



FUNDAMENTOS DEL RECONOCIMIENTO AUTOMÁTICO DE LA VOZ



“Métodos de modelado y clasificación de patrones”

Agustín Álvarez Marquina



Introducción. Modelado y clasificación de patrones



Objetivos:

- Agrupar el conjunto de vectores de rasgos de características en clases.
- Proporcionar una medida del grado de pertenencia de un determinado vector de características a una clase.



Métodos de modelado y clasificación de patrones



- **Cuantificación vectorial.**
- **Redes neuronales.**
- **Funciones de densidad de probabilidad.**



Cuantificación vectorial (I)



- **La cuantificación vectorial [MAK85], [GRA84] es una técnica que permite comprimir la información proporcionada por las plantillas espectrales obtenidas en la fase anterior.**
 - También, permite de paso reducir la dispersión estadística presente en los datos.
- **La principal motivación para realizar este proceso viene determinada por el modelo de producción de la voz.**





Cuantificación vectorial (II)



- Si el número de sonidos elementales que se pueden producir físicamente es limitado en una lengua, debemos poder ser capaces de encontrar la transformación que convierta nuestro conjunto de vectores de valores continuos en un nuevo conjunto con un número de vectores finito.
 - En el caso más extremo los vectores pasan a ser elementos individuales.



Cuantificación vectorial (I)



- **Método supervisado, que asocia para cada vector de rasgos de la entrada de dimensión N , una de las K clases de equivalencia posibles.**
 - El proceso de asociación requiere definir un determinado criterio de distancia: Euclídea o Mahalanobis.

- **Define para cada clase k , con $0 \leq k \leq K$ un vector de dimensión N , que representa el centro de masas dicha clase (centroide).**





Cuantificación vectorial (II)



- La unión de los K centroides constituye el libro de códigos (*codebook*).
- Las distancias entre un vector de la entrada y los vectores contenidos en los libros de códigos permite establecer el grado de pertenencia de dicho vector a cada clase.



Cuantificación vectorial (III)



- **Proceso de obtención de los libros de códigos. Método de las K-medias:**
 - ➊ Partición del conjunto inicial en K grupos o clases.
 - ➋ Cálculo de los centroides de clase iniciales.
 - ➌ Reorganización de los vectores de entrada con respecto al conjunto de centroides.
 - ➍ Volver al paso 2 hasta alcanzar un determinado umbral o número de iteraciones.



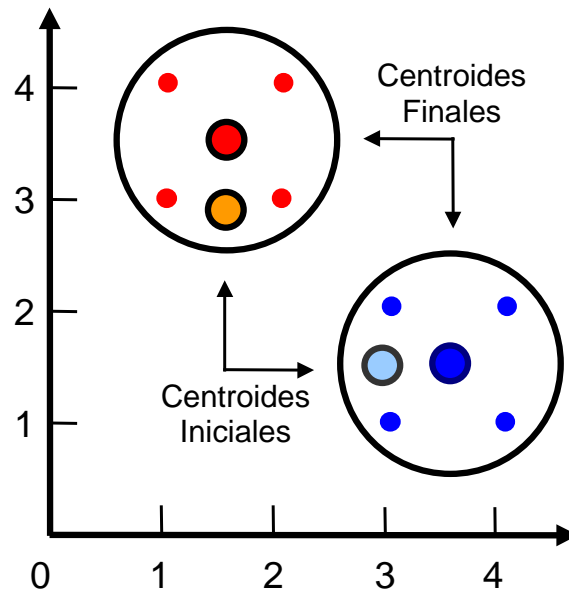


Figura . Reorganización de los centroides de clase.

- Una vez construido el libro de códigos (entrenamiento del sistema), se el grado de pertenencia de un vector al conjunto de clases.
- La pertenencia se expresará en forma de probabilidad.
- Puede definirse de diversas formas:
 - Pertenencia solamente a la clase con el centroide más cercano.
 - Pertenencia proporcional a las C clases más cercanas.



Redes neuronales (I)



- Una red neuronal [LIP87a] [KOH90] [ROB94] [TEB95] es una estructura de cómputo capaz de discriminar y modelar características no lineales.
- Está compuesta por un conjunto de unidades (usualmente grande) de procesamiento sencillas conectadas entre sí, que operan de forma simultánea.
- Aparte de las unidades de proceso o neuronas no existe otro tipo de unidad de control o procesamiento adicional.



Redes neuronales (II)



- Las neuronas de una red se dividen en unidades de entrada, unidades ocultas y unidades de salida.
 - Las células de entrada reciben los datos de entrada al sistema.
 - Las neuronas de salida proporcionan los resultados proporcionadas por la red.
 - Las neuronas de la capa oculta son las encargadas de la transformación de las diferentes representaciones internas de los datos.





Redes neuronales (III)



- El tipo de conexión entre unidades de proceso determina en buena parte el uso de la red neuronal.
- Aunque una red puede presentar cualquier tipo de topología las más usuales son:
 - **Redes desestructuradas.** Son útiles para tareas de completado de patrones.
 - **Redes organizadas en niveles.** Empleadas para asociación de patrones.



Redes neuronales (IV)



- **Redes recurrentes.** Empleadas para secuenciación de patrones a lo largo del tiempo.
- **Redes modulares.** Permiten construir sistemas complejos a partir de componentes más simples.



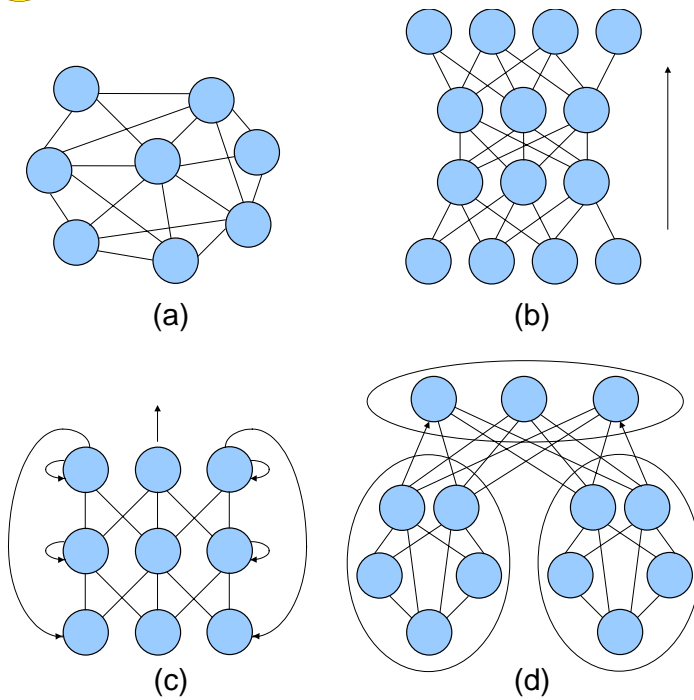


Figura 3. Diferentes topologías de redes neuronales:
(a) Desestructurada.
(b) Organizadas en niveles.
(c) Recurrente.
(d) Modular.

- El funcionamiento de la red viene determinado por el modo de computación de las unidades de proceso.
- La entrada de una célula está formada por la combinación de los valores de salida de las neuronas que la preceden, a través de la fórmula:

$$x_j = \sum_i w_{ij} y_j + \theta_j$$



Redes neuronales (VII)



- ❑ Los valores w_{ij} son los pesos de interconexión entre células de la red.
- ❑ El término θ_j representa un valor constante para la conexión j .
 - En la práctica, este término se considera como una conexión extra donde el peso es precisamente ese valor y la entrada de la conexión vale 1.
- **La célula aplica a la entrada anterior una función que puede ser de tipo lineal o no lineal.**
 - El caso lineal no se usa mucho porque proporciona muy poca potencia de cálculo.



Redes neuronales (VIII)



- **Dentro del grupo de las no lineales encontramos funciones umbral y funciones sigmoide.**
 - ❑ Un ejemplo de función umbral sencillo lo constituye la función escalón.
 - ❑ Las funciones sigmoide suelen ser las más empleadas, siendo algunos ejemplos:

$$y = \frac{1}{1 + \exp(-x)} \quad y = \tanh(x)$$





Redes neuronales (IX)



- El último factor que define a una red neuronal es el tipo de entrenamiento que se realiza.
- El proceso de entrenamiento tiene por objeto modificar el valor de los pesos de interconexión de la red reforzando aquellas conexiones entre dos unidades con salidas correladas.
 - Este mecanismo se conoce como regla de Hebb (donde además se define un factor de aprendizaje ε):

$$\Delta w_{ij} = \varepsilon y_i y_j$$



Redes neuronales (X)



- En cada caso se emplearán diferentes variantes de la anterior regla, siendo una de las más conocidas el método de retropropagación del error (*Backpropagation*).
- Dependiendo del modo en que se entrena una red neuronal se pueden clasificar en:
 - **Redes con aprendizaje supervisado.** Al mismo tiempo que se proporcionan las entradas al sistema se incluye la salida deseada que la red debe producir.





Redes neuronales (XI)



- ❑ **Redes con aprendizaje semi-supervisado.** Junto a la entrada, solamente se indica si el patrón se ha clasificado bien o mal durante el proceso de entrenamiento.
- ❑ **Redes con aprendizaje no supervisado.** La red debe encontrar por sí misma la regularidad presente en los datos de entrada y ordenarlos en consecuencia.



Funciones de densidad de probabilidad (I)



- **La modelización de los vectores de entrada se realiza a partir de la definición de una mezcla de funciones de densidad de probabilidad.**
- **La distribución más empleada en la distribución gaussiana multivariable.**
 - ❑ Queda definida completamente por el vector de media (μ) y la matriz de covarianza (Σ).

$$p(x) = \frac{1}{2\pi \sum} \left| \right|^{-n/2} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)' \Sigma (x_i - \mu)}$$





Funciones de densidad de probabilidad (II)



- Las distribuciones de probabilidad pueden establecerse a partir de los resultados de un proceso de cuantificación vectorial.
- Normalmente la matriz de covarianza se supone triangular.
- Computacionalmente el proceso es costoso.
 - Empleada en sistemas de reconocimiento de tamaño grande.
 - Cuantificación vectorial solamente se emplea en sistemas pequeños.



Bibliografía



- [GRA84] R. M. Gray, "Vector Quantization", *IEEE ASSP Magazine*, abril 1984, pp. 4-29.
- [KOH90] T. Kohonen, "The Self-Organizing Map", *Proceedings of the IEEE*, Vol. 78, N° 9, pp. 1464-1480, septiembre 1990.
- [LIP87a] R. P. Lippmann, "An Introduction to Computing with Neural Nets", *IEEE ASSP Magazine*, pp. 4-23, abril 1987.
- [MAK85] J. Makhoul, S. Roucos and H. Gish, "Vector Quantization in Speech Coding", *Proc. of the IEEE*, Vol. 73, N° 11, noviembre 1985, pp. 1551-1588.
- [ROB94] A. Robinson, "An Application of Recurrent Nets to Phone Probability Estimation", *IEEE Transactions on Neural Networks*, Vol. 5, N° 2, marzo 1994, pp. 298-305.
- [TEB95] J. Tebelskis, *Speech Recognition using Neural Networks*, Tesis Doctoral, Canegie Mellon University, mayo 1995.

