i$$

donde $w^n \in \Re^N \; \forall n$.

Actualizar los pesos

$\Delta w_j = \eta y^n x_j^i$, η es la tasa de aprendizaje.
$w_j^{n+1} = w_j^n + \eta y^n x_j^i$

Arquitectura para Aprendizaje Hebbiano simple



La unidad de salida es lineal $y^n = \sum_{j=1}^{N} w_j^n * x_j^i$.

¿A qué converge?

Oja demostró que

Si la red anterior converge, el vector de pesos resultante w^{final} sería un punto sobre la dirección de máxima variación de los datos, o sea... la primer componente principal.

pero...

el problema es que no converge porque el $||w^n||$ va aumentando en cada paso y se hace tan grande que produce que el algoritmo sea inestable.

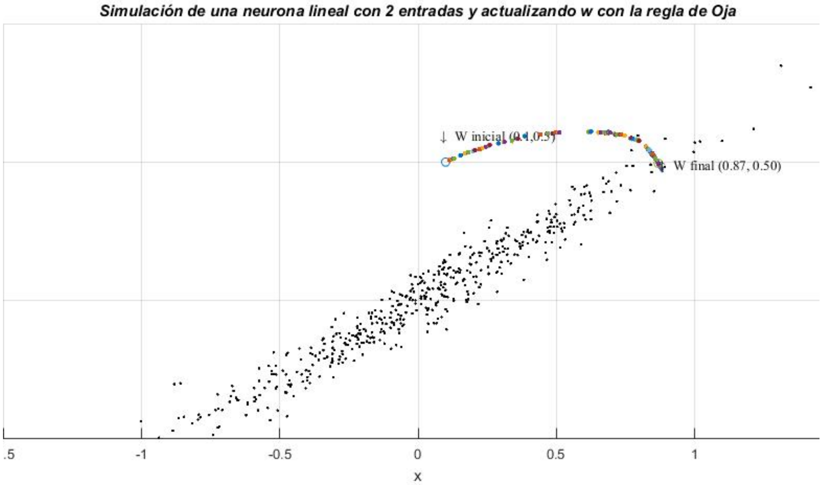
La propuesta de Oja (1982)

Dr. Erkki Oja, Helsinki University of Technology, Finland

$$w_j^{n+1} = \frac{w_j^n + \eta y^n x_j^n}{\left(\sum_{j=1}^{N} (w_j^n + \eta y^n x_j^n)^2\right)^{\frac{1}{2}}}$$

η es la tasa de aprendizaje.

Gráficamente



La propuesta de Oja

Utilizando el polinomio de Taylor en η

$$w_j^{n+1} = \frac{w_j^n}{||\mathbf{w}||} + \eta \left(\frac{y(\mathbf{x^n})\mathbf{x_j^n}}{||\mathbf{w}||} - \frac{w_j \sum_{j=1}^{N} y(\mathbf{x^n})\mathbf{x_j^n}\mathbf{w_j}}{||\mathbf{w}||^3} \right) =$$

$$\frac{w_j^n}{||\mathbf{w}||} + \eta \left(\frac{y(\mathbf{x^n})\mathbf{x_j}}{||\mathbf{w}||} - \frac{w_j^n y(\mathbf{x^n}) \sum_{\mathbf{j=1}}^{\mathbf{N}} \mathbf{x_j^n}\mathbf{w_j^n}}{||\mathbf{w}||^3} \right) =$$

La regla de aprendizaje de Oja en el paso n

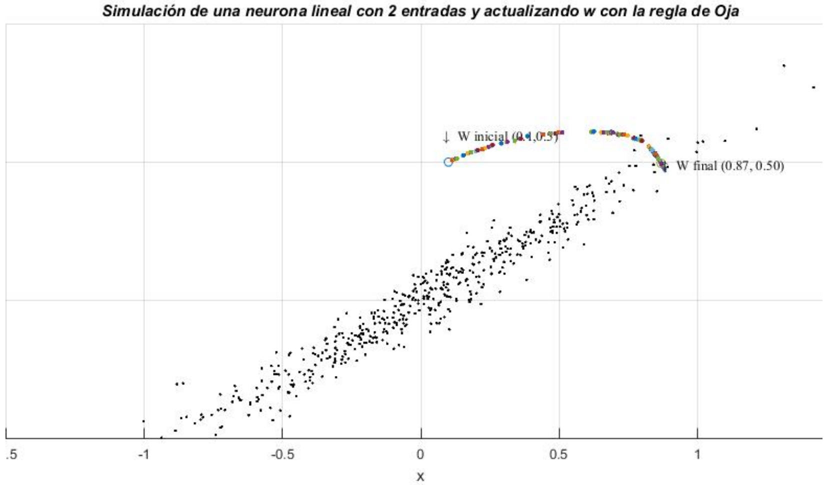
$$\Delta w_j = \eta(y * x_j^n - y^2 * w_j^n)$$

η es la tasa de aprendizaje.

Con esta regla...

- $|w|$ se mantiene acotado.
- Luego de varias iteraciones el método converge al autovector correspondiente al mayor autovalor de la matriz de correlaciones de los datos de entrada.
- Con este vector w_{final} se construye la primer componente principal.

Gráficamente



Algoritmo

input: N datos con media 0; η, w_{inic}

1. for epoch in 1:epoca
   1. for i in 1:N
      1. $s = \sum_j (datos[i,j] * w_j)$
      2. $w = w + \eta * s * (datos[i,] - s * w)$
2. return(w)

$datos[i,]$ significa toda la entrada i

Extensión de la regla de Oja

Regla de Sanger 1989 [5]

- Converge a la matriz de autovectores de la matriz de covarianzas de los datos.
- Permite encontrar todas las componentes principales.

Regla de Sanger

- $\mathbf{y} = W^t\mathbf{x}$ donde $\mathbf{y} = (y_1, \ldots, y_k)$, $\mathbf{x} = (\mathbf{x_1}, \ldots, \mathbf{x_m})$ y $W = (\mathbf{w_1}, \ldots, \mathbf{w_k})$
- Permite encontrar los primeros k autovectores y por lo tanto las primeras k componentes principales, utilizando la regla de Sanger.

Regla de Sanger

- $y_i^{n+1} = \sum_{j=1}^{N} w_{ij}^n x_j^n$
- $w_{il}^{n+1} = w_{il}^n + \eta y_i^n (x_l - \sum_{j=1}^{i} y_j^n w_{jl}^n)$

$1 \leq i \leq k,\ 1 \leq k \leq N,\ 1 \leq l \leq N,\ \eta \geq 0$

En cada paso le resta todas las componentes encontradas en los pasos anteriores.

Referencias

- Metodos Hopfield [1].
- Método de Kohonen [3, 4].
- Para todos los métodos [2].

Referencias

[1] McKay D.J.C. *Hopfield Networks*. Information Theory, Inference and Learning Algorithms, Cambridge, 2003.

[2] Anders Krogh John Hertz and Richard Palmer. *Introduction to the Theory of Neural Computation*. Addison-Wesley, 1991.

[3] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 1(43):59–69, 1982.

[4] T. Kohonen. The self-organizing map. *Neurocomputing*, pages 1–6, 1998.

[5] T. D. Sanger. Optimal unsupervised learning in a singlelayer linear feedforward neural network. *Neural Networks*, 2(6):459–473, 1989.