

**LAS RESPUESTAS NO SON GARANTÍA 100% CORRECTAS**

**FEBRERO 98**

➔ **1) ADALINE EMPLEADO PARA PREDICCIÓN DE SEÑALES (ADAPTH)**

**A) ¿Qué ocurre si la velocidad de aprendizaje es extremadamente baja?**

La red nunca llega a adaptarse a la señal, sino que siempre va por detrás de ella, y predice de una forma muy atrasada. Una solución sería incrementar la velocidad, aunque no demasiado, porque se producirían efectos contrarios

**B) ¿En qué caso se obtiene una salida correcta si hacemos un simulin con la señal y los últimos pesos?**

**2) Las neuronas radbas y las de una estructura competitiva ¿Tienen el mismo valor de activación?**

R: (w-x) las dos

➔ **3) ¿Qué ocurre si en un mapa autoorganizativo decrece excesivamente rápido el área de influencia? ¿Cómo afecta a sus propiedades?**

Pues que se comportará excesivamente rápido como una red puramente competitiva como exponía Kohonen. Afecta sobre todo a la propiedad de colaboración entre neuronas para aproximarse al espacio de entradas, ya que de esta forma se llegará muy pronto al estado en que cada neurona gana siempre, y sólo unas cuantas neuronas serán adaptadas en sus pesos para llegar a representar el espacio de entradas.

**4) Puede resolverse el problema del viajante con un M.A.O.? Como?**

Este problema establece el camino más corto entre una serie de ciudades sin pasar dos veces por la misma.

La red que puede solucionar este problema tiene tantas neuronas de salida como ciudades del recorrido, y dos neuronas de entrada para las coordenadas en el plano (x,y) de cada ciudad. La distribución de las neuronas de salida es circular, estableciendo para cada neurona (j) una zona de vecindad que incluye las neuronas anterior (j-1) y posterior (j+1)

El proceso a seguir para resolver el problema, consiste en entrenar la red para que aprenda la situación geográfica de las N ciudades del recorrido, adaptando los pesos mediante el algoritmo de Kohonen. Cuando el entrenamiento ha concluido, al existir tantas neuronas de salida como ciudades, deberán coincidir los vectores de pesos de la red  $W_j = [w_{jx}, w_{jy}]$  con las coordenadas de las N ciudades. Entonces el problema ha sido resuelto, siendo la solución el recorrido resultante de unir los puntos (ciudades) correspondientes a los pesos de las neuronas adyacentes de la capa de salida. La solución que se obtiene es buena, aunque no puede garantizarse que sea la óptima, ya que el criterio seguido por la red al fijar la zona de vecindad indicada ha sido buscar siempre como siguiente ciudad la más próxima.

Problema: Puede suceder que más de una neurona pueda ser atraída por la misma ciudad, lo que originaría que algunas ciudades quedasen aisladas sin aparecer en el recorrido final. Solución: Aumentando el número de neuronas, para que sean superiores al número de ciudades.

## ¿Como se abordaría con una red de Hopfield el problema del viajante?

También Se puede abordar utilizando una red de Hopfield de  $N^2$  neuronas, siendo  $N$  el número de ciudades. Imaginaremos  $N$  filas y  $N$  columnas de tal forma que cada fila se asocia con una ciudad del recorrido y las columnas con la posición de cada ciudad dentro de la ruta.

La utilización del modelo de Hopfield para la resolución de problemas de optimización se basa en la idea de intentar fijar el objetivo del problema mediante una expresión matemática, denominada función de coste o función objetivo, que haya que minimizar. A continuación se compara con la expresión general de la función de energía de una Red de Hopfield, determinandose los valores de los pesos y de los umbrales en términos de parámetros de la función objetivo para que ambas expresiones sean equivalentes. De esta forma, cuando se pone en funcionamiento una red Hopfield, continua con los valores calculados, ésta itera hasta alcanzar un mínimo de la función de energía, que en este caso coincidirá con un mínimo de la función objetivo, encontrando una posible solución de mínimo coste del problema de optimización. Generalmente no es fácil llegar al mínimo global por la presencia de mínimos locales.

### → 5) BPN con momento ¿ Que reglas heurísticas de incremento de velocidad satisfacen?

A mayor tasa de aprendizaje, mayor es la modificación de los pesos en cada iteración, con lo que el aprendizaje es más rápido, aunque por otro lado puede dar a oscilaciones. Para filtrar estas oscilaciones, se añade a la expresión del incremento de pesos, un término momento beta de manera que influye el cambio en los pesos en el instante  $t$  para la modificación en el instante  $t+1$ .

$$\text{beta} * (w(t) - w(t-1))$$

De esta forma, si en  $t$  el incremento de un peso era positivo y en  $t+1$  también, entonces el descenso por la superficie de error en  $t+1$  es mayor. Sin embargo, si en  $t$  el incremento era positivo y en  $t+1$  negativo, el paso que se da en  $t+1$  es más pequeño, lo cual es adecuado, ya que eso significa que se ha pasado por un mínimo y que los pasos deben ser menores para poder alcanzarlo.

### 6) LVQ. ¿Cuál es la función de la capa oculta? ¿Cuándo tiene sentido emplear en ella un M.A.O. con la filosofía de Matlab?

**Pregunta 7.- ¿Que representan los PEs de una capa oculta en una LVQ con respecto a las salidas deseadas de los patrones de entrenamiento?**

Las salidas de la capa oculta (competitiva) en un LVQ representan la clase en la que el mapa autoorganizativo o estructura competitiva analoga cataloga la entrada actual. Después, a través de la capa de salida, será asignada a dicha clase o región un vector de salida (salida deseada del patrón de entrenamiento).

### 7) Art2 ¿Cuándo no es necesaria fase de búsqueda?

R: Cuando  $T=W$

### ¿Por qué no es tan destructiva como ART1?

Por dos razones fundamentales:

- Se trabaja con la distancia como medida de similaridad, y no con un AND lógico, por lo que no es tan destructiva.
- La actualización de pesos da más importancia a lo aprendido en el pasado mediante el parámetro Num.

**Si queremos mucha similitud entre los valores de una clase ¿Cuánto ha de valer  $\rho$ ?**

R: ha de valer 0, ya que en la ART2, la medida de similaridad es la diferencia (distancia euclídea) entre el patrón de entrada y los pesos de las conexiones entre esta neurona y las de la capa de entrada.

Es diferente a la red ART1, en la que se utilizaba como medida de similaridad un parámetro de vigilancia, que representaba un porcentaje de semejanza entre la entrada y el prototipo de la categoría encuestada (1 sería 100% de semejanza)

**Demostrar que la función de energía de una red de Hopfield decrece de un paso a otro.**

$$R: (\Delta E = \sum w * x)$$

**8) Características Neocognitrón. Esquema de conexiones entre capas.**

**9) Aprendizaje estructura radial ¿influye en el error que los patrones sean linealmente separables?**

**10) Definir: extrapolación, generalización**

**11) Regla de aprendizaje del perceptrón**

## EXAMEN

**1) Principio de mínima perturbación del Madaline**

El entrenamiento equivale a hacer una reducción del número de salidas incorrectas para cada una de las tramas de entrenamiento que se den como entrada (nótese que la salida de la red es una serie de unidades bipolares)

Este método se puede aplicar mediante la idea de que se seleccionan las neuronas más próximas a cero, ya que son las que pueden cambiar su salida binaria con el menor cambio en sus pesos.

**2) ¿Cómo se consigue implementar la función de transferencia competitiva en base a las conexiones? ¿ Como funciona?**

Se consigue mediante conexiones laterales de inhibición y mediante conexiones hacia la propia célula de amplificación. Así, cuando una neurona se activa, produce una inhibición de las demás en su capa, y una amplificación de si misma. La célula que tenga una mayor activación gana y las demás no se activan.

**3) Estructuras Competitivas**

**A) Inicialización de las estructuras competitivas**

- Pesos aleatorios y pequeños
- Región de interés grande.

**B) ¿Qué problemas suelen abordar?**

Clasificación de patrones.

**C) ¿Qué representan los EP's de la capa de salida y sus pesos?**

Los Eps en la capa de salida representan el número de neuronas asociadas a cada clase, y sus pesos representan el vector característico que las define, es decir, el centro del clúster de datos.

**D) ¿En qué consiste el proceso de olvido?**

En hacer que ciertas neuronas que ganan de una forma continuada, sean suspendidas temporalmente para que otras neuronas puedan adaptar sus pesos. Así evitaremos que ciertas neuronas se queden sin entrenar.

→ 4)

**C) ¿Qué significa el momento?**

Es un parámetro que tiene en cuenta el comportamiento de la red en pasos anteriores y contribuye en una fracción al cambio de pesos para que los matenga en una misma dirección.

→ **Pregunta 3.- ¿Por que el Momento supone una mejora en el aprendizaje de un Perceptron Multicapa?**

La introduccion del termino Momento  $[+aAW_{ji}(t-1)]$  contribuye a mejorar la velocidad de aprendizaje de la red. Supone tener en cuenta la tendencia anterior en la modificacion de los pesos, una inercia. Además, esto le hace satisfacer las 4 heurísticas (...).

→ **D) Limite de la regla Delta generalizada**

- Puede caer en mínimos locales
- No extrapola bien si la red ha sido entrenada mal o insuficientemente

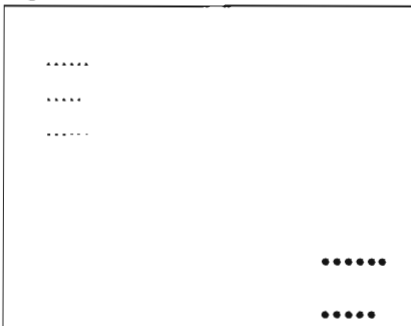
→ **E) ¿Qué se debe tener en cuenta a la hora de escoger el conjunto de entrenamiento?**

Que represente todo el conjunto de entredada.

→ **F) Capacidad de generalización. Las redes Perceptrón Multicapa no extrapolan bien.**

Tienen capacidad de generalización porque aunque no hayan sido entrenadas con un patrón determinado cerca de un conjunto de entrenamiento, aprenderá a ver las similitudes entre ellas. Sin embargo, ante entradas que se salen fuera del espacio de entradas, la red no es capaz de extrapolar

**5) En una estructura competitiva con una situación de entrada como la siguiente**



**¿Qué salida tendrá?**

Los puntos grandes representan neuronas, por tanto en una red competitiva, sólo una o dos llegarán a estar en los patrones de entrada, dejando a las demás células sin entrenar. Esto se puede mejorar entrenando con una mayor región de interés.

→ 6) **Consecuencias que podría tener que la región de interés en los mapas autoorganizativos decrezca muy rápidamente.**

Pues puede que dejemos a muchas neuronas sin entrenar, porque a medida que avanza el entrenamiento, propiciamos que unas cuantas neuronas ganen siempre, no modificándose las vecinas.

7) **Diferencias existentes entre el adaline y el perceptrón en lo que se refiere a la superficie de error.**

**¿Qué consecuencias tienes en el proceso de aprendizaje?**

Ya vista

## EXAMEN DICIEMBRE 97

1) **¿Cómo se consigue en la red ART2 que durante su funcionamiento no pierda lo aprendido?**

R: Cada vez que introduzco un patrón, influye un porcentaje del patrón en relación a los patrones anteriormente introducidos

→ 2) **Perceptrón Multicapa con 3 capas ocultas. Una vez que se ha entrenado la estructura, es posible que presentando en distintos momentos el mismo patrón, la red produzca salidas diferentes. Si es así, cuando, si no es así, comentar con que red se obtiene una situación de ese tipo.**

R: Los pesos, una vez entrenada la red, son fijos, por lo tanto no se pueden obtener salidas diferentes. Se podría obtener con una red de Melman.

3) **¿Con qué problema nos podríamos encontrar si para el entrenamiento de un perceptrón multicapa se utilizara la misma filosofía que se utiliza en la función adaphn sobre un adaline?**

R: Los pesos se van moviendo y van tendiendo hacia los últimos patrones. Corro el riesgo de viciar a la red y en el ciclo siguiente podemos minimizar el error dependiendo de únicamente los últimos patrones introducidos.

→ 4) **Contrapropagación ¿Cuándo se consigue un error 0 en el aprendizaje? ¿Es posible obtener ese error 0? ¿Por qué el error se puede minimizar mejor con una red radial?**

R: Patrones que se asocian a la misma capa competitiva tienen salidas parecidas. En una red radial la función de transferencia no es una hardlimiter, sino que la salida está entre 0 y 1, y no existe sólo un elemento de procesamiento activo. Parto de diferentes patrones de entrada y salidas más diferentes. Parto de situaciones diferentes y consigo ecuaciones diferentes.

→ 5) **BPN: El entrenamiento es como sigue?? Se determinan los términos de error de la capa de salida y se modifican las conexiones correspondientes. A continuación se determinan los términos de error de la capa oculta anterior a la salida y se realizan los cambios en los pesos asociados a dichas neuronas. Este proceso se realiza para todas las capas ocultas para llegar a la entrada. ¿es cierto que este es el proceso de backpropagation? Propagación hacia atrás? Coméntelo y Justifique la respuesta.**

R: No es cierto, ya que no se pueden modificar los pesos antes de calcular el error de las capas ocultas, porque no conseguiría el error en función de los pesos anteriores, sino en función de los pesos actualizados, lo cual no es correcto.



6) Trabajamos sobre un aprendizaje sin supervisar con una ley de Oja, la estructura tiene 2 capas (entrada y salida). La capa de entrada deja pasar los patrones tal cual. La salida de los PE's es un hardlimiter ( $W \cdot P + B$ ) Totalmente conectada. Las condiciones son:

- La matriz de pesos es la identidad
- Los bias tienen un valor fijo de ceros

Si introduzco los siguientes patrones dos veces ¿Cómo queda la matriz de pesos?

(1 0 0 0)  
 (0 0 1 0)  
 (0 1 0 1)

R: La ley de oja es  $\Delta w = l \cdot x \cdot y - dr \cdot w(t)$

Al final 
$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 2 \\ 0 & 0 & 3 & 0 \\ 0 & 2 & 0 & 3 \end{pmatrix}$$

### EXAMEN

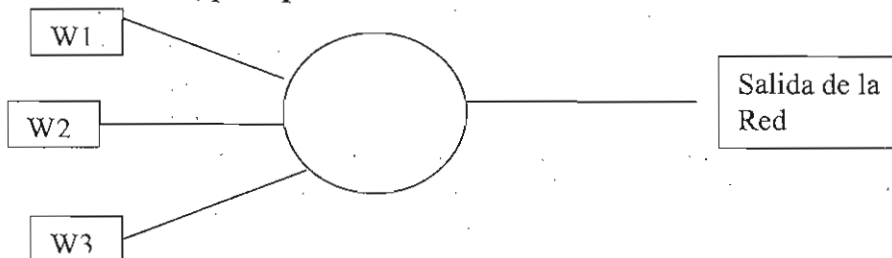
➤ 1) Cuando utilizar CPN y cuando BPN para clasificar patrones??

La CPN es un modelo de red que permite acelerar el orden de 10 a 100 veces el proceso de aprendizaje con respecto al algoritmo BPN en redes feedforward multicapa. Sin embargo, no se trata de un algoritmo óptimo, como en el caso de BPN, y por tanto sólo se deberá utilizar cuando se trabaje con pocos patrones de entrenamiento, menor o igual número que el de neuronas ocultas o con un número mayor si los patrones más parecidos generan salidas también parecidas. Las aplicaciones más comunes para CPN se refieren al procesado de imágenes, clasificación de patrones, compresión de datos, aproximación de funciones matemáticas y análisis estadístico.

2) Resolviendo un problema de clasificación de patrones. ¿Cuándo utilizaré LVQ y cuando un perceptrón multicapa?

LVQ es una red muy parecida a CPN, por lo que será un problema igual al anterior.

➤ 3) Dibuja un perceptrón. Justificar e indicar que entrenamiento es más fácil / fiable, perceptrón o adaline



(Se entiende la pregunta acerca de la grafica del error) A diferencia de la del Adaline, que es una superficie suavizada y con un unico minimo (hiperparaboloide), la grafica del error de un perceptron esta llena de minimos locales. Es por ello que su aprendizaje es mas arduo, pues se corre el permanente riesgo del estancamiento en un minimo local.

4) **¿Cuándo el error que se comete al predecir una señal es cero y no se produce modificación de pesos y utilizo lo que he obtenido para entrenar la red con un patrón de aprendizaje pasandole una señal.?**

5) **ART: Función general de los pesos feed – forward**

Son una normalización de los pesos feed – back. Los feed – back almacenan las clases.

→ 6) **¿Por qué una red de tipo competitivo es muy sensible al orden de los patrones de aprendizaje?**

Porque los primeros en entrenar, tienen más opciones de modificarse luego, ya que se adaptan al principio y no dejan a los demás entrenar. Por eso si comenzamos con otra secuencia, la situación llegaría a ser completamente diferente.

→ 7) **Casos por los que un adaline puede no llegar a converger nunca. ¿Puede suceder que una vez llegado al mínimo de la función de aprendizaje no funcione luego?**

Puede no llegar a converger por múltiples razones:

- porque tenemos el parámetro de aprendizaje muy alto y se va de un lado a otro del paraboloides, etc.
- Podemos haber llegado al mínimo, pero como la regla se basa en el error cuadrático medio, los errores serán compensados con aciertos. Por ejemplo, si tenemos muchos patrones de una clase y pocos de otra, llegaremos a entrenar bien para la mayoría, pero luego si utilizamos patrones de la segunda clase, no funcionaría bien. Aumentar el número de muestras en los patrones poco representados.
- Mal muestreo, conjunto de entrenamiento no representativo. Se soluciona aumentando el número de muestras.
- No resuelve el problema (el minimo del paraboloides no esta en el cero o el problema es no lineal).

→ 8) **¿Cuándo en una red de CPN el error es cero?**

Cuando se trabaja con pocos patrones de entrenamiento, es decir, menor o igual al número de neuronas ocultas. De esta forma, si el entrenamiento ha sido satisfactorio y no existen problemas de más de un representante para cada clase o problemas de vector pegado, tendremos un representante para cada entrada, y al terminar el entrenamiento, cada uno dispondrá de su salida deseada, con error cero.

Esto también se produciría si patrones muy parecidos generasen la misma salida, ya que ésta sería almacenada en los pesos de la última capa, y no existiría error siempre que la clasificación fuese correcta.

→ 9) **Explica porque en mapas autoorganizativos se representa una densidad de los patrones de entrenamiento.**

Porque en zonas donde existen más patrones de entrada, se ven representadas con un mayor numero de neuronas.

10) **¿Qué representan los EP's de la capa .... de una red ART?**

→ **¿Cómo haces la funcion inversa en CPN una vez tienes entrenada la funcion normal?**

La instar puede clasificar o reconocer un vector de entrada, y la outstar puede reconstruir un vector de entrada o recalcularlo a partir de un valor. Por eso si se aprende  $y=f(x)$  se tendrá también  $x=f^{-1}(y)$   
La forma de actuar sería introduciendo

DICIEMBRE 2000

1) **diferencias entre la ART 1 y la ART 2**

La red ART2 es capaz de trabajar con entradas reales en lugar de binarios.

En cuanto a su arquitectura, es igual a la ART1, solo que las conexiones feedforward y feedback son iguales  $W=T$ .

En cuanto a su funcionamiento,

La neurona vencedora en la competición es aquella que verifica una mínima diferencia (distancia euclídea) entre el patrón de entrada y los pesos de las conexiones entre esta neurona y las de la capa de entrada.

La relación de semejanza también es diferente, siendo en ART2: Entradas – pesos

La regla de modificación de pesos también es algo diferente, basandose en  $\text{Num}_j(t)$

(Numero de vectores de entrada que han sido considerados hasta el instante t como de la clase j)

2) **para q servian los pesos fw y fb en la ART**

Inicialmente, cuando la red todavía no ha aprendido nada, se asigna el valor 1 a los pesos de las conexiones hacia atrás (T) . En estos pesos, se distribuyen los M prototipos de las M diferentes clases que es capaz de diferenciar la red. Los pesos de las conexiones hacia delante (W) tienen el mismo valor que los (T), pero normalizado.

4) **explicar la metodologia para resolver el problema del viajante con un mapa autoorganizativo**

Ya visto.

→ 5) **q pasa si ponderabas la entrada de los pe de un mapa autorganizativo (solo una de las componentes de la entrada**

Se refiere a multiplicar una de las componentes de la entrada x por un numero. Finalidad que sea mas importante???



→ 6) **que son las funciones de merito y como se pueden aplicar al crecimiento dinamico de mapas**

funciones que representan el valor resource general de toda la red como sumar todos los resources de la red

→ 7) **q representa el termino de error en las capas ocultas en un perceptron multicapa**

Es una aportación del error de la neurona en la capa oculta al error general de la red. Este debe ser calculado antes de actualizar los pesos.

→ 8) **Puedes entrenar un mapa con ep bidimensionales y entradas tridimensionales??**

Si, es reducción de dimensionalidad

**Es ART sensible a lorden de patrones??**

Tanto ART1 como ART2, son sensibles a este orden, ya que la comparación se hace con los nuevos prototipos de las clases y porque la actualización de éstos siempre se realiza eliminando información no común, con lo que se pierden valores que no se pueden volver a recuperar.

→ **¿ Que función realiza la capa intermedia en la CPN?**

Es la capa de instars, que realizan la función competitiva.

\*\*\*\*\*

CONTROL DE SISTEMAS CONEXIONISTAS - DICIEMBRE 1999

**Pregunta 1.- Principio de Minima Perturbacion.**

Rige el algoritmo MR2, de entrenamiento del Madaline.  
[ver Freeman].

→ **Pregunta 2.- ¿Como se implementa la funcion de transferencia competitiva mediante conexiones y como funciona?**

Conexiones laterales: excitadoras (cada PE a si mismo)  
inhibidoras (al resto de PEs de su misma capa)

Cuando uno gana, se pone su salida a 1 e inhibe a los demas.

← **Pregunta 3.- (a) Limitaciones de las estructuras competitivas.**

- Dependen del orden de presentacion de patrones.
- Pueden quedar PE sin entrenar.

**(b) Problemas que suelen abordar. ¿Que representan los PE en la capa de salida? ¿Que representan los pesos asociados a los PE?**

Problemas de clusterizacion (categorizacion, clasificacion).

Los PE simbolizan clases y sus vectores de pesos, un vector representativo de las mismas.

**(c) ¿Que representa el proceso de olvido?**

Obliga a que un PE sea penalizado si gana muchas veces. Politica para evitar vectores de pegado.

→ **Pregunta 4.- (a) ¿Que significa el Momento? Limitaciones de la Regla Delta Generalizada.**

Significa tener en cuenta la tendencia de cambio del peso anterior.

La regla delta generalizada es un algoritmo muy lento, puede caer en minimos locales, puede resultar inestable o/y oscilar...

→ **(b) ¿Que se debe tener en cuenta al escoger el conjunto de entrenamiento?**

Que sea representativo, que contenga algun caso anomalo...

→ **(c) ¿Que es la capacidad de generalizacion? ¿Que significa que no extrapola bien?**

Capacidad de generalizacion significa que respondera bien aun ante entradas desconocidas.

Ahora bien, no lo hara si no ha sido suficientemente entrenada o mal entrenada (no extrapolara bien).

→ **Pregunta 5.- Indicar, usando estructuras competitivas, como seria el estado final, representando los PE, dado un mapa de entradas donde todas ellas estan situadas en un extremo, aglomeradas, y los pesos iniciales de los PE los situan todos juntos tambien en el otro extremo.**

[Se activarian una o a lo sumo dos neuronas que se

desplazarían al centro del cumulo de entradas, quedando todas las demas inactivas y donde estaban].

→ **Pregunta 6.- ¿Que puede pasar si la region de vecindad varia muy rapidamente?**

Se pierden las propiedades de los mapas, que surgen precisamente de esta característica.

→ **Pregunta 7.- Diferencias entre las superficies de error del Adaline y el Perceptron. Consecuencias en el aprendizaje.**

El Adaline tiene como superficie de error un paraboloides, una superficie suave con un solo minimo global. El Perceptron en cambio tiene una superficie de error llena de minimos locales. Durante el aprendizaje existe, pues, el riesgo de estancarse en uno de ellos, no así en el Adaline.

Ninguno de los dos resuelve problemas no linealmente separables.

→ **Pregunta 8.- ¿Que heurísticas cumple el Momento?**

Cumple las cuatro heurísticas (...).

---

#### CUESTIONES DE SISTEMAS CONEXIONISTAS

**Pregunta 1.- (A) ¿Como se consigue en la red ART2 que durante su entrenamiento no se pierda lo aprendido?**

La red fue diseñada para eso, almacena los prototipos aprendidos. No se produce el deterioro que se da en ART1 porque se analizan distancias.

**(B) ¿En base a que se puede afirmar que la red de Hopfield converge en un tiempo finito?**

En base a su función de energía:

$$E = (-1/2) \sum_i \sum_j (x_i * x_j * w_{ij}) + \sum_i (x_i * T_i)$$

a las propiedades que esta cumple y cierto teorema (...).

**(C) ¿Es cierto que todos los planos S de una capa del Neocognitron responden a la misma característica? Justificalo. ¿Como es el entrenamiento de los pesos en**

dicho plano?} [\*]



**(D) Trabajando con el Perceptron Multicapa con 3 capas ocultas, es posible que presentando en distintos momentos el mismo patron una vez entrenada la estructura de la red, produzca salidas diferentes. Si es cierto comentar a que es debido y sino indicar una red con la que se pueda dar tal situacion.**

No es posible que una vez entrenada una red ocurra algo así, para dar resultados diferentes debería estarse modificando algo.

**(E) ¿Con que problema nos podemos encontrar si para el entrenamiento de un Perceptron Multicapa se utiliza la misma filosofia que utiliza el adapt con el Adaline?**

El adapt presenta un patron y modifica pesos, presenta otro y los vuelve a modificar... y así sucesivamente. Esto es un perceptron multicapa, además de ser muy costoso, probablemente nos conducirá a un mínimo local, impidiéndonos salir de él con toda seguridad.

Pregunta 2.- Trabajando con contrapropagacion (CNP), ¿cuando se conseguira un error 0? {¿Por que el error que se comete durante la fase de aprendizaje se puede mimimizar utilizando una red radial?} [\*]

(ver respuesta a pregunta similar en fichero EX-SC)



**Pregunta 3.- En un Perceptron Multicapa, el entrenamiento es el siguiente:**

- Se determinan los terminos de error de la capa de salida y se modifican las conexiones correspondientes.
  - A continuacion se determinan los terminos de error de la capa oculta anterior a la salida y se realizan los cambios en los pesos asociados a las neuronas de dicha capa. Este proceso se repite para todas las capas ocultas hasta llegar a la entrada.
- ¿Es cierto que el proceso se refiere a la propagacion hacia atras del error? Justificar la respuesta.**

Este proceso es erroneo si entre cada modificacion de pesos de capa y capa se calcula un nuevo error con los nuevos pesos. Ha de propagarse la señal del todo, modificando a continuacion los pesos sucesivamente capa por capa, pero con ese valor, sin recalculer nada hasta que se hayan modificado los de todas las capas, ya que de lo contrario el error que se estaria propagando no seria el debido a la señal original.

**Pregunta 4.- ( 1 1 -1 -1 1 )**

**( 1 1 1 1 -1 )**

**(-1 -1 1 1 1 )**

**Con Hopfield se pretende almacenar la informacion de estos tres patrones. Cuando se les presenta**

**( 1 1 -1 1 -1 ), ¿que devuelve?**

**Pregunta 5.- Trabajando bajo aprendizaje sin supervisar y aplicando**

**la regla de Hebb tenemos: Una estructura de 2 capas**

**(entrada, salida) totalmente conectadas. La entrada**

**deja pasar el patron de entrada tal cual. Los PEs de**

**la capa de salida ponderan la salida mediante la**

**funcion:  $\text{hardlimiter}(w * p + b)$ .**

**Las condiciones iniciales son:**

- **Bias de cada PE de +0.5.**
- **La matriz de pesos es la identidad.**
- **El numero de PEs por capa es 4.**
- **El conjunto de entrenamiento es:**  
**(1 0 0 0), (0 0 1 0), (0 1 0 1)**

**$L_r = 0.1$  y el termino de decaimiento es:  $0.1/3$ .**

**Indicar cual es la matriz de pesos que tenemos al final.**

---

[\*] Las preguntas entre llaves son relativas a cuestiones que, si bien estan en el temario, no las hemos dado este año.

.....

## EXAMEN DE SISTEMAS CONEXIONISTAS - SEPTIEMBRE 1998

➤ **Pregunta 1.- Señalar como serian las graficas de error para un Adaline usado como filtro adaptativo en prediccion de señales en 2 casos:**

- a) **Los patrones de entrada se toman de muestras muy alejadas en el tiempo.**
- b) **Los patrones de entrada se toman de muestras muy proximas en el tiempo.**

La superficie del error es siempre nuestro famoso paraboloide. En cuanto a la grafica del error, en el caso (a) parece evidente que no se puede predecir adecuadamente una señal muestreandola de esa manera, el error sin duda seria elevado y con grandes



variaciones. En el segundo caso, empero, se augura un buen trabajo de la red, y un error mínimo.

Pregunta 2.- {Pregunta sobre el Neocognitron: función que calcula S en las capas ocultas e importancia de la localización espacial en las células c} [\*]

Pregunta 3.- {Red Radial: Estructura, motivos de su funcionalidad y manera de obtener un buen aprendizaje.}

**Pregunta 4.- Señalar los casos en los que se usaría un LVQ en vez de un perceptrón multicapa. Justificarlo.**

Suponiendo problemas de clasificación (aplicación común), el LVQ ganaría el terreno al perceptrón prácticamente en todos los aspectos: mayor velocidad en el aprendizaje, respuestas más precisas, mejor generalización (en base a distancias), sin peligro de mínimos locales...

Salvo si el espacio de entradas es poco homogéneo y prácticamente se requiere un PE por clase, solo entonces sería mejor Perceptrón Multicapa.

**Pregunta 5.- ART2. Fases con las señales de control y parámetros. ¿Cómo verifica el principio de Estabilidad y Plasticidad?**

Va almacenando los prototipos que aprende y no se degrada, como en el caso del ART1, porque aquí no se hace AND, se calculan distancias.

Por lo demás, es igual que el ART1 (...).

 **Pregunta 6.- Dado un perceptrón multicapa:**

**a) Señalar como es la secuencia realizada para la modificación de un peso  $w$  oculto.**

**b) Indicar que parámetros individuales del PE afectan a la modificación del peso.**

Dada una entrada, se propaga y se obtiene la salida. Como se tiene la salida deseada, se calcula el error a la salida (suma de los cuadrados de los errores para cada nodo de salida). Se puede, pues, aplicar la regla Delta. Las capas ocultas reciben una señal de error de las posteriores, una señal proporcional a la cantidad en que pueden contribuir al mismo (ponderada por los pesos) y así sucesivamente.

En cuanto a los parámetros individuales del PE que afectan a la modificación del peso son: el peso anterior, la ganancia, el momento:

$$W_{ji}(t+1) = W_{ji}(t) + uA W_{ji}(t) + aA w_{ji}(t-1)$$

➔ **Pregunta 7.- Señalar las diferencias existentes entre las superficies de error del Adaline y el Perceptron.**

Adaline -> suavizada, minimo unico  
Perceptron -> minimos locales

➔ **Pregunta 8.- Indicar con un dibujo como funciona el mapa autoorganizativo para las distintas estadísticas de distribución de entrada.**

[Se trata de dibujar todos los patrones juntos y mostrar que el mapa autoorganizativo quedaria con todas las neuronas juntas en el mismo lugar, dibujar dos nucleos de patrones, y mostrar que el mapa quedaria repartido entre los dos nucleos con neuronas en el medio, de "enlace" y dibujar todos los patrones distribuidos uniformemente y que el mapa autoorganizativo quedaria completamente extendido por todo el espacio de entradas].

➔ **Pregunta 9.- ¿A que tienden los pesos en la capa oculta y en la capa de salida en contrapropagacion?**

Capa oculta -> a la media de todas las entradas que se asocian a una clase en concreto. Como en cualquier estructura competitiva.

Capa de salida: 2 opciones

(i) todas las entradas de 1 clase tienen la misma salida -> vector de salida son los pesos de la capa de salida.

(ii) distintas -> media de los vectores representativos de las clases que tienen una misma salida

➔ **Pregunta 10.- ¿Como es la superficie de error de un Adaline al que le metemos como entradas (2,1) (4,2) y salidas deseadas (1,2)?**

Como podemos ver, son patrones linealmente dependientes, de modo que hay infinitas soluciones, situadas sobre una recta. La superficie de error tiene la forma de un plano doblado en forma de U.

**Pregunta 11.- Madaline: ¿que principio usa? ¿en que se basa?**

Usa el principio llamado MRH y se basa en el principio

de minima perturbacion [ver Freeman].

---

[\*] Las preguntas entre llaves son relativas a cuestiones que, si bien estan en el temario, no las hemos dado este año.

.....

## EXAMEN DE SISTEMAS CONEXIONISTAS - DICIEMBRE 1998

### **Pregunta 1.- En una red ART indica cual es la funcion general de los pesos.**

Como en cualquier red, se puede decir que los pesos representan lo que la red ha aprendido (memoria asociativa, conocimiento distribuido). Concretando, sabemos que en la ART hay dos tipos de pesos, los de las conexiones FEEDFORWARD (que solemos llamar  $w$ ) y los de las conexiones FEEDBACK (que suelen denominarse  $t$ ).

Los pesos  $t$  representan los vectores representativos de las clases que ha aprendido la red. Los pesos  $w$  tienen su mismo valor pero normalizado y simplemente trasladan la entrada a la capa competitiva.

En general, los pesos marcan siempre el orden de seleccion de las clases.



### **Pregunta 2.- ¿Por que una red competitiva es sensible al orden de presentacion de los patrones?**

El segundo de los principios relativos a la autoorganizacion nos dice que a raiz de la competitividad se generan fuertes interconexiones a expensas de otras.

Esto puede traducirse de la siguiente forma: aplicado un patron de entrada, se va a determinar un ganador, que es la neurona que va a modificar sus pesos para aproximarse a ese vector de entrada. Si el siguiente patron es muy alejado, se activaran otros PE, si es proximo, es probable que se active el mismo.

Asi, un mismo conjunto de entrenamiento puede llevarnos a situaciones muy dispares, algo que no ocurriria, por ejemplo, con mapas autoorganizativos.

### **Pregunta 5.- Explicar cuando en una red de contrapropagacion (CNP) el error es cero.**

Cuando a todos los vectores pertenecientes a una misma clase se les asigna como salida deseada una y solo una, de modo que el vector de pesos que une la competitiva y la de salida es el vector de salida deseado.



**Pregunta 6.- Explica por que los mapas autoorganizativos representan bien la densidad de aparicion de los patrones de entrenamiento.**

Debido a las relaciones de vecindad que caracterizan este modelo de red, un lugar del espacio de entradas especialmente prolífico en vectores/patrones para la red activara en mayor numero de ocasiones un mayor numero de neuronas, que seran "atraidas" a esa zona, aproximados sus vectores de pesos, arrastrando consigo a sus vecinas.

**Pregunta 9.- ¿Cuando el error que se comete al predecir una señal es cero?**

Cuando despues de presentar el mismo conjunto de entrenamiento a la red en sucesivas ocasiones. En ese caso no se produce cambio de pesos ya que se utilizan los ultimos obtenidos despues de un largo periodo de entrenamiento.