

Seminario de solución de problemas de IA - II

1. Aplicación de redes neuronales para el reconocimiento de patrones



Dr. Diego Oliva

Redes Neuronales.

Modelo computacional



Redes neuronales artificiales.

Aprendizaje

El aprendizaje es la parte mas importante de una RNA.

Consiste en la determinación de los valores precisos de los pesos en todas sus conexiones que la capacite para la resolución del problema.

La modificación de pesos puede hacerse despues de introducir cada ejemplo o una vez introducidos todos ellos.



Redes neuronales artificiales.

Aprendizaje

La capacidad de una red para resolver un problema estará ligada al tipo de ejemplos de que dispone en el proceso de aprendizaje.

- Estos ejemplos deben poseer las siguientes características:
 - Ser significativo: Número suficiente de ejemplo, o no será capaz de adaptar sus pesos de forma eficaz.
 - Ser representativo: Ejemplos diversos, o se especializará con el conjunto más representativo.



Redes neuronales artificiales.

Aprendizaje

El proceso general de aprendizaje consiste en:

1. Ir introduciendo todos los ejemplos de entrenamiento y modificando los pesos siguiendo un determinado paradigma de aprendizaje (supervisado, no supervisado, por reforzamiento).
2. Una vez introducidos todos los ejemplos de entrenamiento se comprueba si se ha cumplido cierto criterio de convergencia, de no ser así, se repite el proceso y todos los ejemplos de entrenamiento se vuelven a introducir.



Redes neuronales artificiales.

Aprendizaje

La finalización del periodo de aprendizaje se puede terminar:

Con un número fijo de ciclos:

- Se decide “a priori” cantas veces se introduce el conjunto de entrenamiento y cuando se supere dicho número, se detiene el aprendizaje y se da por aceptada la red resultante.

Cuando el error descende por debajo de una cantidad preestablecida:

Habría que definir una función de error y decidir el valor aceptable de dicho error, se puede disponer de un criterio adicional de parada.

Cuando la modificación de pesos sea irrelevante:

En algunos modelos se hace que las conexiones vayan modificándose cada vez con menor intensidad hasta llegar a un momento en que ya no haya variaciones de pesos de ninguna conexión, entonces se dice que la red ha convergido



Redes neuronales artificiales.

Aprendizaje

Paradigmas de aprendizaje

- Supervisado:

Se usan un conjunto de ejemplos de entrenamiento entrada-salida.

Utiliza etiquetas para la distinción de las clases, el esquema del aprendizaje supervisado utilizará esta información para modificar las conexiones.

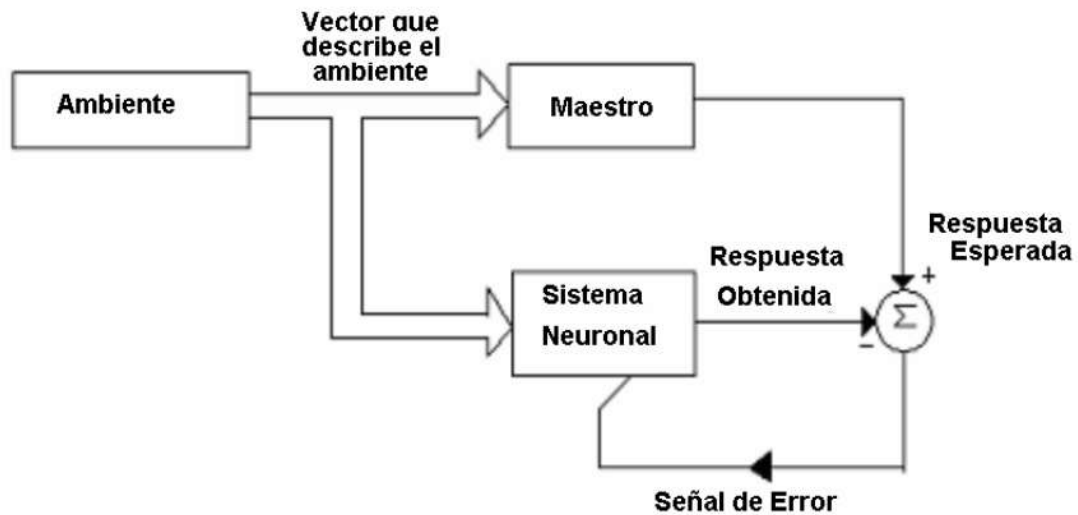
Se dice que hay un maestro encargado de determinar si la red se está comportando de forma adecuada, mediante la comparación en la salida y actuar en consecuencia modificando los pesos.



Redes neuronales artificiales. Aprendizaje

Paradigmas de aprendizaje

- Supervisado:





Redes neuronales artificiales.

Aprendizaje

Paradigmas de aprendizaje

- No Supervisado:

El conjunto de aprendizaje solo tiene información de la entrada, no hay nada que permita guiar el proceso de aprendizaje.

La red tratará de determinar características de los datos del conjunto de entrenamiento: rasgos significativos, regularidades o redundancias.

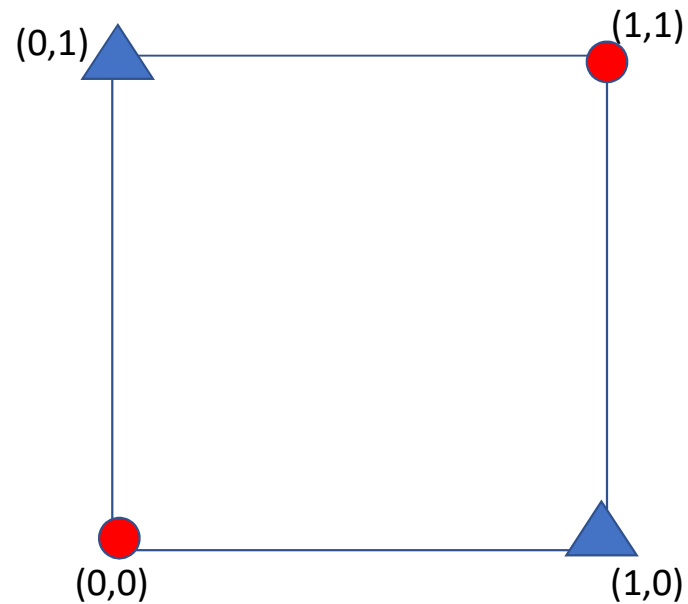
Se le conoce también como sistemas autoorganizados, debido a que la red se ajusta a partir de información interna.



PERCEPTRON MULTICAPA (MLP)

Clasificación de la function XOR.

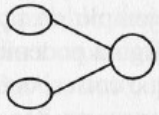
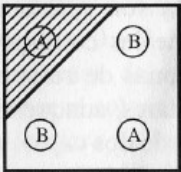
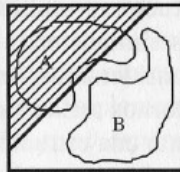
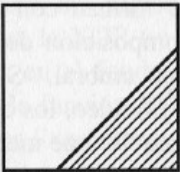
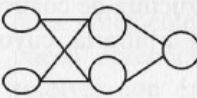
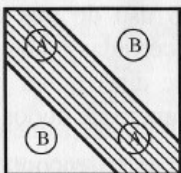
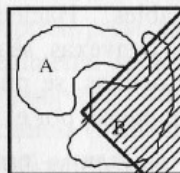
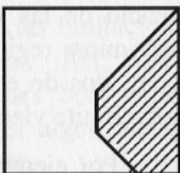
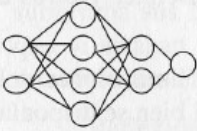
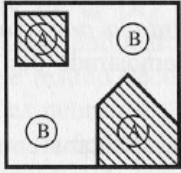
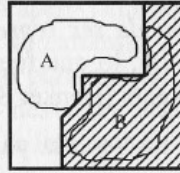
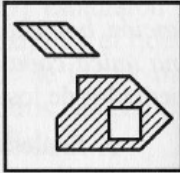
| Funcion XOR | |
|-------------|--------|
| ENTRADA | SALIDA |
| (0,0) | 0 |
| (0,1) | 1 |
| (1,0) | 1 |
| (1,1) | 0 |





PERCEPTRON MULTICAPA (MLP)

El Perceptron multicapa con dos capas de neuronas ocultas es capaz de discriminar regiones de forma arbitraria.

| Arquitectura | Región de decisión | Ejemplo 1: XOR | Ejemplo 2: clasificación | Regiones más generales |
|--|--------------------------------|--|---|---|
| Sin capa oculta  | Hiperplano (dos regiones) |  |  |  |
| Una capa oculta  | Regiones polinomiales convexas |  |  |  |
| Dos capas ocultas  | Regiones arbitrarias |  |  |  |



PERCEPTRON MULTICAPA (MLP)

Características:

Surgió como consecuencia de las limitaciones del perceptron en lo referente al problema de la separabilidad no lineal.

Diferentes autores han demostrado que el MLP es un aproximador universal, en el sentido de que cualquier función continua puede aproximarse con un MLP con al menos una capa oculta.

La habilidad del MLP para aprender a partir de un conjunto de ejemplos, aproximar relaciones no lineales, etc. hace que sea un modelo adecuado para abordar problemas reales.



PERCEPTRON MULTICAPA (MLP)

Estas redes han sido aplicadas con éxito para la resolución de problemas en una gran variedad de áreas diferentes, como:

- Reconocimiento de habla
- Reconocimiento de caracteres ópticos
- Reconocimiento de caracteres escritos
- Control de procesos
- Conducción de vehículos
- Diagnósticos médicos
- Predicción de series temporales, etc.



MLP: Arquitectura

Tiene sus neuronas agrupadas en capas de diferentes niveles:

- **Capa de entrada**

No actúan como neuronas, se encargan únicamente de recibir las señales o patrones que proceden del exterior y propagar dichas señales a todas las neuronas de la siguiente capa.

- **Capas ocultas**

- **Capa de salida**



MLP: Arquitectura

Tiene sus neuronas agrupadas en capas de diferentes niveles:

Capa de entrada

Capas ocultas

Las neuronas de las capas ocultas realizan un procesamiento no lineal de los patrones recibidos.

Capa de salida



MLP: Arquitectura

Tiene sus neuronas agrupadas en capas de diferentes niveles:

Capa de entrada

Capas ocultas

Capa de salida

Actúa como salida de la red, proporcionando al exterior la respuesta de la red para cada uno de los patrones de entrada.



MLP: Arquitectura

Las conexiones del MLP siempre están dirigidas hacia adelante.

Todas las neuronas de la red llevan asociado un umbral, que en el caso del MLP suele tratarse como una conexión más a la neurona, cuya entrada es constante e igual a 1.

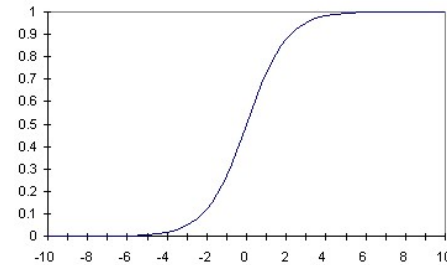
Generalmente, es una red totalmente conectada.



MLP: Arquitectura

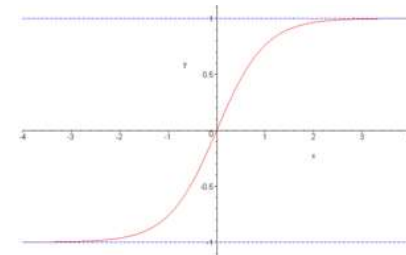
Las funciones de activación más utilizadas son:

Función sigmoideal $y = \frac{1}{1 + e^{-x}}$



Función tangente hiperbólica

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$



Ambas son funciones crecientes con dos niveles de saturación: el máximo, que proporciona salida 1, y el mínimo, salida 0 para sigmoideal y -1 para tangente hiperbólica.



MLP: Arquitectura

Generalmente, la función de activación en MLP es común en todas las neuronas de la red y es elegida por el diseñador, elección que se realiza únicamente basándose en los valores de activación que se desee que alcancen las neuronas. $[0,1]$ o $[-1,1]$

En ocasiones, y dependiendo de la naturaleza del problema, las neuronas de la salida se distinguen del resto de las neuronas de la red, utilizando otro tipo de función de activación, las más usadas son la función umbral y escalón. $\{0,1\}$



MLP: Arquitectura

MLP tiene 3 características distintivas:

- 1.El modelo de cada neurona de la red incluye una función de activación no lineal.
- 2.Las neuronas de las capas ocultas permiten que la red aprenda tareas complejas por la extracción progresiva de las características principales de los patrones de entrada.
- 3.La red presenta altos grados de conectividad, determinados por las sinápsis de la propia red.



MLP: Aprendizaje

La regla o algoritmo de aprendizaje es el mecanismo mediante el cual se van adaptando y modificando todos los parámetros de la red.

El objetivo del aprendizaje o entrenamiento es ajustar los parámetros de la red(pesos) con el fin de que las entradas presentadas produzcan las salidas deseadas, es decir, con el fin de minimizar la función de error E .



MLP: Aprendizaje

Las redes neuronales tipo MLP, han sido aplicadas satisfactoriamente para resolver problemas muy diversos y difíciles por medio del algoritmo conocido como retropropagación, el cual consta de dos etapas.

1. Etapa hacia adelante.

Se fijan los parámetros de la red y se presenta una señal de entrada a la red, que se propaga hacia adelante para producir una salida.

2. Etapa hacia atrás.

El error entre la salida deseada y la red se propaga hacia atrás.

Los parámetros de la red se modifican para minimizar el cuadrado de dicho error.



MLP: Aprendizaje

Los pasos que componen el proceso de aprendizaje del MLP son los siguientes:

1. Se inicializan los pesos de la red. Generalmente, esta inicialización es aleatoria y con valores cercanos al cero.
2. Se toma un patrón n del conjunto de entrenamiento $(X(n), S(n))$ y se propaga hacia la salida de la red obteniendo $Y(n)$



MLP: Aprendizaje

3. Se evalúa el error cuadrático cometido por la red para el patrón n utilizando:

$$e(n) = \frac{1}{2} \sum_{i=1}^n (s_i(n) - y_i(n))^2$$

4. Se modifican los pesos, entre mayor error mas modificación de pesos.

5. Se repiten los pasos 2, 3 y 4 para todos los patrones de entrenamiento, completando así una iteración o un ciclo de aprendizaje.



MLP: Aprendizaje

6. Se evalúa el error total E cometido por la red, llamado error de entrenamiento.

$$E = \frac{1}{N} \sum_{n=1}^N e(n)$$

7. Se repiten los pasos 2, 3, 4, 5 y 6 hasta alcanzar un mínimo del error de entrenamiento, para lo cual se realizan m ciclos de aprendizaje.



MLP: Diseño de la arquitectura

Cuando se aborda un problema utilizando MLP, uno de los primeros pasos a realizar es el diseño de la arquitectura de la red.

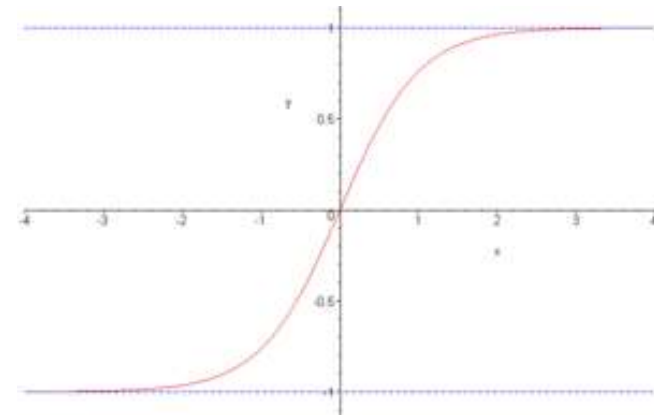
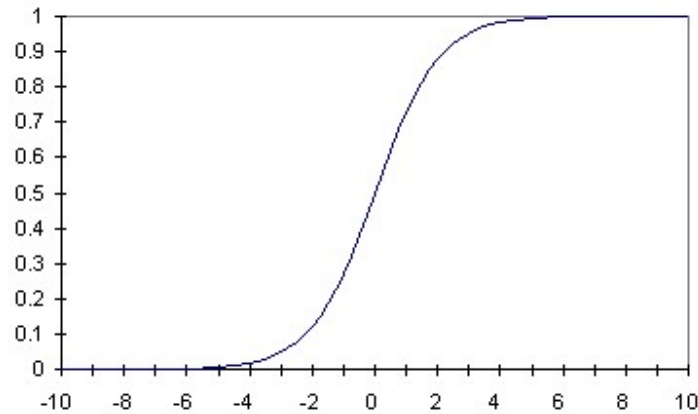
Este diseño implica la determinación de:

- La función de activación
- El número de neuronas
- El número de capas en la red



MLP: Diseño de la arquitectura

La elección de la función de activación se suele hacer basándose en el recorrido deseado, y el hecho de elegir una u otra, generalmente, no influye en la capacidad de la red para resolver el problema.





MLP: Diseño de la arquitectura

En lo que respecta al número de neuronas y capas, algunos de estos parámetros vienen dados por el problema y otros deben de ser elegidos por el diseñador.

Por ejemplo, tanto el número de neuronas de la capa de entrada y salida, vienen dadas por las variables que definen el problema.

Puede darse el caso que se dispone de un gran número de variables , algunas no podrían aportar información relevante a la red, y su utilización podría complicar el aprendizaje, por lo que sería conveniente realizar un análisis previo de las variables de entrada



MLP: Diseño de la arquitectura

El número de capas ocultas y de neuronas en cada capa, debe ser elegido por el diseñador.

No existe un método o regla que determine el número óptimo de neuronas ocultas para resolver un problema dado.

En la mayor parte de las aplicaciones prácticas, éstos parámetros se determinan por prueba y error.



MLP: Capacidad de generalización

A la hora de evaluar el comportamiento de un MLP es importante:

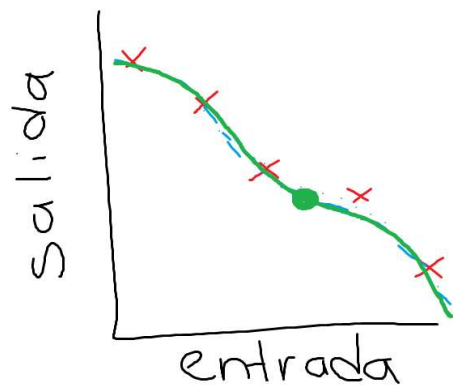
1. Saber si la red ha aprendido con éxito los patrones utilizados durante el aprendizaje.
2. Conocer el comportamiento de la red ante patrones que no se han utilizado durante el entrenamiento.



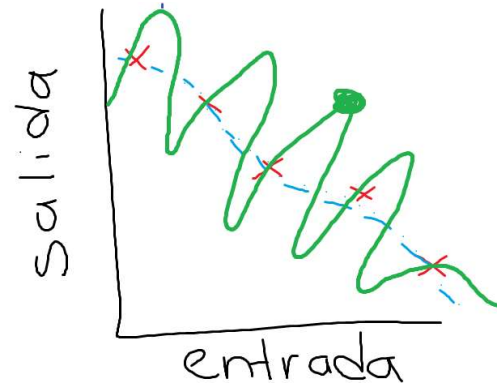
MLP: Capacidad de generalización

Es necesario que durante el aprendizaje, la red extraiga las características de las muestras, para poder así responder correctamente a patrones diferentes.

Esto se conoce como la capacidad de generalización de la red.



Buenas propiedades



Escasas propiedades

- Las cruces representan los patrones de entrenamiento
- La línea punteada, la función a aproximar
- La línea sólida, la salida que proporcionó el MLP ya entrenada



MLP: Capacidad de generalización

El sobreaprendizaje ocurre cuando la red ha aprendido correctamente los patrones de aprendizaje, pero no es capaz de responder adecuadamente ante los patrones de validación.

Este hecho puede producirse por:

- Un número elevado de ciclos de aprendizaje
- Por la utilización de demasiadas neuronas ocultas en la red

En problemas en los que las muestras poseen ruido, la red se ajustaría al ruido de los patrones, impidiendo la generalización.



Tarea 1

Investigar las distintas métricas que se utilizan para verificar si una red neuronal ha logrado identificar apropiadamente los patrones de un conjunto de datos.

* La tarea se entrega en formato PDF en la pagina de Schoology.