



TRATAMIENTO DIGITAL DE SEÑALES

Ingeniería de Telecomunicación (4º, 2º c)

Unidad 13^a: Redes Neuronales

Aníbal R. Figueiras Vidal
Jesús Cid Sueiro
Ángel Navia Vázquez

Área de Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70



alidades

Neuronal (NN) es una máquina:

arquitectura paralela

unidades no lineales (“neuronas”) sencillas

pero dichas unidades dispuestas normalmente en capas

la típica:
$$o = \sum_j w_j \phi_{w_j}(\underline{x}) + w_0$$

$$\phi_j = \text{th}\left(\sum_i w_{ji} x_i + w_{j0}\right) : \text{Perceptrones (iterables (MLP))}$$

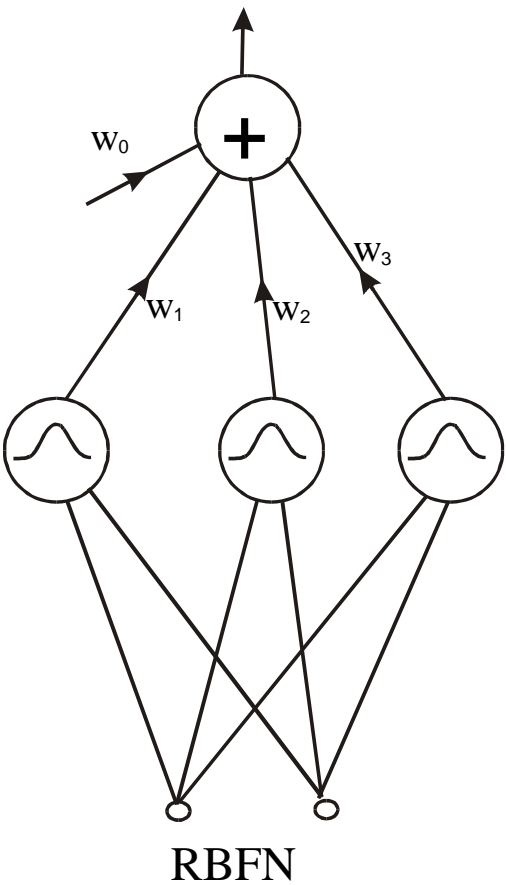
$$\phi_j = G(\mathbf{x} - \mathbf{m}_j, V_j)$$
 (o similares con centros \mathbf{m}_j): Red de **Funciones Base**

Radiales (RBFN) (puede prescindirse del factor de la exponencial)

ATSC-DTC/UCIIM

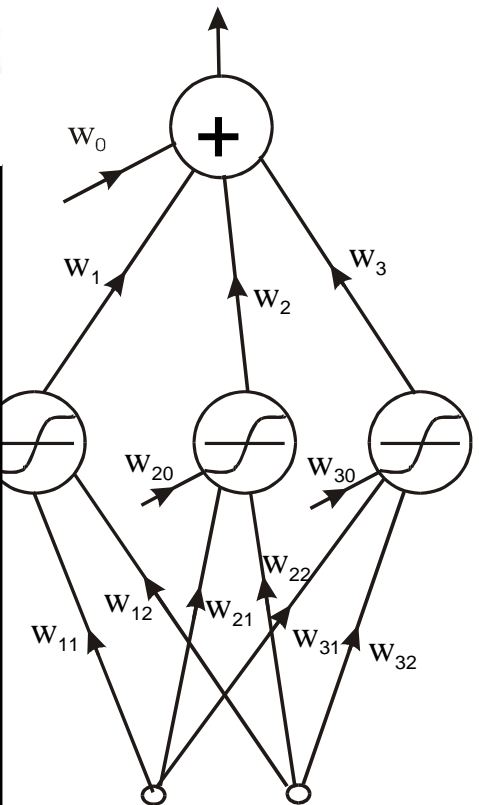
CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP:689 45 44 70



RBFN

(Aproximadores locales)



MLP

(Sustituir sumador por sigmoide)

(Aproximadores globales)

ATSC-DTC/UCIIM



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP: 689 45 44 70



icas

s:

- rapidez (implementadas en paralelo)
- degradación suave ante fallos (por distribución)
- gran potencia expresiva (sin explosión dimensional)

bles:

- difícil elección
- difícil dimensionado
- difícil y lento entrenamiento (dan lugar a muchos mínimos locales) (consecuencia: baja capacidad de seguimiento)
- difícil interpretación

ATSC-DTC/UCHIM

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP:689 45 44 70



isto que derivan de los Perceptrones con activación blanda.

isto que tienen alta capacidad expresiva; en puridad,

LP con una capa oculta (dos capas) del adecuado número de unidades
establecer cualquier correspondencia entre entrada y salida

no es constructivo

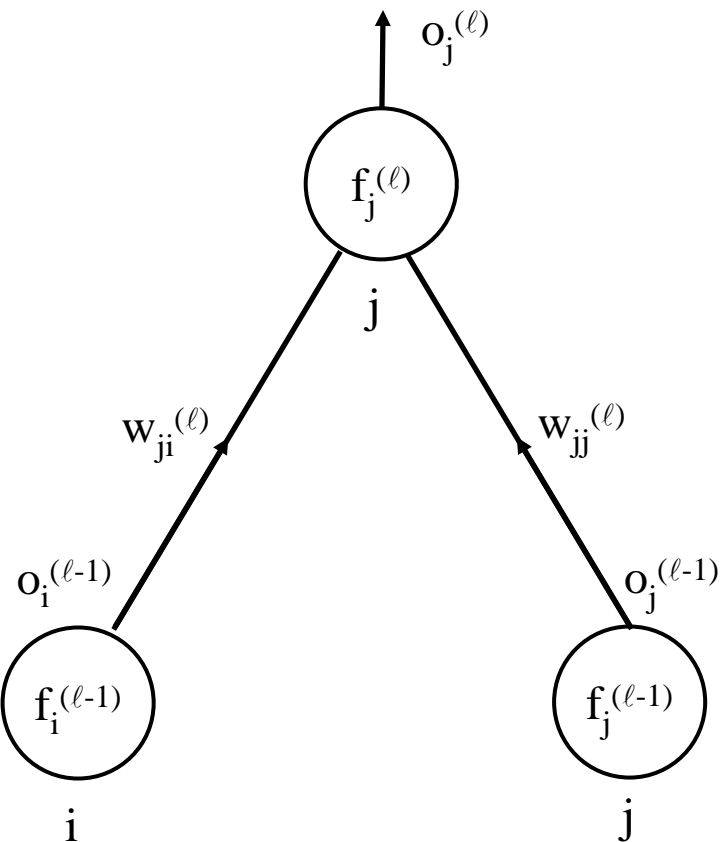
no implica que no pueda ser más eficaz utilizar dos capas ocultas)

ATSC-DTC/UCHIM

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP: 689 45 44 70

ento del MLP: **El algoritmo de Retropropagación (BP)**
(Werbos; Parker; Rumelhart)



ATSC-DTC/UCHIM

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP:689 45 44 70

la utilización de la regla de la cadena para calcular gradientes:

$$\frac{1}{2} \sum_{j'=1}^{N_L} (d_{j'} - o_{j'}^{(L)})^2$$

$$= \frac{\partial C(\mathbf{w})}{\partial o_j^{(\ell)}} \frac{\partial o_j^{(\ell)}}{\partial w_{ji}^{(\ell)}} = \frac{\partial C(\mathbf{w})}{\partial o_j^{(\ell)}} o_i^{(\ell-1)} f_j'^{(\ell)} \quad (\mathbf{w}_{j0}^{(\ell)} : \sin o_i)$$

$$\left(o_j^{(\ell)} = f_j^{(\ell)} \left(w_{j0}^{(\ell)} + \sum_{i'} w_{ji'}^{(\ell)} o_{i'}^{(\ell-1)} \right) \right)$$

$$= \begin{cases} -(d_j - o_j^{(L)}) & , \ell = L \\ \sum_{n_{\ell+1}=1}^{N_{\ell+1}} \frac{\partial C(\mathbf{w})}{\partial o_{n_{\ell+1}}^{(\ell+1)}} \frac{\partial o_{n_{\ell+1}}^{(\ell+1)}}{\partial o_j^{(\ell)}} = \sum_{n_{\ell+1}=1}^{N_{\ell+1}} \frac{\partial C(\mathbf{w})}{\partial o_{n_{\ell+1}}^{(\ell+1)}} w_{n_{\ell+1}j}^{(\ell+1)} f_{n_{\ell+1}}'^{(\ell+1)} & , \ell < L \end{cases}$$

$$\left(o_{n_{\ell+1}}^{(\ell+1)} = f_{n_{\ell+1}}^{(\ell+1)} \left(w_{n_{\ell+1}0}^{(\ell+1)} + \sum_{j'=1}^{N_L} w_{n_{\ell+1}j'}^{(\ell+1)} o_{j'}^{(\ell)} \right) \right)$$

ATSC-DTC/UCHIM



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP: 689 45 44 70

la última expresión:

puede calcular $\partial C/\partial o$ para la capa de salida directamente consecutivamente, para $\ell = L-1, L-2, \dots, 1$ (¡retropropagación!)

Introducir los valores resultantes en la expresión de $\partial C/\partial w$ para completar el

Se observa que, si se emplea $\Delta_j^{(\ell)} = -\frac{\partial C(\mathbf{w})}{\partial o_j^{(\ell)}} f_j'^{(\ell)}$, se tiene

$$w_{ji}^{(\ell)}(k+1) = w_{ji}^{(\ell)}(k) + \eta^{(\ell)} \Delta_j^{(\ell)}(k) o_i^{(\ell-1)}(k)$$

regla análoga al LMS, con $\Delta_j^{(\ell)}$ en el papel del error: por eso se llama regla Delta Generalizada.

Se recomienda que η esté entre 0 y 1: 0 para el entrenamiento)

Recuérdese que, para la th, $f' = 1-o^2$)

ATSC-DTC/UCIIM



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP: 689 45 44 70



ones

El algoritmo BP se puede aplicar para cualquier función de coste sin más que calcular $\partial C/\partial o$ para $\ell = L$.

El algoritmo análogo al visto se pueden extender los otros métodos de búsqueda para el MLP.

Indicar el algoritmo que ha de aplicarse para el cálculo de las derivadas parciales.

ATSC-DTC/UCIIM

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP:689 45 44 70



Aplicación del MLP

Dimensionado

En la mayoría de los casos basta con una capa oculta; la necesidad de 2 es muy infrecuente en estimación, y muy infrecuente para decisión.

Para la decisión (como habitualmente) conviene un nodo por clase (uno solo en binarios).

Se dan algunas prácticas para el número de unidades:

Para una capa oculta: $2N+1 \lesssim N_1 \lesssim 3N$ (Kudrycki)

Para dos capas ocultas: $N_1 \lesssim 3N, N_2 \lesssim 2N_3$ (Lippmann)

Se recomienda ir de menor a mayor.

ATSC-DTC/UCIIM

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP: 689 45 44 70



do caso, el número de muestras disponibles limita el número de pesos:
ene no bajar de 10 muestras por peso (y un límite absoluto es 1 muestra
so)

en
goritmos constructivos o de crecimiento: establecen la arquitectura a
rtir de una versión subdimensionada, añadiendo elementos;

goritmos de reducción o de poda: justo al revés;

proporcionan ventajas en prestaciones y, obviamente, en carga
acional para el entrenamiento (sobre todo, los primeros).

Familias de) Algoritmos de Poda.

ATSC-DTC/UCIIM

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70
ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

ritmo BP

pasos $\eta(\ell)$ han de tantearse (como en el de gradiente) entre valores pequeños (típicamente, décimas a milésimas); y debe notarse que conviene usar valores mayores en las capas más lejanas de la salida: $\Delta(\ell)$ va incorporando valores f' .

Se debe notar que pueden producirse efectos de parálisis por saturación temprana de neuronas internas (por lo que hay que cuidar mucho los pasos).

Se puede combatir tal saturación temprana:

- incrementando artificialmente las derivadas;
- reduciendo el nivel de las entradas;
- modificando la función de coste (p.ej.: no corrigiendo momentáneamente para muestras bien clasificadas).

ATSC-DTC/UCIIM

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP: 689 45 44 70

én es obvio que no pueden inicializarse los pesos a cero (se hacen nulas si no hay entrenamiento); típicamente se inicializan en valores aleatorios pequeños (p.ej., tomados de una distribución $U[-0.5, 0.5]$);

Por último, ha de señalarse que, dado lo complicado de la topografía del espacio de búsqueda (C vs. w), es muy habitual que los algoritmos de gradiente descendente o Newton no den ventaja respecto al BP (suele ser preferible bajar el escalón).

ATSC-DTC/UCHIM



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

--

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

as

e lo habitual

nuestras representativas del problema,

posibilidad de selección, ...

ene normalizar las componentes de las muestras (típicamente a potencia a o valor de pico unitario) para evitar el dominio de alguna(s) de ellas en o de la búsqueda.

ene aleatorizar el orden de presentación de las muestras, pero ciclar la tación de la secuencia resultante hasta la detención (se evitan así mas de inestabilidad en la búsqueda).

ATSC-DTC/UCIIM



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

--

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

problemas sencillos (MLP resultantes de tamaño reducido), suele facilitar el aprendizaje representar las muestras mediante códigos sencillos (tipo 1 bit: Barómetro, y análogos).

que ser cuidadoso con posibles preprocesados de las muestras: sólo son defendables cuando se tiene la certeza de eliminar información que se sabe que es irrelevante para la resolución del problema abordado.

ATSC-DTC/UCHIM



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

--

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP:689 45 44 70



Características intrínsecas de los MLP

Reutilización es mala: cuando se cambia una arquitectura (aunque sea simplemente para dividir una clase en dos, p. ej.), no se puede reaprovechar el trabajo ya hecho.

MLP no son adecuados para problemas espaciotemporales (datos de la forma (x, t)), al tratarse de estructuras estrictamente progresivas.

Características:

- preprocesado temporal
- versiones realimentadas (son delicadas)

CONTACTO: ATSC-DTC/UCHIM

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP: 689 45 44 70



Hay tres tipos de parámetros:

- Centroides m_j
- Varianzas V_j (suele haber circularidad : v_j)
- Costos de salida w_j

...

Efectos de las unidades son locales.

Métodos (básicos) de entrenamiento:

- Selección aleatoria de los centroides;
- Preestablecimiento de los centroides;
- Algoritmos de búsqueda convencionales (gradiente).

ATSC-DTC/UCIIM

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP:689 45 44 70



ento por selección aleatoria

centroides se seleccionan aleatoriamente de entre las muestras;

varianzas de las Gauss se estiman agrupando muestras por proximidad, cuando

$$v'_j = \frac{1}{\#j} \sum_{k_j} \left\| \mathbf{x}^{(k_j)} - \mathbf{m}_j \right\|_2^2$$

dimensionando este valor: $v_j \cong (1.5 - 2) v'_j$

de, al agrupar, no se consideran “muestras lejanas”)

de salida se entrenan convencionalmente (LMS, p. ej.)

ATSC-DTC/UCHIM

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP: 689 45 44 70

ento por preselección

ntroides se fijan mediante un proceso previo : p. ej., un modelado EM o
upamiento tipo K-medias;
zas y pesos, como en el caso anterior.

ntuario de lo que puede parecer, no deben esperarse grandes ventajas: EM
edias tienden a modelar densidades de probabilidad, y los costes típicos a
a la salida estimaciones de las probabilidades “a posteriori”; y entre unas
no hay una correspondencia directa: recuérdese la forma del clasificador
en EM presentada en la Unidad 8).

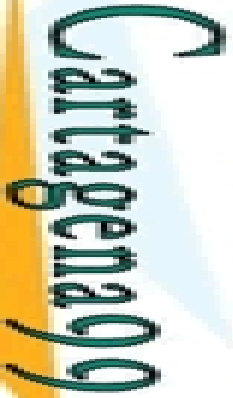
ATSC-DTC/UCIIM



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

--

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70



ento por búsqueda directa (gradiente)

alida y supuesto coste cuadrático (las generalizaciones son inmediatas):

$$\frac{1}{2}(d - o)^2 = \frac{1}{2} \left(d - \sum_j w_j \exp \left[-\frac{\|\mathbf{x} - \mathbf{m}_j\|_2^2}{2v_j} \right] \right)^2$$

$$\frac{\partial C}{\partial w_j} = -(d - o) \exp \left[-\frac{\|\mathbf{x} - \mathbf{m}_j\|_2^2}{2v_j} \right]$$

$$\frac{\partial C}{\partial \mathbf{m}_j} = -(d - o) \exp \left[-\frac{\|\mathbf{x} - \mathbf{m}_j\|_2^2}{2v_j} \right] \frac{\mathbf{x} - \mathbf{m}_j}{v_j}$$

$$\frac{\partial C}{\partial v_j} = -(d - o) \exp \left[-\frac{\|\mathbf{x} - \mathbf{m}_j\|_2^2}{2v_j} \right] \frac{\|\mathbf{x} - \mathbf{m}_j\|_2^2}{2v_j^2}$$

particularizadas para la iteración k y con sus correspondientes escalones, rtan en la expresión general del algoritmo de gradiente.

ATSC-DTC/UCIIM

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP: 689 45 44 70



ones

queda directa no suele dar buenos resultados: tanto por la existencia de los locales como por la dificultad de asignar η_w , η_m , η_v .

General, la dependencia de los resultados con $\{v_j\}$ es grande: puede resultar conveniente expresarlas de la forma $\{\alpha_j v\}$ y realizar una exploración de v (o puede procederse así por regiones).

cabe esperar, las RBFN dan muy buenos resultados en problemas con muchas variaciones locales.

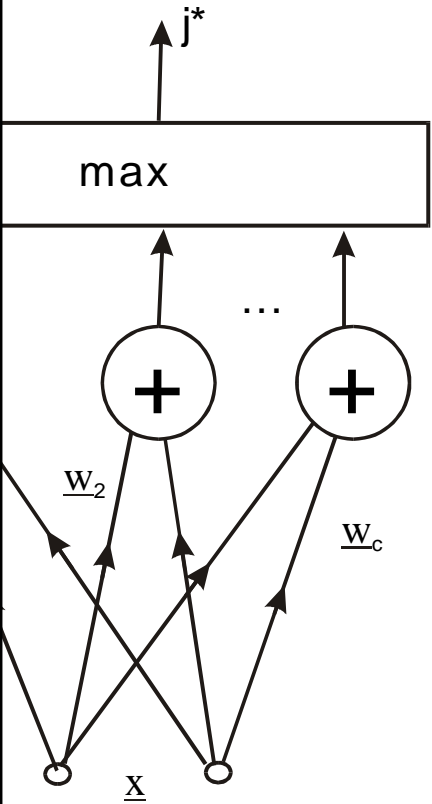
ATSC-DTC/UCHIM

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP:689 45 44 70

neuronales del Agrupamiento

mediata la estructuración en paralelo de los métodos competitivos; así, el (Básico) de Kohonen es asociable con la **Red (Básica) de Kohonen**:



$$w_j(k+1) = \begin{cases} w_j(k) + \eta(k)[\mathbf{x}^{(k)} - w_j(k)] & , j = j^* \\ w_j(k) & , j \neq j^* \end{cases}$$

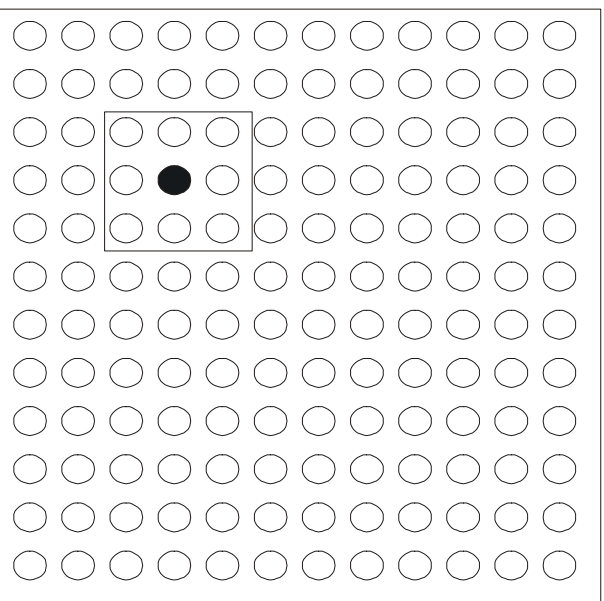
ATSC-DTC/UCHIM



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP:689 45 44 70

De esta forma (NN) propuso Kohonen la aplicación de algoritmos iterativos con inicialización por arrastre, que denominó Mapas autoorganizables (de Características)) (SO(F)M: Self-Organizing (Feature) Map) porque recubren el espacio de las muestras con una “red” de neuronas interconectadas según los entornos que se definan.



ATSC-DTC/UCHIM

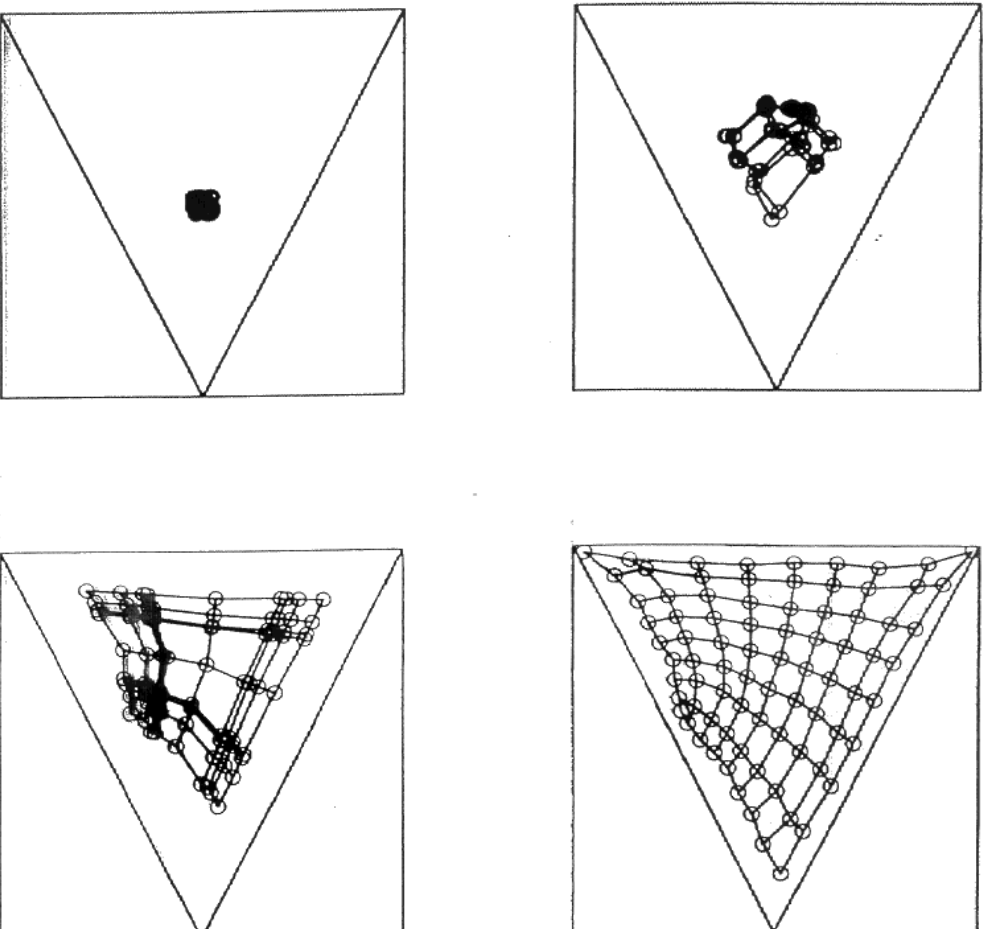
Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

--

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP:689 45 44 70

e del SOFM



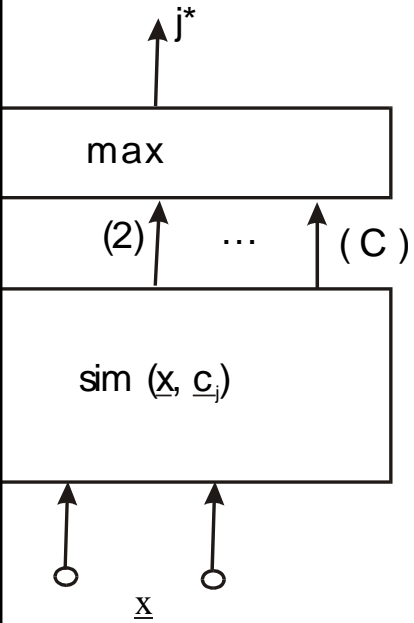
ATSC-DTC/UCHIM



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP:689 45 44 70

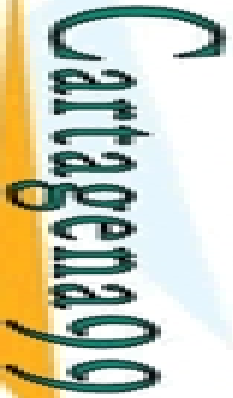
generalizaciones son obvias para otras medidas de similitud: basta sustituir a de productos escalares por otra capa de entrada en que se calculen las similitudes con los representantes $\{c_j\}$



actualizando los $\{c_j\}$ del modo (acretivo) que corresponda.

también lo son para modos no acretivos, e incluso para redes no competitivas (optimización de un objetivo).

ATSC-DTC/UCIIM



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP:689 45 44 70



Clasificadores Vectoriales con Aprendizaje (“Learning Vector Quantization”, LVQ)

Propusieron por Kohonen como paso siguiente a su Red Básica para clasificación: es un caso claro de aprovechar el agrupamiento para reducir el tiempo de diseño de un clasificador.

Algunas versiones fundamentales (y algunas variantes):

$$\mathbf{w}_j(k+1) = \begin{cases} \mathbf{w}_j(k) + \eta(k)[\mathbf{x}^{(k)} - \mathbf{w}_j(k)], & \text{si } \begin{cases} j = \arg[\max_i \{\mathbf{w}_j^T(k)\mathbf{x}^{(k)}\}], \\ \text{y } \mathbf{x}^{(k)} \text{ está en su clase} \end{cases} \\ \mathbf{w}_j(k) - \eta(k)[\mathbf{x}^{(k)} - \mathbf{w}_j(k)], & \text{si } \begin{cases} j = \arg[\max_i \{\mathbf{w}_j^T(k)\mathbf{x}^{(k)}\}], \\ \text{y } \mathbf{x}^{(k)} \text{ no está en su clase} \end{cases} \\ \mathbf{w}_j(k), & \text{en otro caso} \end{cases}$$

CONTACTO: ATSC-DTC/UCHIM

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP: 689 45 44 70



ay adaptación si hay error y además $\mathbf{x}^{(k)}$ está en una cierta ventana entre el ganador \mathbf{w}_j y el más próximo de los que pertenecen a su clase, \mathbf{w}_m ; entonces:

$$\mathbf{w}_j(k+1) = \mathbf{w}_j(k) - \eta(k)[\mathbf{x}^{(k)} - \mathbf{w}_j(k)]$$

$$\mathbf{w}_m(k+1) = \mathbf{w}_m(k) + \eta(k)[\mathbf{x}^{(k)} - \mathbf{w}_m(k)]$$

etenerlo)

adaptación si hay clasificación correcta y además $\mathbf{x}^{(k)}$ está en una cierta ventana entre el peso ganador \mathbf{w}_j y el más próximo, \mathbf{w}_m ; entonces

$$\begin{cases} \mathbf{w}_j(k) + \eta(k)[\mathbf{x}^{(k)} - \mathbf{w}_j(k)], & \text{si } \mathbf{w}_m \text{ representa a otra clase} \\ \mathbf{w}_m(k) - \eta(k)[\mathbf{x}^{(k)} - \mathbf{w}_m(k)], & \text{si } \mathbf{w}_m \text{ representa a la misma clase} \\ \mathbf{w}_m(k) - \varepsilon\eta(k)[\mathbf{x}^{(k)} - \mathbf{w}_m(k)], & \end{cases}$$

($0.1 \leq \varepsilon \leq 0.5$)

ATSC-DTC/UCHIM

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

 ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP: 689 45 44 70