APLICACIÓN DE LAS REDES DE NEURONAS DE COMPRESIÓN A LA EXTRACCIÓN DE CARACTERÍSTICAS PARA EL RECONOCIMIENTO A PARTIR DE IMÁGENES DE LA OREJA

AUTOR: MIGUEL Á. CARREIRA PERPIÑÁN

TUTOR: ÁNGEL SÁNCHEZ CALLE

FACULTAD DE INFORMÁTICA

UNIVERSIDAD POLITÉCNICA DE MADRID

SEPTIEMBRE DE 1995

Índice

Introducción al procesamiento facial

Problemas básicos

- *Representación* de las caras para su almacenamiento y transmisión
  - Por rasgos geométricos
  - Basada en la imagen de tonos de gris
- *Detección* de la cara en una imagen. Si ésta tan sólo contiene la cara, el problema se llama de *reconocimiento*
- *Identificación*: asignar un nombre (posiblemente extraído de una base de datos) a la cara detectada
- *Análisis de expresiones faciales*
- *Clasificación* en cuanto a sexo, edad, raza, etc.

Aplicaciones

- Principalmente de índole legal: sistemas de seguridad, identificación de delincuentes, vigilancia mediante cámara de vídeo, etc.
- Otras: interfaces de usuario personalizados, videotelefonía, acceso a bases de datos usando una imagen facial como clave, etc.

Introducción al procesamiento facial (cont.)

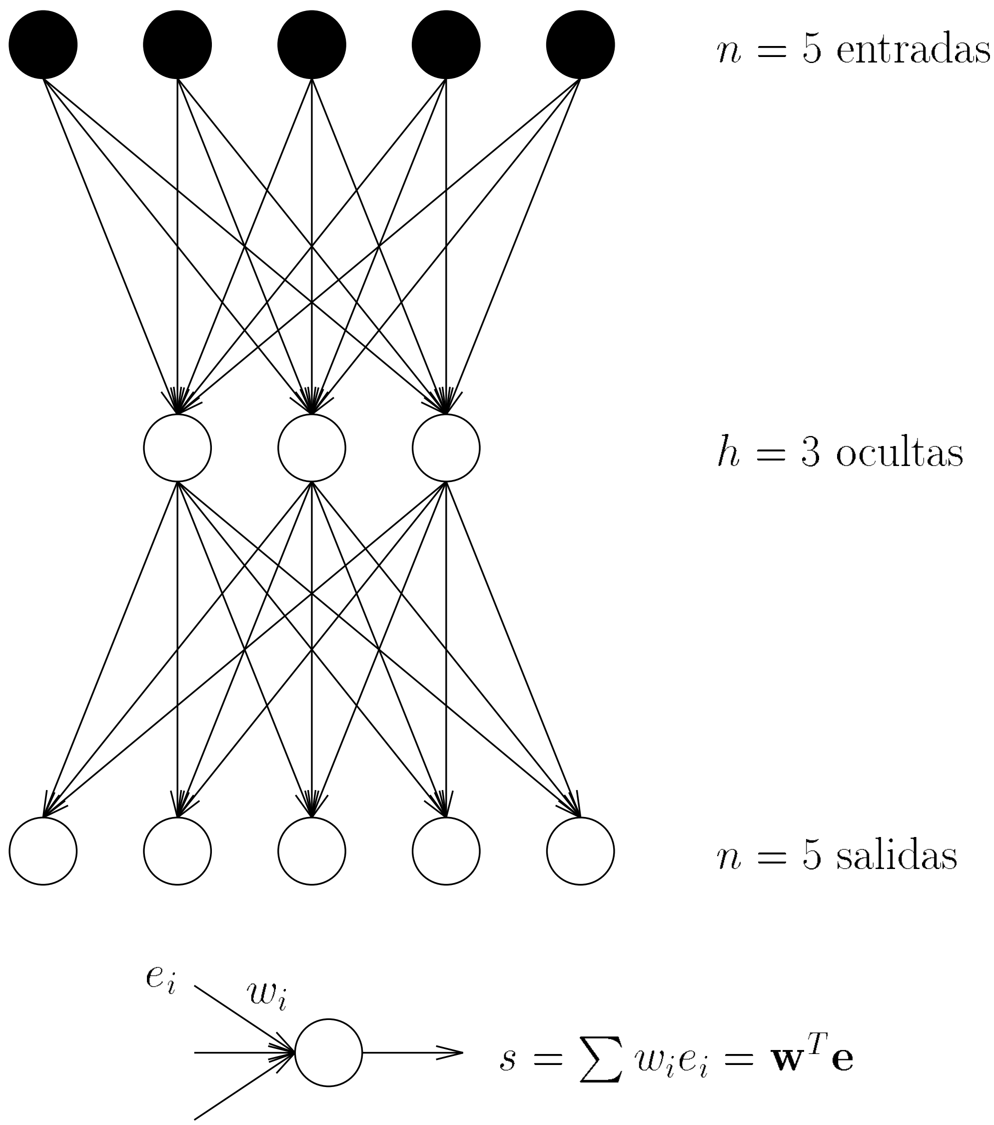
Uso de imágenes de la oreja para la identificación

- La oreja posee más información discriminante que cualquier parte del cuerpo (excepto los dibujos papilares)
- No cambia significativamente a partir de la edad adulta
- Forma y aspecto fijos (la cara varía con la expresión)
- Distribución uniforme del color
- Superficie pequeña ($\approx \frac{1}{20}$ de la de la cara)
- Desventaja: ocultación por pelo, pendientes, etc. (pero también ocurre con la cara: gafas, barba...)

Las redes
de neuronas artificiales (RNAs)

El perceptrón multicapa autoasociativo lineal

- Proceso paralelo hacia adelante del cálculo del patrón de salida: $\mathbf{x}_{\text{salida}} = \mathbf{A}\mathbf{B}\mathbf{x}_{\text{entrada}}$



Las redes de neuronas artificiales (cont.)

- Proceso paralelo hacia atrás de *aprendizaje por retropropagación*:
  - Aprendizaje autosupervisado: el patrón de salida debe ser lo más parecido al de entrada
  - Corrección iterativa de pesos: los errores en la capa de salida se propagan hacia las capas previas y se modifican los pesos respecto a la regla de aprendizaje
  - Criterio de aprendizaje: minimizar la suma de errores cuadráticos para todos los patrones de entrenamiento:

$$E = \left\langle \|\mathbf{x} - \mathbf{A}\mathbf{B}\mathbf{x}\|^2 \right\rangle$$

  - Equivale a un descenso de gradiente sobre la superficie E. Problemas:
    * Mínimos locales
    * Puntos de silla
  - El aprendizaje se detiene cuando E es suficientemente pequeño
  - Otro algoritmo de aprendizaje, *quickprop*, aproxima localmente E por un paraboloide y salta directamente al mínimo de éste. El criterio de aprendizaje sigue siendo el mismo

Las redes de neuronas artificiales (cont.)

- Conjuntos de patrones
  - *Conjunto de entrenamiento, TS*:
    * Sobre él aprende la red, es decir, sobre él se produce la minimización
    * Consta de 85 imágenes de $30 \times 48 = 1440$ píxeles y 256 tonos de gris, correspondientes a 17 sujetos distintos
    * Las imágenes fueron tomadas por el autor en condiciones muy controladas de iluminación y posición, para obtener un conjunto homogéneo. Posteriormente fueron procesadas (mejora del brillo, recorte, orientación normalizada)
  - *Conjunto de validación del entrenamiento, VTS*:
    * Debe ser parecido al TS
    * No se usa durante el entrenamiento sino al final del mismo
    * El error por patrón de la RNA (una vez entrenada) sobre el VTS debería ser parecido al obtenido sobre el TS
  - Otros conjuntos de patrones empleados:
    * *Conjunto adicional de validación, VTS2*
    * *Conjunto de patrones alterados, AS*
    * *Conjunto para evaluar la capacidad de reconocimiento, RS*

Figura 1: Patrones obtenidos en el proceso de captación. Para cada uno de los 17 individuos se tienen 6 fotos, 5 para el TS y 1 para el $VTS1$.

Figura 2: Las 17 imágenes que conforman el conjunto VTS2. Son imágenes sobrantes del proceso de captación y otras digitalizadas de revistas.

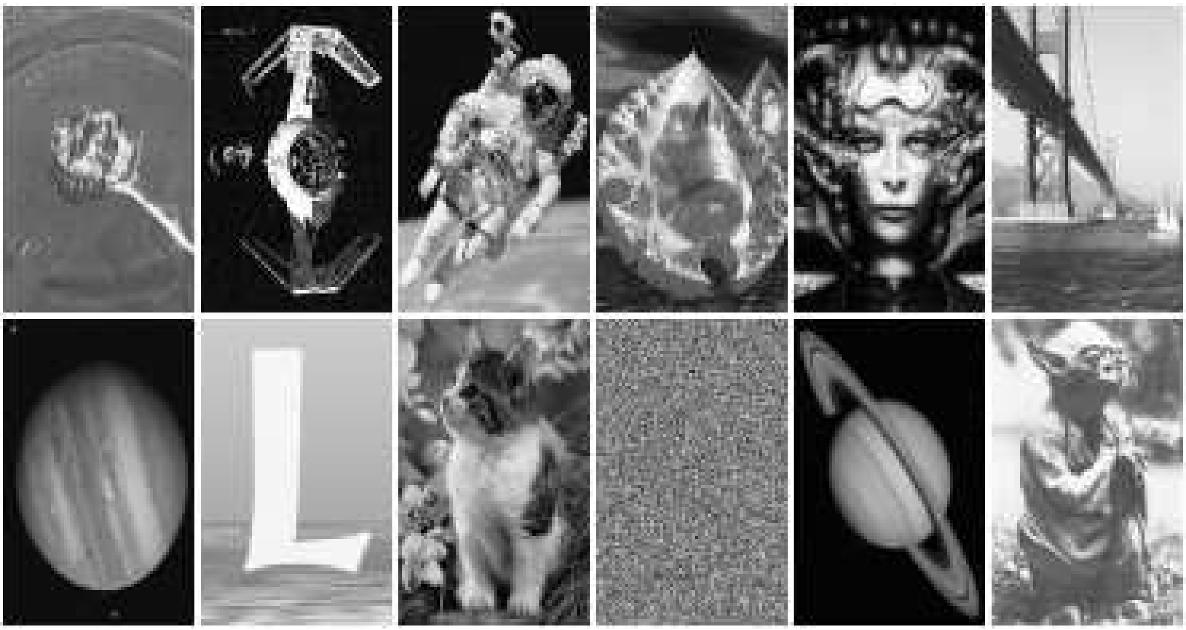


Figura 3: Fotos de diversos objetos que la red debería rechazar. Entre ellos se encuentra una imagen de ruido blanco, generada mediante una distribución uniforme $[0,1]$.

Figura 4: Las 49 imágenes que forman el conjunto AS. Son imágenes transformadas por (ordenadas por filas): adición de intensidad, multiplicación de intensidad, adición de ruido, rotación, homotecia y traslación. El patrón original separado a la derecha.

Relación con el análisis de componentes principales

El análisis de componentes principales (ACP)

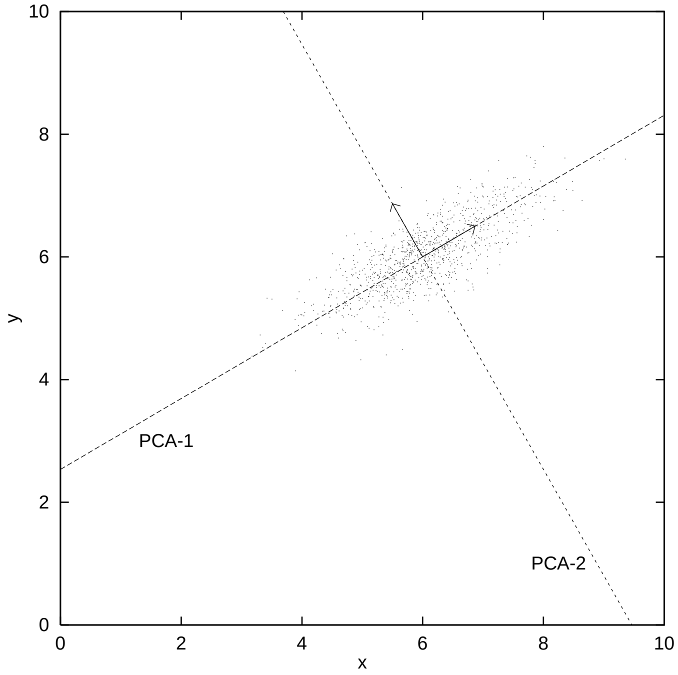


Figura 5: Nube de puntos normal en dos dimensiones con sus direcciones principales.

- Sea un conjunto de p vectores n-dimensionales $\{\mathbf{y}_i\}_{i=1,\ldots,p}$
- Obtiene un sistema de referencia ortonormal $\mathbf{u}_1, \ldots, \mathbf{u}_n$ tal que la varianza direccional a lo largo de cada nuevo eje $\mathbf{u}_i$ es máxima

Relación con el análisis de componentes principales (cont.)

- $\mathbf{u}_1, \ldots, \mathbf{u}_n$ coinciden con los autovectores (ordenados en orden descendente de su autovalor asociado λ_i) de la matriz de covarianzas de los vectores $\mathbf{y}_i$, $\mathbf{\Sigma} = \frac{1}{p}\mathbf{Y}\mathbf{Y}^T - \overline{\mathbf{y}}\overline{\mathbf{y}}^T = (\sigma_{ij})$, donde $\sigma_{ij} = \frac{1}{p}\left\langle(\mathbf{y}_i - \overline{\mathbf{y}}_i)(\mathbf{y}_j - \overline{\mathbf{y}}_j)\right\rangle$
- Los autovalores $\lambda_1 \geq \cdots \geq \lambda_n$ coinciden con las varianzas direccionales: $\lambda_i = \mathrm{var}_{\mathbf{u}_i}\{\mathbf{y}\}$
- Es más cómodo usar directamente la matriz $\mathbf{X}\mathbf{X}^T$
- Tomando sólo los h primeros componentes principales (CPs) $\mathbf{u}_1, \ldots, \mathbf{u}_h$ tenemos una representación lineal en h coordenadas lo mejor posible (en el sentido de mínimos cuadrados)
- Es decir, comprimimos la información en una fracción $\frac{h}{n}$

Otros enfoques

- El ACP es equivalente a la transformada de Karhunen-Loève
- Desde el punto de vista de la teoría de la información, el ACP maximiza la información mutua entre los vectores originales y los comprimidos (minimiza su entropía condicional)

Relación con el análisis de componentes principales (cont.)

Características de la superficie de error
$E(\mathbf{A},\mathbf{B}) = \left\langle \|\mathbf{x}-\mathbf{ABx}\|^2 \right\rangle$

- Sea una red de compresión con n unidades de entrada, h ocultas y p patrones de entrenamiento
- $E(\mathbf{A},\mathbf{B})$ presenta un único mínimo local y global en el punto $\mathbf{W} = \mathbf{AB} = \mathbf{U}_h\mathbf{U}_h^T$, con $\mathbf{U}_h = (\mathbf{u}_1,\ldots,\mathbf{u}_h)$
- En ese punto $E = \operatorname{tr}\mathbf{XX}^T - \sum_{i=1}^h \lambda_i = \sum_{i=1}^n \lambda_i - \sum_{i=1}^h \lambda_i$
- $E(\mathbf{A},\mathbf{B})$ presenta puntos de silla para cualquier otro conjunto de autovectores en $\mathbf{U}_h$ (distinto del formado por los h principales)
- Observemos que:
  - $\mathbf{A}$ y $\mathbf{B}$ no son únicas ($\mathbf{A} = \mathbf{U}_h\mathbf{C}$ y $\mathbf{B} = \mathbf{C}^{-1}\mathbf{U}_h^T$ hacen $\mathbf{W} = \mathbf{U}_h\mathbf{U}_h^T$ para $\mathbf{C}$ invertible)
  - Por tanto, el aprendizaje de la red de compresión obtiene una base $\mathbf{A}$ (o $\mathbf{B}$) del subespacio generado por los CPs $\mathbf{u}_1,\ldots,\mathbf{u}_h$, pero no coincidente —en general— con esta última
  - En la capa oculta queda el vector $\mathbf{c} = \mathbf{By}$, de h componentes, que es la proyección sobre el subespacio de los CPs del vector de entrada $\mathbf{y}$. $\mathbf{c}$ es, pues, el vector de características buscado
  - Existen otros tipos de RNAs capaces de extraer los autovectores $\mathbf{u}_1,\ldots,\mathbf{u}_h$, basados en la regla de Oja: redes de Sanger, Földiák, APEX, etc.

Experimentos realizados

Extracción de características

- La distribución de autovalores cae rápidamente con el orden de los mismos, lo que justifica la compresión
- La distribución de E sigue a la de autovalores (la integra)
- *Holones*: representaciones pictóricas de los autovectores y de los vectores obtenidos por la red
- $\mathbf{A}$ aproxima el subespacio principal en las primeras iteraciones
- Análogamente, E alcanza su valor mínimo en las primeras iteraciones
- La varianza direccional se distribuye de manera aproximadamente uniforme entre los vectores de la base obtenida (a diferencia de lo que ocurre con $\mathbf{u}_1, \ldots, \mathbf{u}_h$)

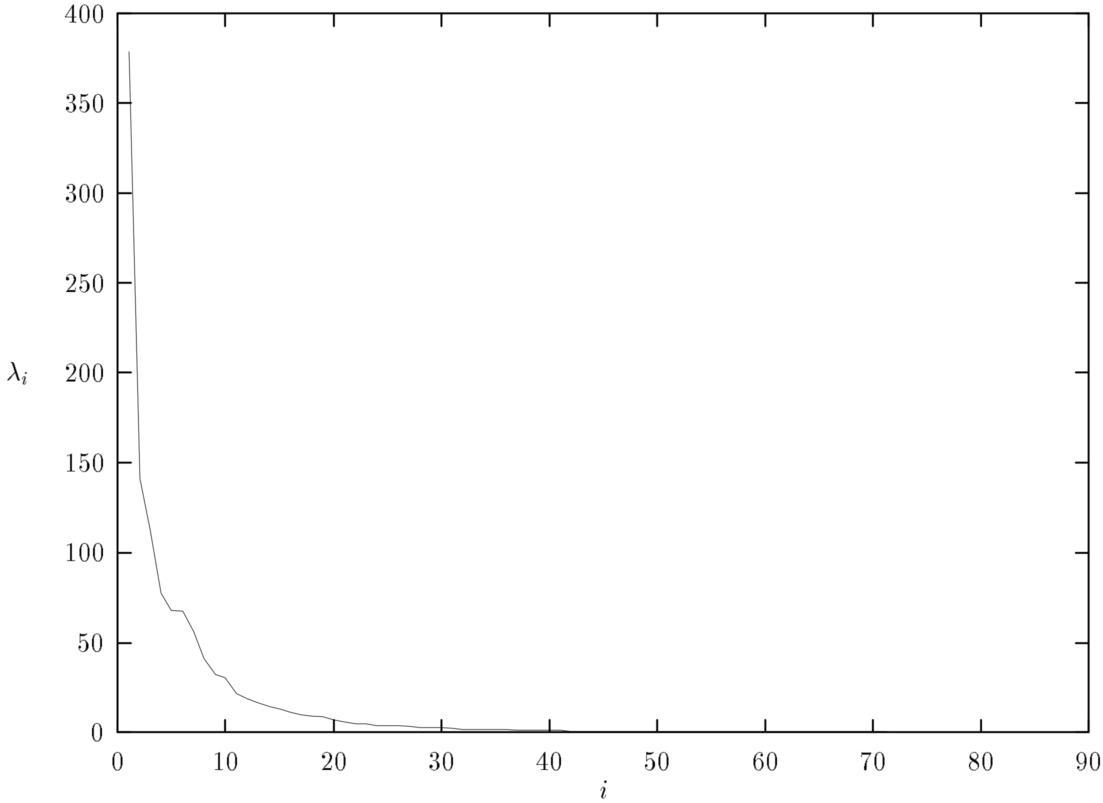


Figura 6: Autovalores de $\mathbf{XX}^T$

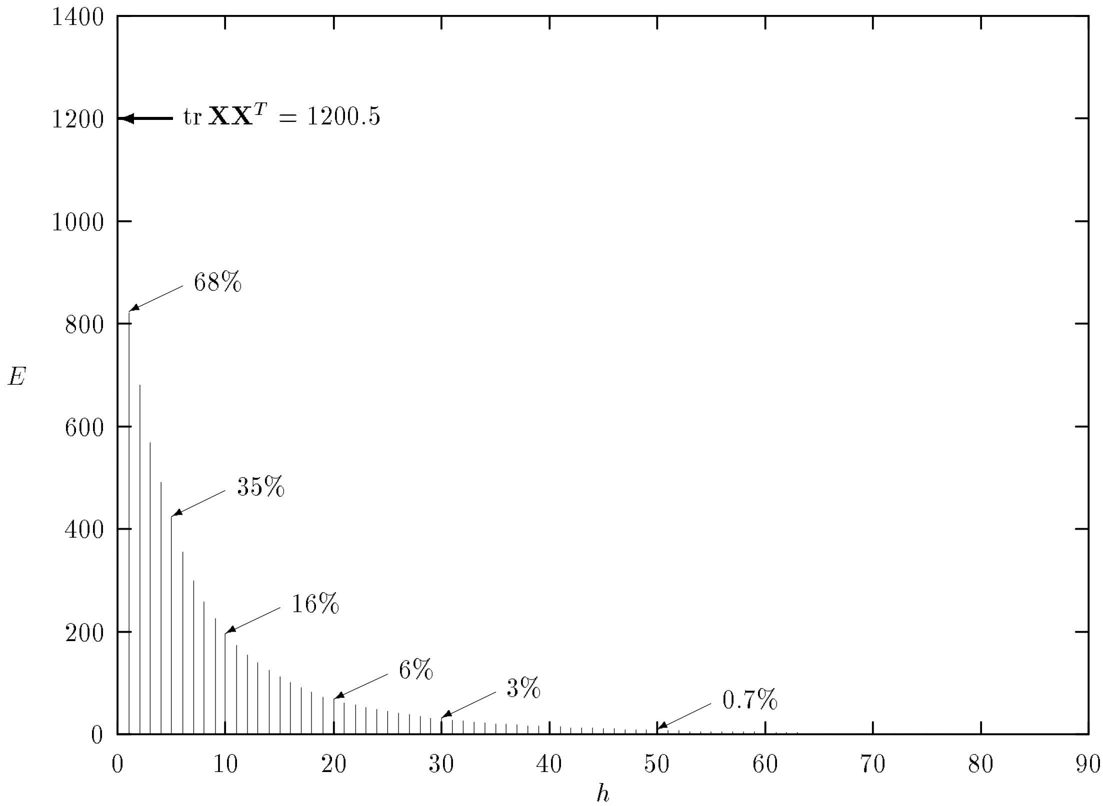


Figura 7: Errores cuadráticos E

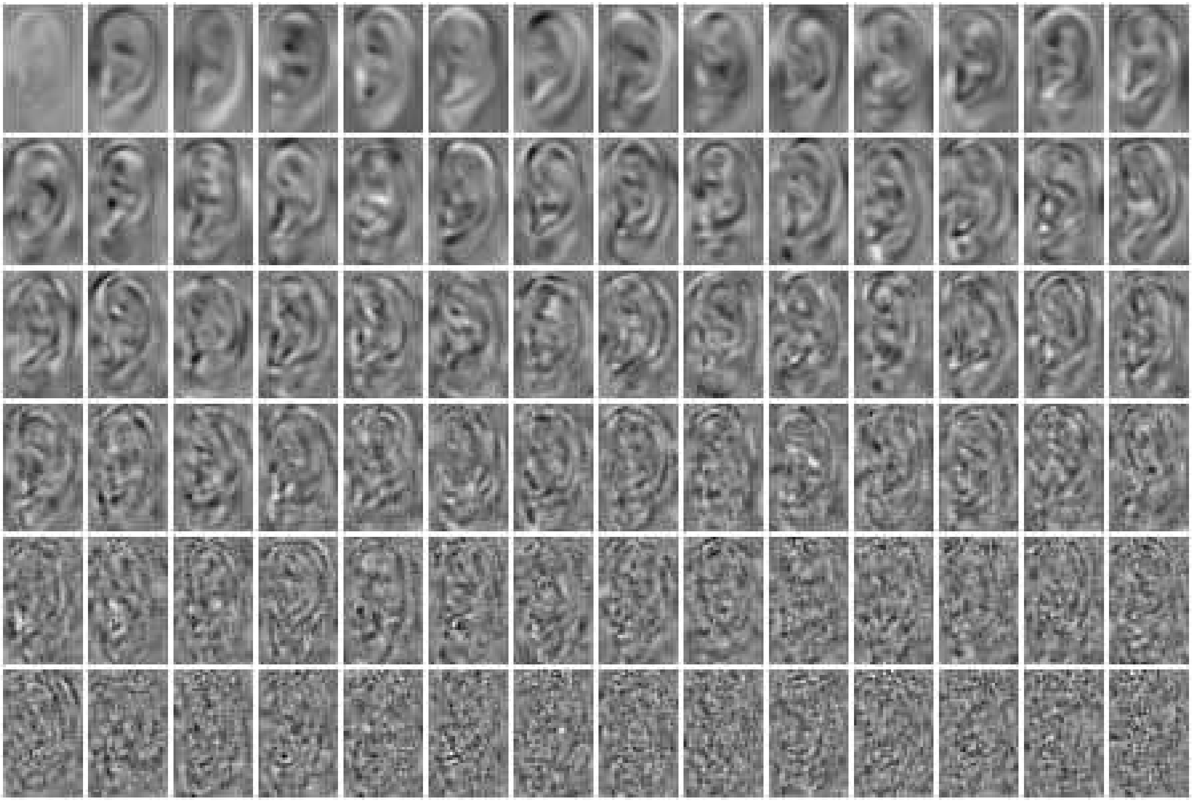


Figura 8: *Holones* de los autovectores principales $\mathbf{u}_i$

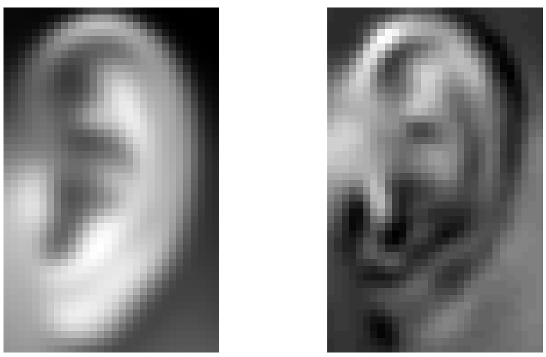


Figura 9: Media $\overline{\mathbf{y}}$ (a la izquierda) y autovector principal $\mathbf{u}_1$ (a la derecha, normalizado a 256 tonos de gris)

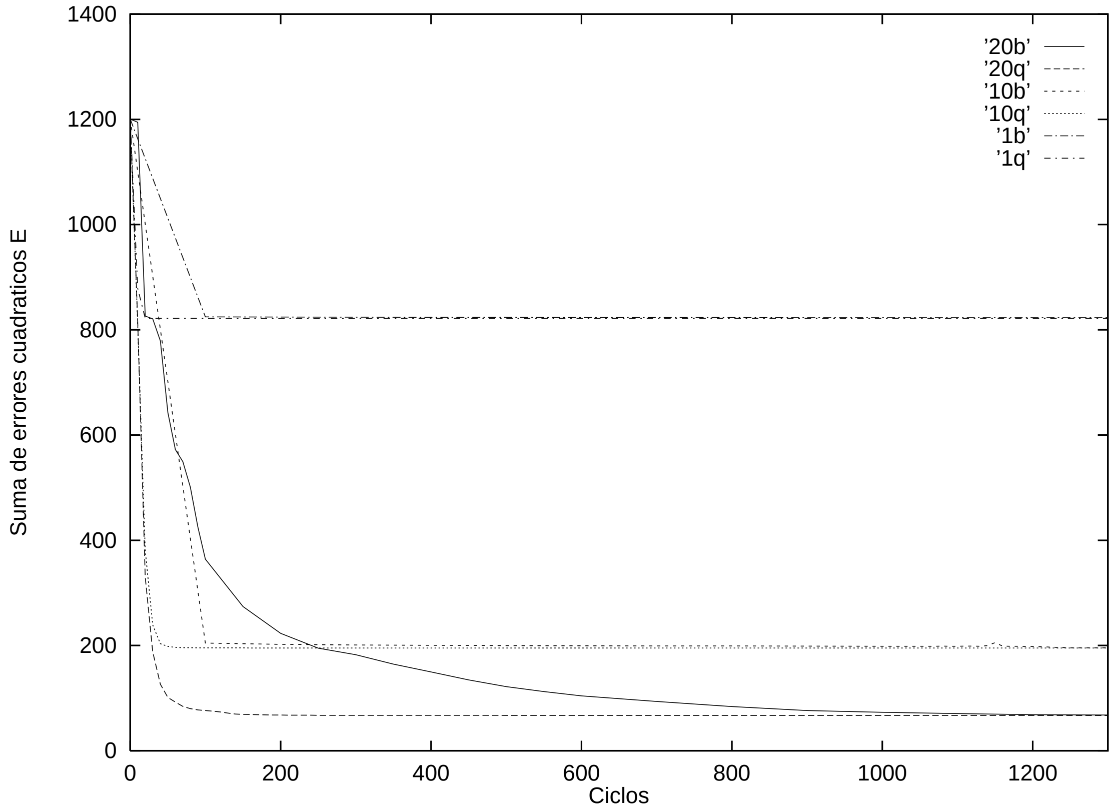


Figura 10: Curvas de aprendizaje para $h = 1$, 10 y 20, con los algoritmos de retropropagación y *quickprop*.

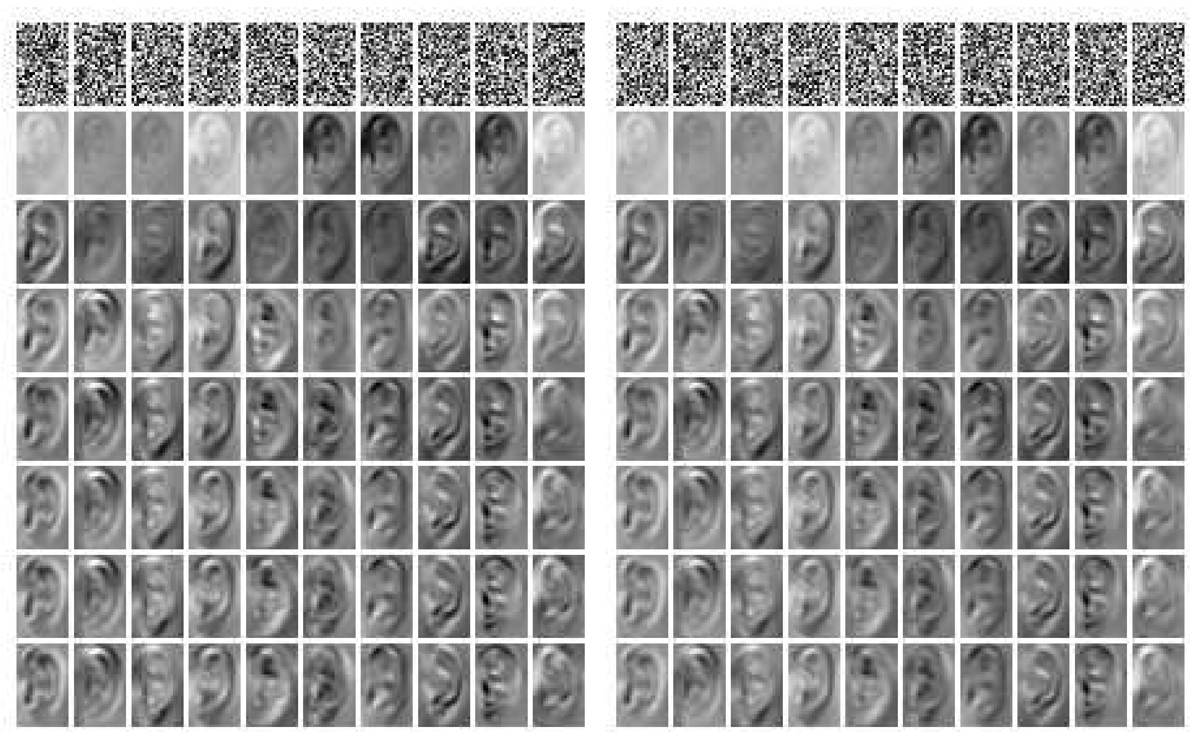


Figura 11: $h = 10$: evolución de los *holones* durante el aprendizaje con retropropagación. El bloque izquierdo corresponde a $\mathbf{B}$ y el derecho a $\mathbf{A}$. La primera fila a las matrices iniciales y la última a las finales.

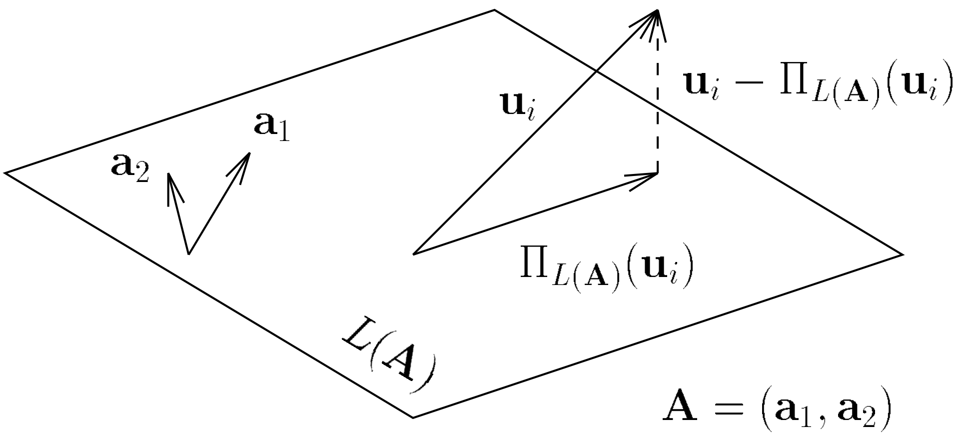


Figura 12: Proyección sobre el subespacio $L(\mathbf{A})$ y error generado.

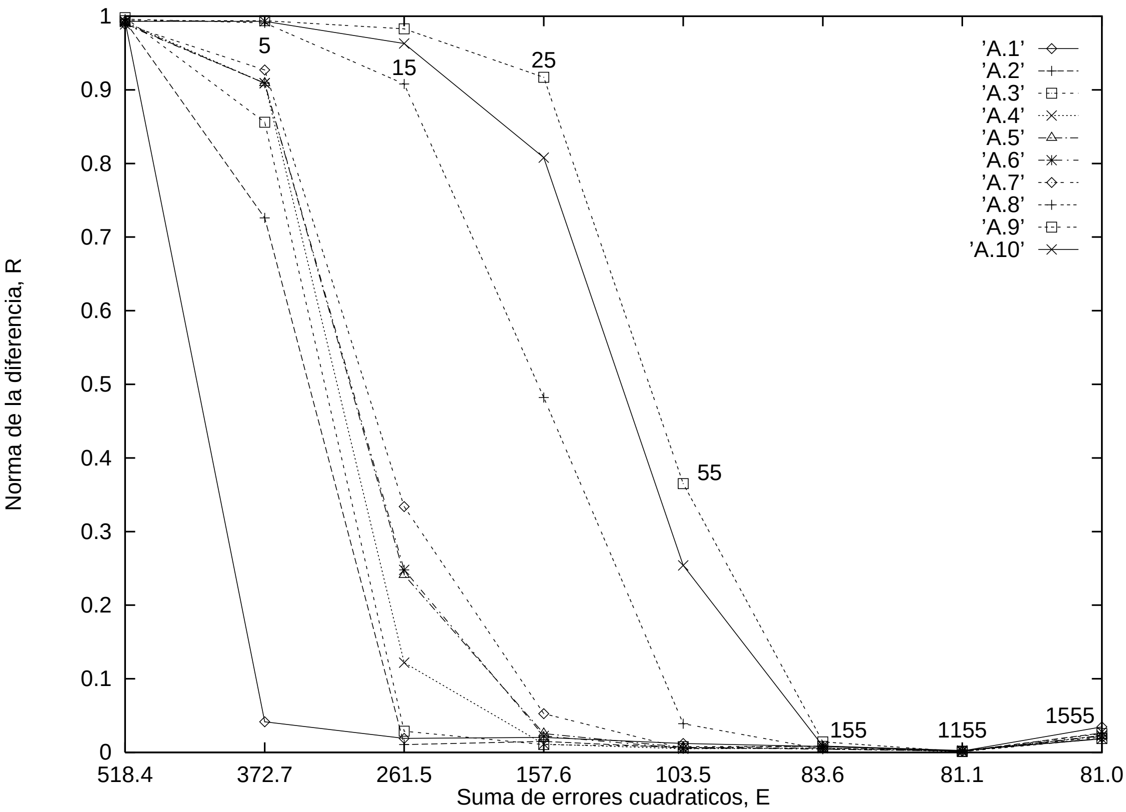


Figura 13: $h = 10$: evolución de $\|\mathbf{u}_i - \Pi_{L(\mathbf{A})}\mathbf{u}_i\|$, $i = 1, \ldots, 10$ durante el aprendizaje.

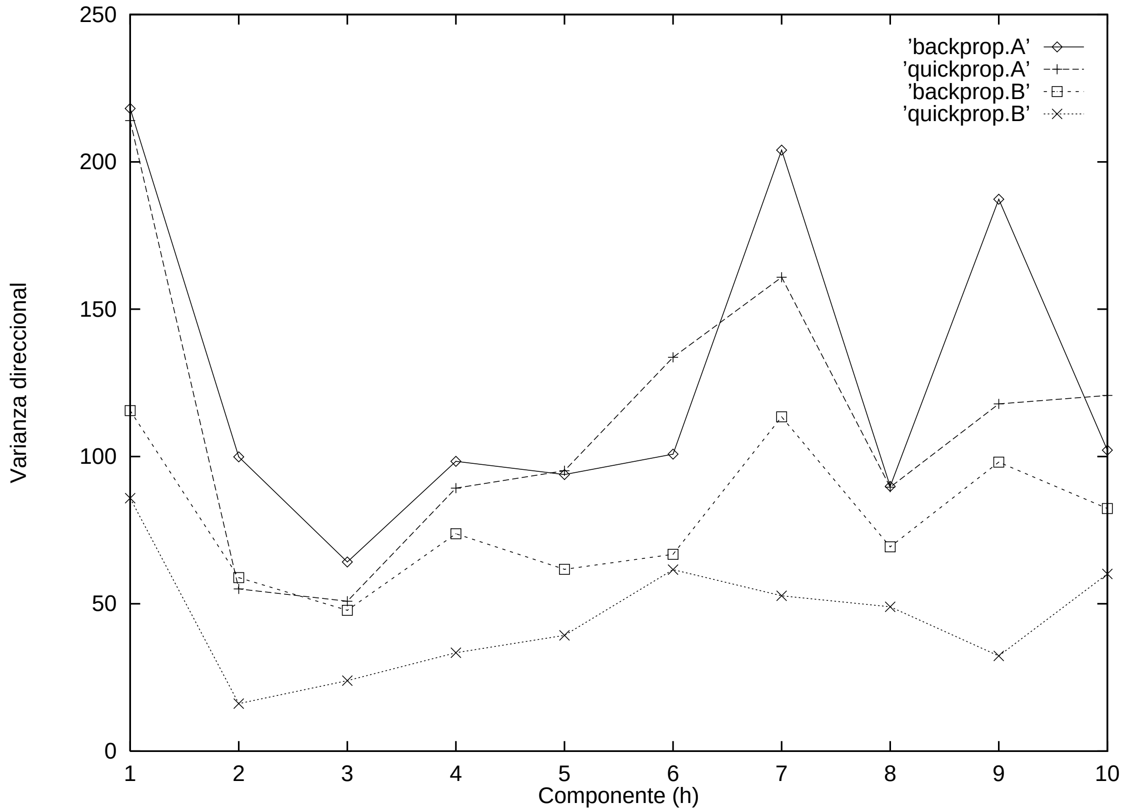


Figura 14: $h = 10$: distribución de p-varianzas en las bases finales, para las dos matrices **A** y **B** y los dos algoritmos de aprendizaje, retropropagación y *quickprop*.

Experimentos realizados (cont.)

Aplicación al reconocimiento

- Regla de rechazo:

$$E_{\mathbf{y}} = \|\mathbf{y} - \mathbf{W}\mathbf{y}\|^2 = \begin{cases} > E_0 \Rightarrow \text{No es reconocido (rechazar)} \\ < E_0 \Rightarrow \text{Sí es reconocido (aceptar)} \end{cases}$$

- Interpretación geométrica: franja de grosor $2\sqrt{E_0}$ alrededor del hiperplano principal de dimensión h
- Fácil implementación, pero puede ser difícil fijar el umbral E_0
- Robustez de la red: soporta aceptablemente transformaciones en la intensidad de la imagen y la adición de ruido; falla ante transformaciones geométricas (rotación, homotecia, traslación)
- La red sirve también como memoria autoasociativa

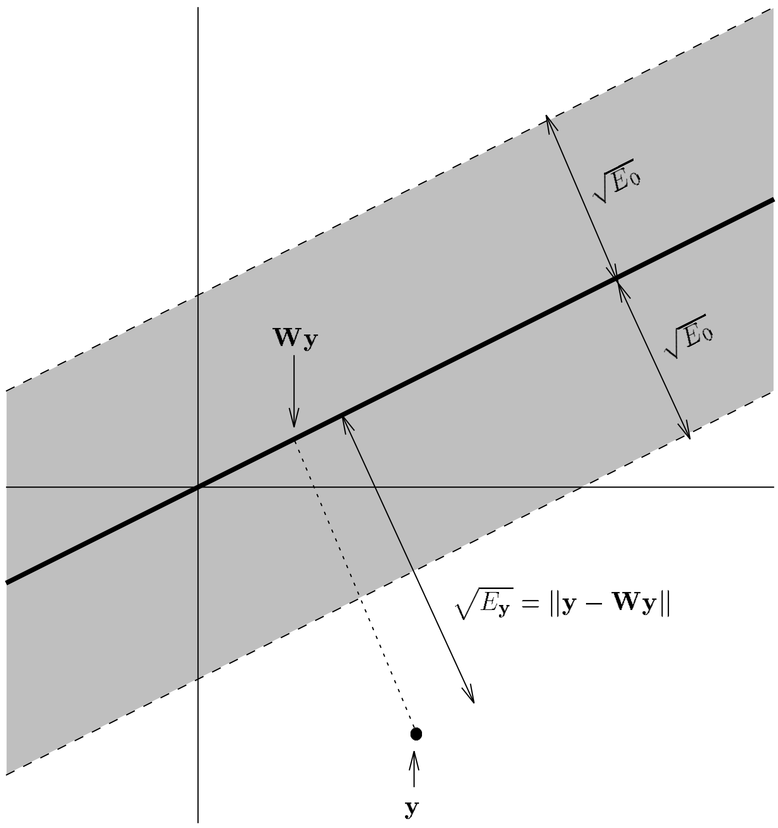


Figura 15: Región reconocida por la regla de rechazo y su complemento.

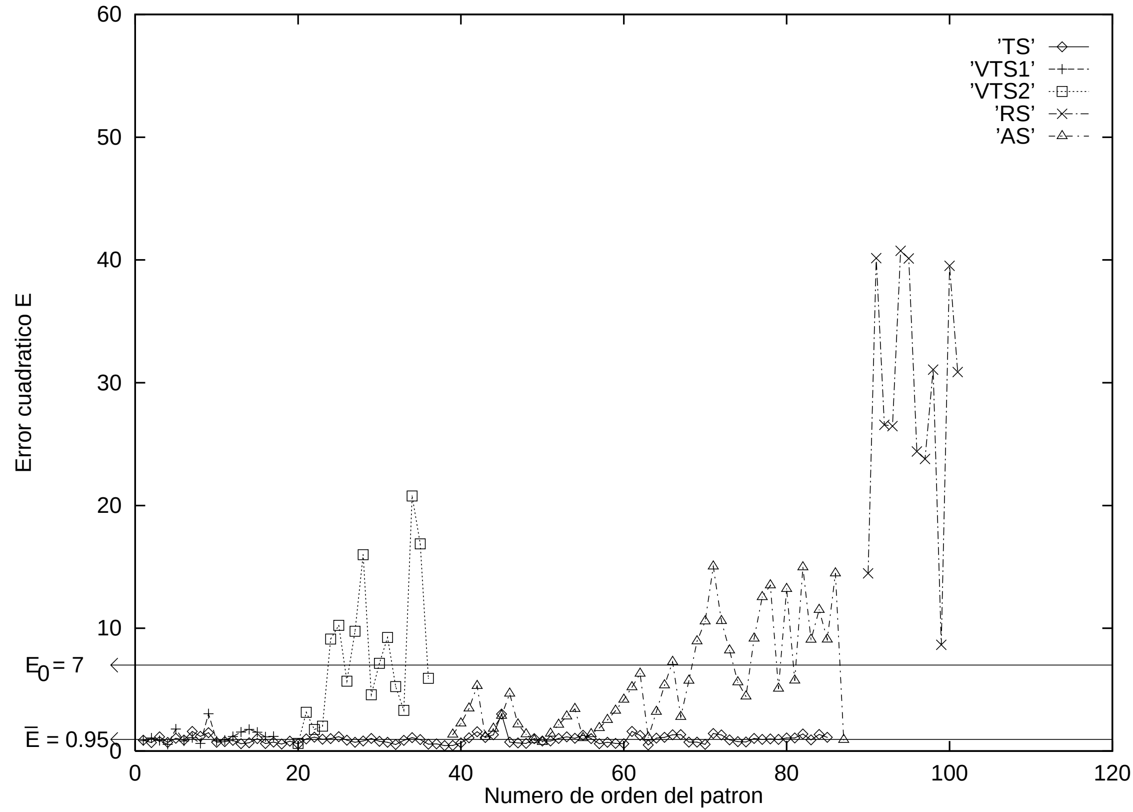


Figura 16: $h = 10$: errores para los conjuntos TS, VTS1, VTS2, AS y RS, valor medio del error para el conjunto TS ($\overline{E}$) y posible umbral de rechazo (E_0).

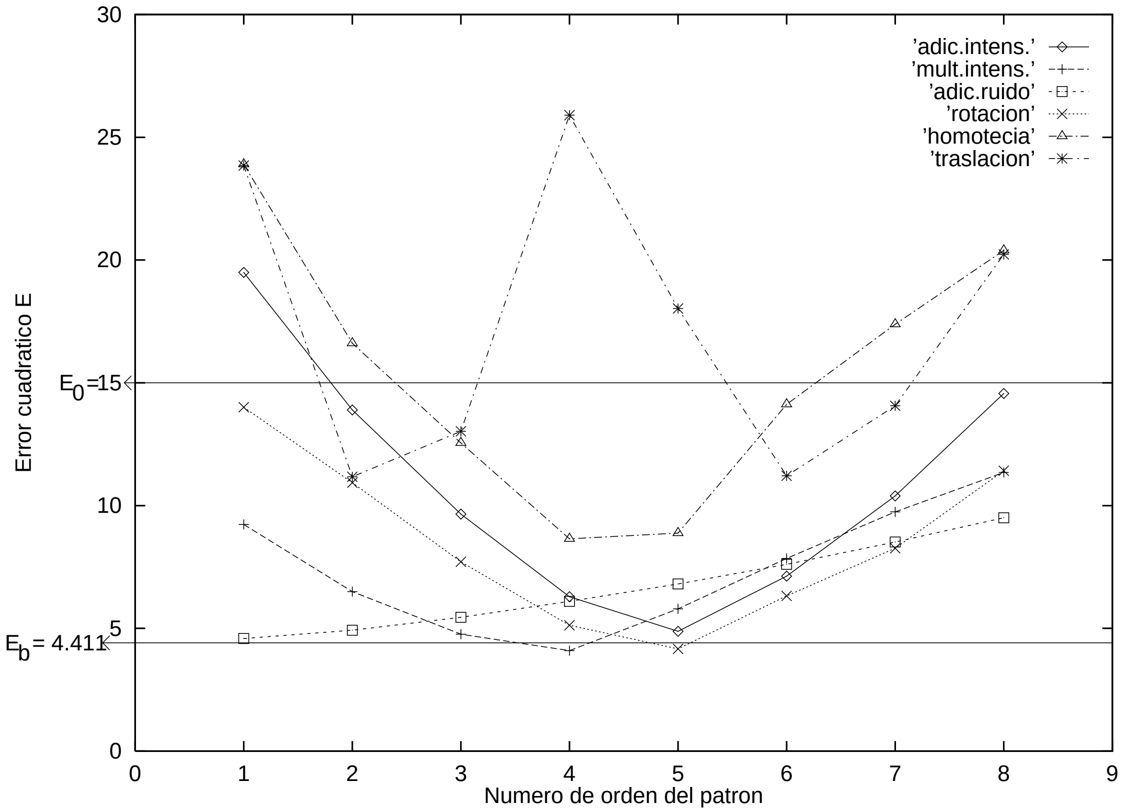


Figura 17: $h = 1$: errores para el conjunto AS, error base (E_b) y umbral de rechazo fijado anteriormente (E_0).

Figura 18: Imágenes de entrada y su reconstrucción por la red de compresión.

Conclusiones

Resultados principales

- La red de compresión realiza durante su entrenamiento un proceso muy similar al análisis de componentes principales del conjunto de patrones de aprendizaje
- Las activaciones de sus unidades ocultas sirven como vector de características para un vector (imagen) de entrada
- El error de la red al reconstruir un patrón permite usar una regla de rechazo para el reconocimiento de un patrón

Comparación con otros métodos

- El análisis espectral completo de la matriz $\mathbf{X}\mathbf{X}^T$ puede efectuarse con métodos tradicionales de análisis numérico de manera más eficiente que la red de compresión, si $\mathbf{X}\mathbf{X}^T$ es pequeña
- Cuando el orden de $\mathbf{X}\mathbf{X}^T$ es grande (p. ej. imágenes de resolución $30 \times 48 = 1440$ dan un orden de 1440×1440) la red aventaja a estos métodos, particularmente con el uso del aprendizaje *quickprop*

Otras observaciones

- El número de patrones empleados en el entrenamiento es muy pequeño

Desarrollos futuros

- Usar la red de compresión como fase de extracción de características de un sistema de identificación
- Usar un conjunto de patrones realista: mayor número y resolución
- Estudiar la distribución de autovalores de $\mathbf{XX}^T$ para diversas resoluciones y familias de imágenes
- Emplear otras arquitecturas de RNAs
- Dotar al sistema de invarianza a ciertas transformaciones
- Aplicar todo lo anterior a imágenes faciales
- Seleccionar rasgos geométricos relevantes de la oreja
- Implementar un sistema práctico (p. ej. un sistema de control de acceso)