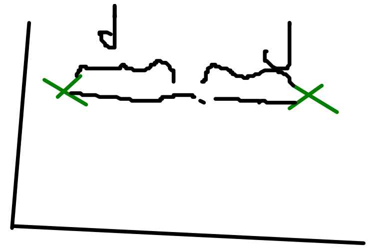
REDES DE BASE RADIAL

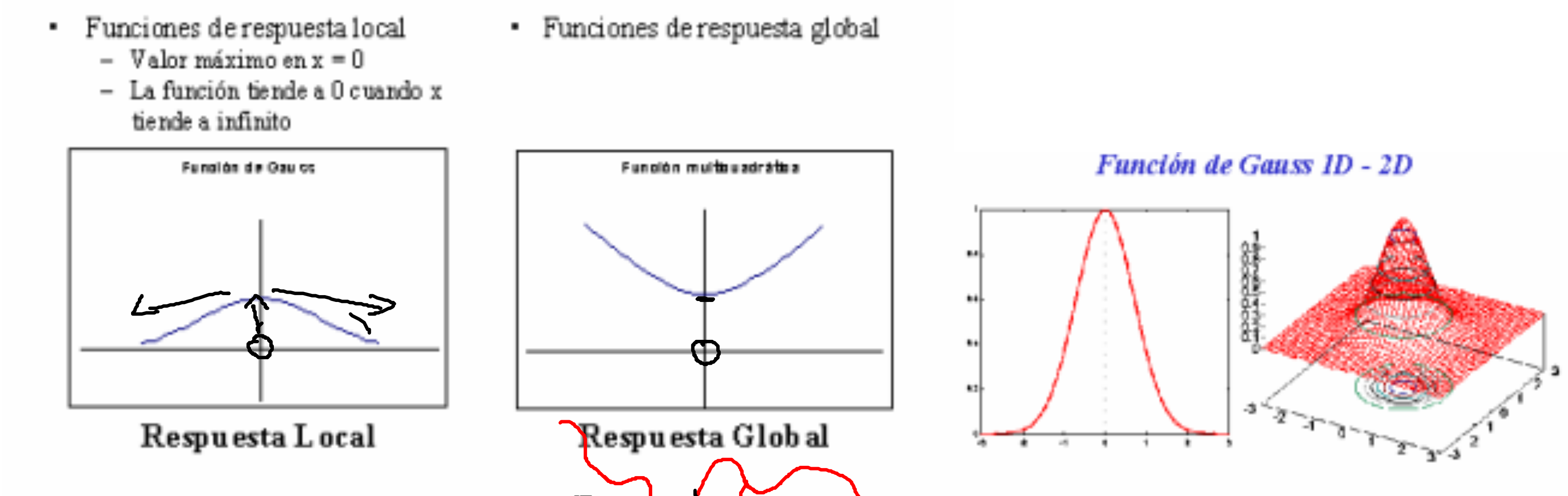
FUNCIONES DE BASE RADIAL

Son funciones cuya salida depende de la distancia a un punto denominado ***Centro***.



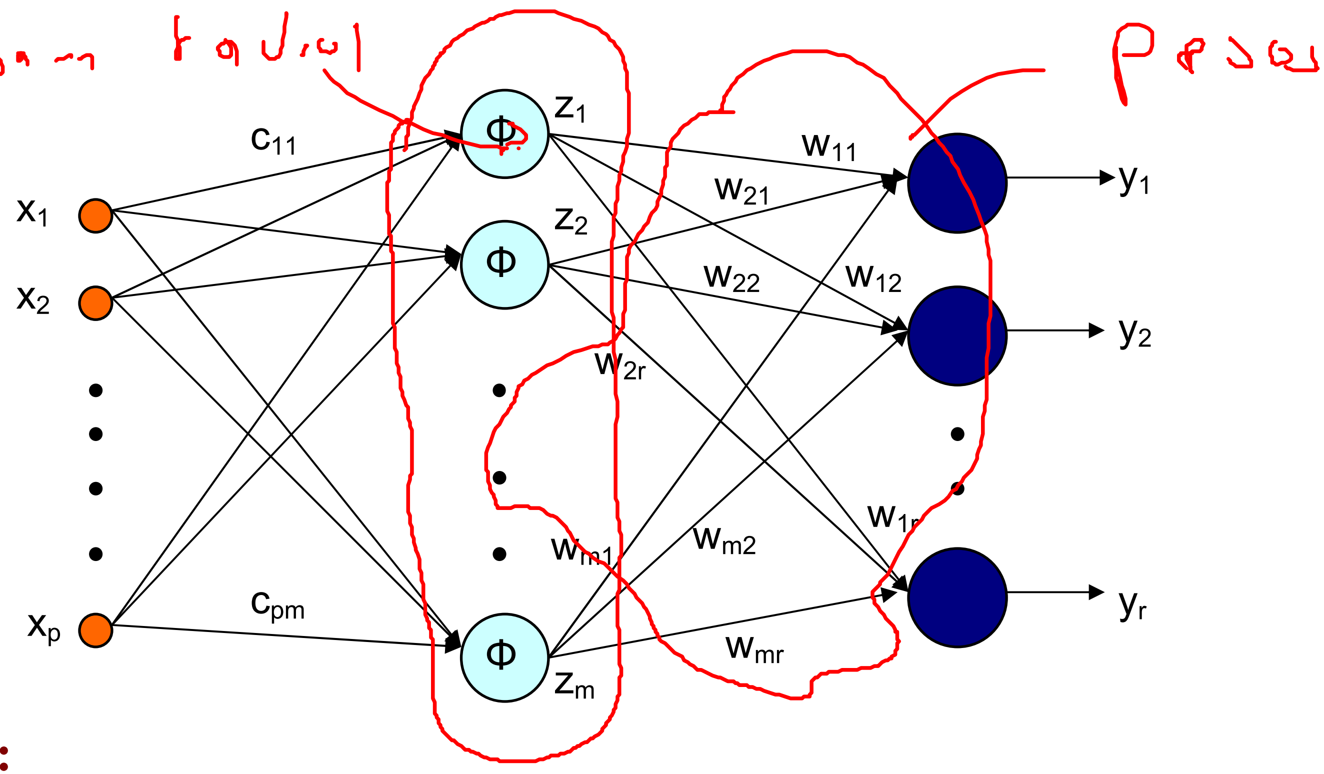
Características:

- Simétricas respecto de x=0
- Se definen con al menos dos parámetros.
  - ***Centro***: Punto donde la función posee un extremo.
  - ***Anchura***: Magnitud de la variación de la función según se aleja del centro.



ARQUITECTURA DE UNA RBFN

Función Base radial Pesos



Caracterísicas:

- Capa de entrada: reciben las señales del exterior, no realizan ningún preprocesado.
- Capa Oculta: reciben las señales de la capa de entrada y realizan una ***transformación local*** y no lineal sobre dichas señales (Diferencia con el MLP).
- Capa de Salida: Se realiza una combinación lineal de las activaciones de las neuronas de la capa oculta y actúa como salida de la red.

Activaciones de las neuronas en una RBFN

Capa Salida: Cada elemento de procesado calcula su valor neto como una combinación lineal de las salidas de los elementos de procesado de la capa oculta. La función de activación y transferencia es lineal, por lo tanto:

Para un patrón n, X(n)=(x1(n),x2(n),…..xp(n)), la salida de la red asociada a cada elemento *k* de la capa de salida se obtiene de la siguiente manera:

$$y_k(n) = \sum_{i=1}^{m} w_{ik} z_i(n) + \mu_k \qquad para \quad k = 1,2,...,r$$

Donde los W_{ik} son los pesos asociados al elemento k de la capa de salida y el elemento i de la capa oculta, que ponderan cada uno las salidas $z_i(n)$ del elemento de procesado de la capa oculta correspondiente.

El término μ_k es un término denominado umbral y está asociado a cada una de los elementos de procesado de la capa de salida.

Capa Oculta: Cada elemento de procesado, *i*, de la capa oculta tiene asociada una función de base radial de tal manera que representa una clase o categoría, donde dicha clase viene dada por (C_i, d_i). C_i representa un centro de cluster (pesos asociados a cada neurona i) y d_i representa la desviación, anchura o dilatación de la función de base radial asociada a dicho elemento.

La salida de cada elemento de la capa oculta $z_i(n)$ se calcula como la distancia que existe ente el patrón de entrada X(n) al centro del cluster C_i ponderada inversamente por d_i y aplicando después a ese valor una función de base radial.

Base radial

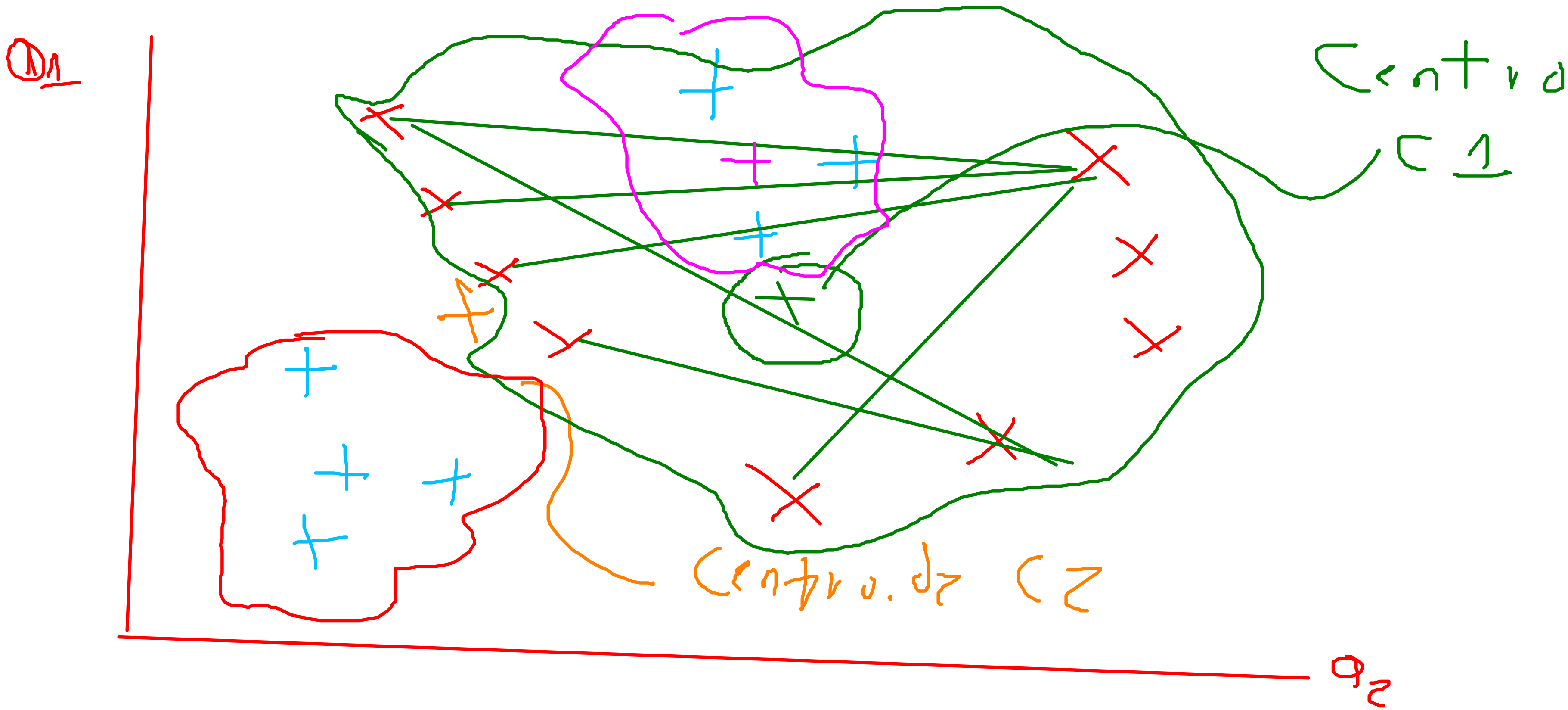
centroide

$$z_i(n) = \Phi\left(\frac{\left(\sum_{j=1}^{p}\left(x_j(n) - c_{ji}\right)^2\right)^{1/2}}{d_i}\right) \qquad para \qquad i = 1,2....m$$

Distancia

Donde Φ es una función de base radial, dentro de éstas la más utilizada es la función

Gausiana: $\Phi(r) = e^{\left(-r^2/2\right)}$.



a_1
Centro
C 1
Centro de C2
a_2

Resumen

1) FUnciones de base radial, Depende de un centroide y una distancia
Funciones locales: UN sólo centroide
Funciones globales: Sobre todo el conjunto de centroides, en 3D se verian como un plano con un huecos.

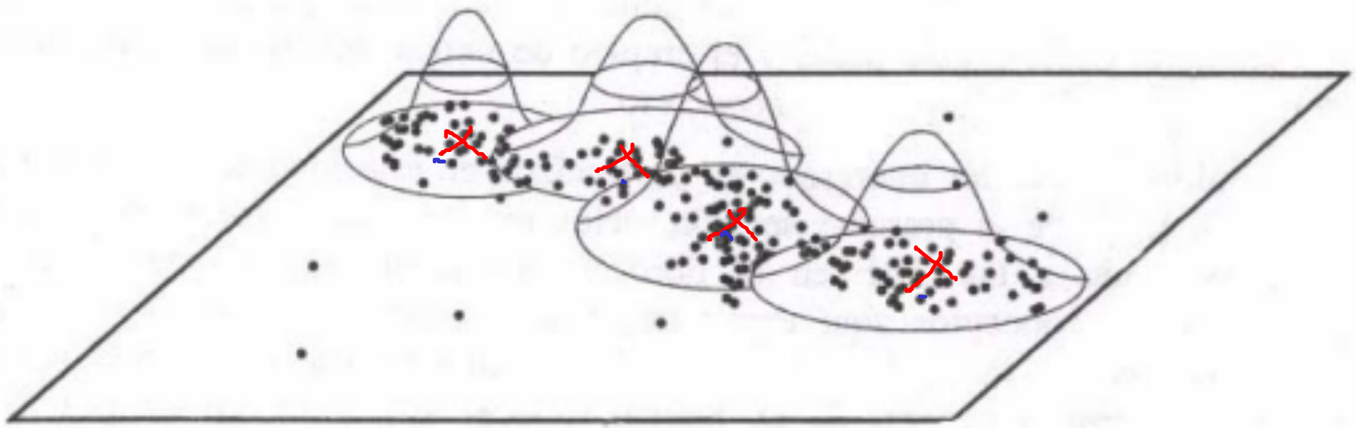
2) Estructura RN

a) 3 capas, Capa entrada: SOlo se encarga de direccionar las entradas hacia la capa oculta
Capa oculta: Funciones de base radial: Centroide y la distancia
Capa salida: TOma las salidas de la capa oculta y las multiplica por los pesos correspondiente

CARACTERÍSTICAS

Las funciones de base radial tienen todas ellas un **carácter Local** pues son funciones que alcanzan un nivel cercano al máximo de su recorrido cuando el patrón de entrada X(n) está próximo al centro de la neurona. A medida que el patrón se aleja del centro, el valor de la función va tendiendo al valor mínimo de su recorrido.

Las salidas de las redes de neuronas de base radial son, por tanto una combinación lineal de gausianas, cada una de las cuales se activa para una determinada porción del espacio definido por los patrones de entrada.



APRENDIZAJE. ENTRENAMIENTO

El entrenamiento de este tipo de redes, determina todos los parámetros de la red.

- Parámetros de la capa de salida: Pesos, W
- Parámetros de la capa Oculta: Centros, C y desviaciones asociadas d.

La determinación de los parámetros de la capa oculta, se realiza mediante la optimización en el espacio de entradas, ya que cada neurona va a representar una zona diferente en dicho espacio.

La determinación de los parámetros de la capa de salida, la optimización se realiza en base a las salidas que se desea obtener (salidas deseadas), ya que en su globalidad, las redes de base radial se utilizan para aproximar relaciones entre el conjunto de variables de entrada y salida que definen el problema.

Aprendizaje Híbrido

- Fase No supervisada: Determinación de parámetros de la capa oculta.
- Fase Supervisada: Determinación de pesos en capa de salida.

Fase No Supervisada

Puesto que las neuronas ocultas se caracterizan porque representan zonas diferentes del espacio de entradas, los centros y las desviaciones deben de ser calculados con este objetivo (clasificar el espacio de entradas en diferentes clases).

Determinación de Centros:

K = # centroids

- Algoritmo ~~K-medias~~ K-Means
- Mapas de Kohonen

Determinación de Desviaciones

Se deben de calcular de manera que cada neurona de la capa oculta se active en una región del espacio de entradas y de manera que el solapamiento de las zonas de activación de una neurona sea lo más ligero posible, para suavizar así la interpolación

Determinación de Desviaciones

Varias aproximaciones:

- Media Uniforme de las distancias euclídeas del centro C_i a los p centros más cercanos.

$$d_i = \frac{1}{p}\sum_{p}\left\|C_i - C_p\right\|$$

- Media geométrica de la distancia del centro a sus dos vecinos más cercanos.

$$d_i = \sqrt{\left\|C_i - C_t\right\|\left\|C_i - C_s\right\|}$$ con C_t y C_s los más cercanos a C_i

Fase Supervisada

Se utiliza la técnica de correción de Error (Adaline, Perceptrón Multicapa.)

Minimización de la función error dada por la salida de la red. $$E = \frac{1}{N}\sum_{n=1}^{N} e(n)$$

Donde N es el número de patrones, y e(n) es el error cometido por la red para el patrón X(n), que viene dado por:

$$e(n) = \frac{1}{2}\sum_{k=1}^{r}\left(s_k(n) - y_k(n)\right)^2$$

Donde Y(n) es la salida de la red, S(n) es la salida deseada para un patron X(n)

Determinación de pesos:

$$w_{ik}(n) = w_{ik}(n-1) - \mu \frac{\delta e(n)}{\delta w_{ik}}$$

$$u_k(n) = u_k(n-1) - \mu \frac{\delta e(n)}{\delta u_k}$$

Teniendo en cuenta la expresión del error y que el peso w_{ik} y el umbral u_k únicamente afectan a la neurona de salida k se obtiene que:

$$\frac{\delta e(n)}{\delta w_{ik}} = -\left(s_k(n) - y_k(n)\right)\frac{\delta y_k(n)}{\delta w_{ik}} \qquad \frac{\delta y_k(n)}{\delta w_{ik}} = \Phi_i(n)$$

$$\frac{\delta e(n)}{\delta u_k} = -\left(s_k(n) - y_k(n)\right)\frac{\delta y_k(n)}{\delta u_k} \qquad \frac{\delta y_k(n)}{\delta u_k} = 1$$

Salida deseada

Salida obtenida

Quedando las ecuanciones de cambio como:

Factor Aprendiz

$$\boxed{w_{ik}(n) = w_{ik}(n-1) + \mu\left(s_k(n) - y_k(n)\right)\phi_i(n)}$$

$$u_k(n) = u_k(n-1) + \mu\left(s_k(n) - y_k(n)\right)$$

para k = 1,2,...,r y para i = 1,2,...m

Aprendizaje Totalmente Supervisado

A diferencia con el método anterior, este tipo de aprendizaje no conserva, en principio, las propiedades o caracterísicas locales de las redes de base radial. En este caso, todos los parámetros de la red, centros, amplitudes, pesos y umbrales, se determinan de manera completamente supervisada y con el objetivo de minimizar el error cuadrático medio.

En este proceso, en ningún momento se tiene en cuenta que las amplitudes alcancen valores tales que el solapamiento de las activaciones de las neuronas de la capa oculta sea lo más suave posible. Así que en principio, esa característica de localidad se puede perder.

Cálculo de parámetros

$$w_{ik}(n) = w_{ik}(n-1) - \mu \frac{\delta e(n)}{\delta w_{ik}} \qquad c_{ij}(n) = c_{ij}(n-1) - \mu \frac{\delta e(n)}{\delta c_{ij}}$$

$$u_k(n) = u_k(n-1) - \mu \frac{\delta e(n)}{\delta u_k} \qquad d_i(n) = d_i(n-1) - \mu \frac{\delta e(n)}{\delta d_i}$$

Redes de Base Radial Frente a Perceptron Multicapa

Perceptron Multicapa

Uso de Funciones de Transferencia Sigmoidales. Relaciones globales entre los datos de entrada y la salida.

Aprendizaje Lento: Cambio de un solo peso ante un patrón, provoca cambios en la salida para todos los patrones presentados anteriormente, reduciéndose así el efecto de previos ciclos de aprendizaje y retrasando la convergencia.

Redes de Base Radial

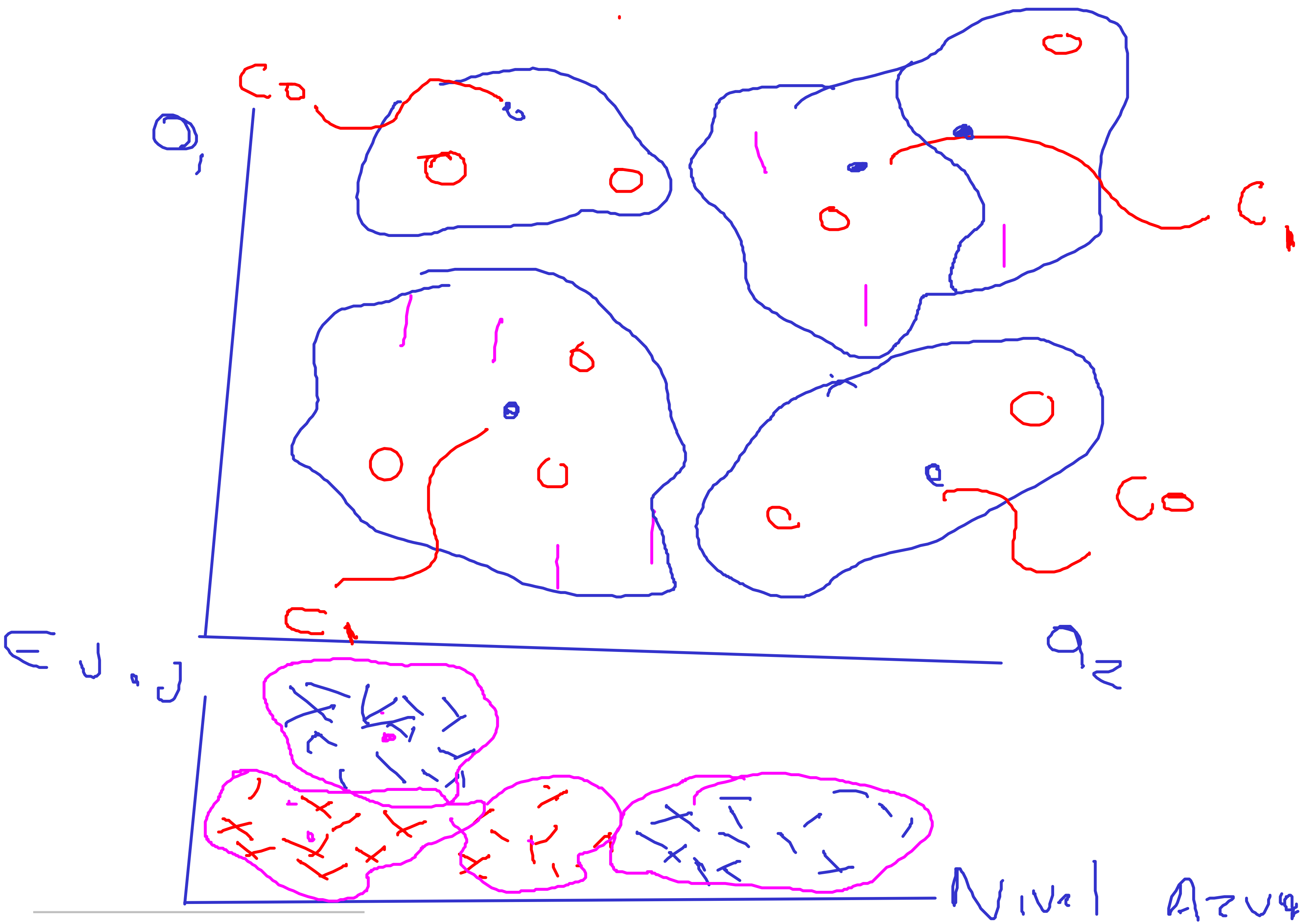
Cada neurona de la capa oculta se espcializa en una determinada región del espacio de entradas. La relación entre la entrada y la salida es una suma de funciones no lineales y locales.

Aprendizaje más rápido: el cambio de peso sólo afecta a la neurona oculta asociada a dicho peso (sólo a un grupo de patrones, pertenecientes a la clase que respresenta dicha neurona oculta.

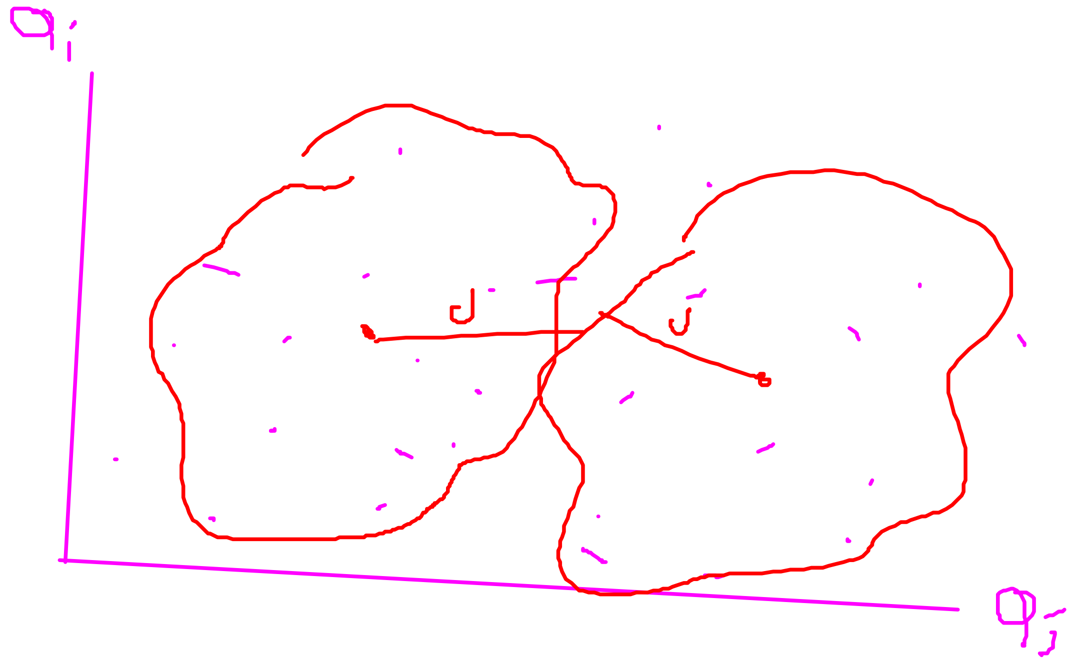
Es menos sensible al orden de presentación de patrones.

Inconvenientes:

- En ciertos casos, es necesario un elevado número de neuronas en la capa oculta. Pérdida de generalización.
- El número de neuronas ocultas aumenta exponencialmente con la dimensión del espacio de entradas.



C0
a1
C1
C0
C1
a2
Nivel


qi
J
J
qj