

# INTRODUCCION AL RECONOCIMIENTO DE PATRONES

*2015*

---

# Equipo docente:

- Pablo Muse
  - Pablo Cancela
  - Guillermo Carbajal
  - Sergio Martinez
  - Alicia Fernández
-

# Objetivos generales

- Brindar un marco teórico para el análisis y resolución de problemas de reconocimiento, clasificación e identificación de patrones, en forma automática.
  - Realizar hincapié en la evaluación de desempeño de los modelos propuestos y su aplicación en la resolución de problemas prácticos.
-

# Objetivos específicos

- Identificar los principales componentes de un Sistema de Reconocimiento de Patrones
  - Estudio de algoritmos asociados a las distintas etapas y modelos de un SRP
  - Herramientas de evaluación de desempeño
  - Análisis de casos prácticos
-

# Algunas preguntas que busca responder el curso

- Cual es la mejor estrategia para resolver un problema concreto?
  - Cuales son las características más significativas?
  - Cuantas muestras necesito?
  - Cual es el grado de corrección de mi solución?
  - De las soluciones propuestas cual es más eficiente y eficaz?
-

# Estrategia

- Presentar un diagrama de bloques completo de un SRP
  - Presentar distintos algoritmos que resuelven los bloques.
  - Analizar métodos de evaluación
  - Analizar estrategias de fusión y combinación de clasificadores.
-

# Actividades específicas

- Clases teóricas: Martes y Jueves de 8:15-10:00
  - Responder cuestionario
  - Entrega de ejercicios obligatorios- clases de consulta.
  - Defensa trabajo final de curso.
-

# Materiales y métodos

- Guías de clase
  - Bibliografía
  - Software de simulación (Matlab, Weka)
-



# Bibliografía

- Pattern Classification (2nd. Edition) Duda, Hart Stork– John Wiley & Son 0471056693-2001
- Pattern Classification. A unified view of statistical and neural approaches. Jürgen Schürmann. John Wiley & Son. 1996
- Pattern Recognition: A Statistical Approach Devijver y Kittler- Prentice-Hall - 1982
- Combining Pattern Classifiers Methods and Algorithms- Ludmila I Kuncheva 047121078 –2004
- Pattern Recognition and Machine Learning C. M Bishop, 2006
- Anil Jain-Statistical Pattern Recognition: A Review - 2000
- Notas del Curso de Reconocimiento de Patrones y Análisis de Imágenes de Francisco Cortijo. Decsai UGR

# **1- Introducción**

---

# Percepción



¿Porqué vemos caras en la luna?

# Percepción



Pareidolia: la percepción imaginada de un patrón o un significado donde no lo hay.

# Percepción



# Percepción vs SRP

- ¿Por qué los humanos son tan buenos reconociendo patrones?

Dra. Nouchine Hadjkhani de la Universidad de Harvard: Los humanos llegan a este mundo “con los cables preconectados” para detectar caras. Un buen sistema de RP proporciona una ventaja competitiva a la hora de la supervivencia (discriminar amigo/enemigo) y reproducción.

- ¿Por qué no imitarlos al diseñar sistemas automáticos de reconocimiento de patrones?
  - Aún no se sabe como los humanos reconocen los patrones.
-

# Que es el reconocimiento de patrones?

- “El estudio de cómo las máquinas pueden observando el ambiente aprender a distinguir **patrones** de interés de un fondo y realizar decisiones razonables sobre las categorías de los mismos”. Anil Jain.
  - “El acto de tomar datos crudos y hacer una acción basados en la categoría de los patrones”. Duda et al.
  - Encontrar regla de decisión automática que transforma medidas en asignaciones a clases.
  - Dar nombres a los objetos basados en observaciones.
-

# Percepción

- A la edad de cinco años la mayoría de los niños pueden reconocer dígitos y letras: – Grandes, pequeños, escritos a mano, rotados, incompletos... Variaciones de la letra R



- Reproducir esta capacidad por medio de los ordenadores es sorprendentemente difícil, de hecho, tras 50 años de investigación no existe actualmente ningún sistema de R.P. de propósito general que sea capaz de aproximarse a la capacidad humana.



# Percepción vs SRP

- ¿Que ventajas tienen los SRP?:
    - Pueden realizar tareas monótonas y repetitivas con menos errores.
    - Trabajar en zonas peligrosas
    - Funcionar en lugares inhóspitos.
    - Económicos y autónomos.
    - Realizar cálculos precisos
-

# Aplicaciones Interesantes

- Maquinas de recomendaciones: Ej: Netflix. El 75% de los videos seleccionados por un usuario promedio provienen de las recomendaciones producidas por su sistema de propuesta.
- Sistemas de reconocimiento: de escritura manual (servicio postal USA), biometría (facial, huellas, andar)
- Aplicaciones de Big Data: descubrir patrones y correlaciones en datos no estructurados, encontrar relaciones dependencias, anomalías.

---

# **Analisis de un ejemplo de SRP**

---

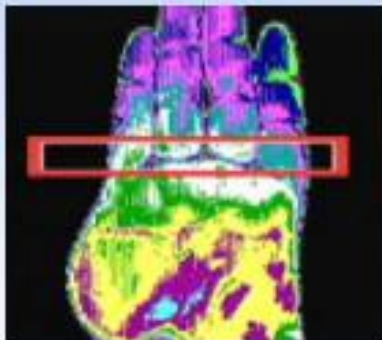
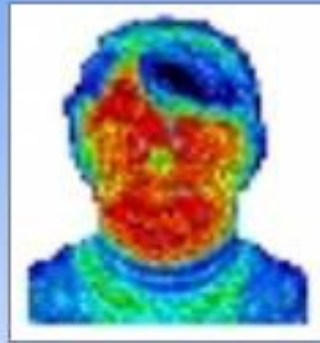
# CSI (Crime Scene Investigation)

- Identificación de un criminal a través de la observación de la escena del crimen.
  - procesar la escena: adquisición de objetos
  - analizar sus características
  - clasificarlos
  - identificarlos
  - post-procesar – aumentar confiabilidad
-

# Sistema Biométrico

- Huellas dactilares
- ADN
- Iris
- Forma y tamaño de la mano
- Escritura
- Rasgos faciales
- Forma de caminar

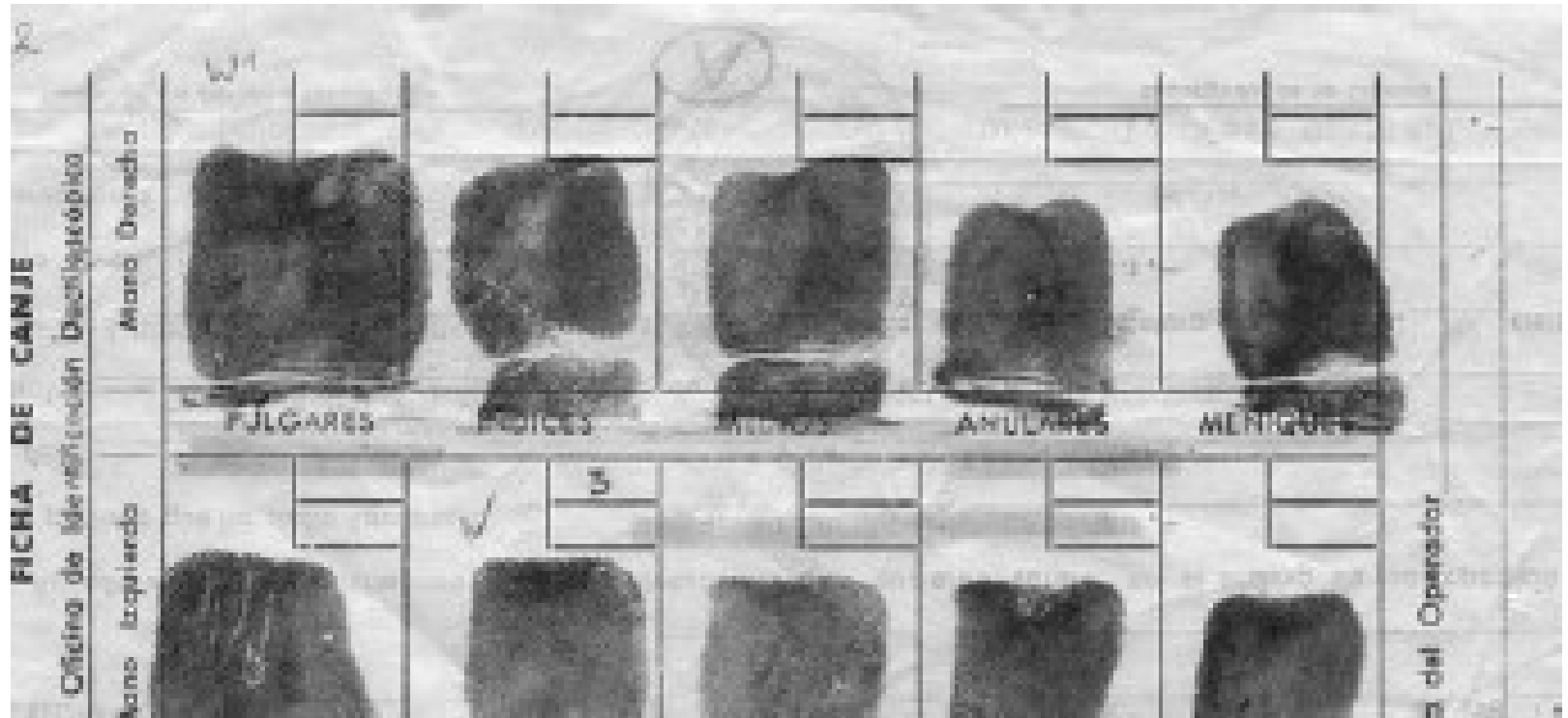
# Biometric Characteristics



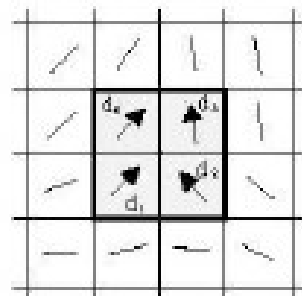
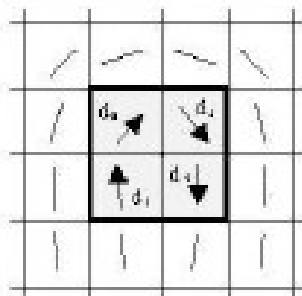
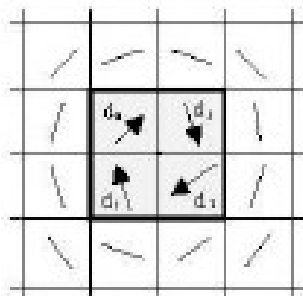
# Adquisición- Sensado

- Toma de la huella
- Toma directa entintado
- Sensores capacitivos
- Critico calidad y tipo de información adquirida que condiciona el procesamiento posterior. Ej: rodada o plana, manchones (entintado)
- **Complementariedad:** Cada característica física tiene ventajas y limitaciones:
  - Poder de discriminación
  - Invasiva vs no invasiva
  - Colaborativa vs no colaborativa

# Problemas de adquisición

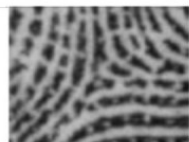
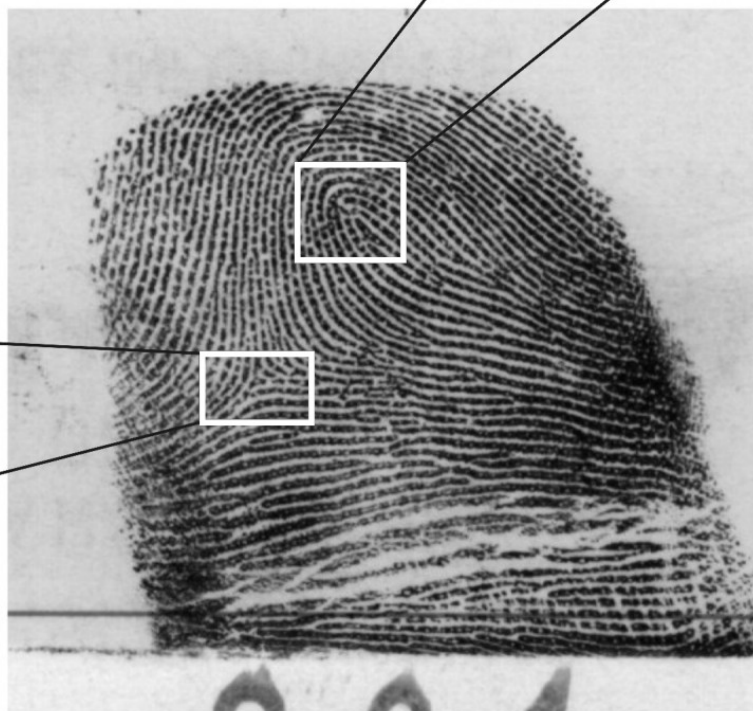






**CORE**

$$P_{G,C}(i,j) = -180^\circ$$

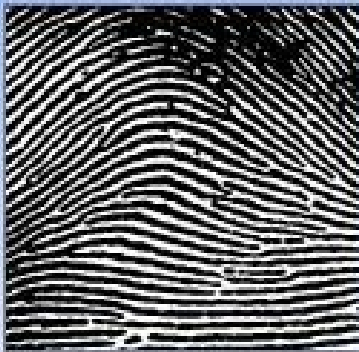


**DELTA**

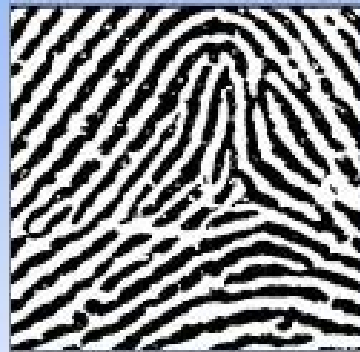


# Fingerprint Classification

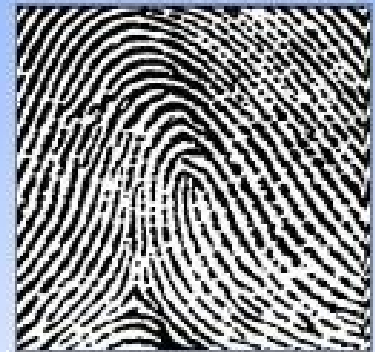
- Assign fingerprints into one of pre-specified types



Arch (A)



Tented Arch (T)



Right Loop (R)



Left Loop (L)

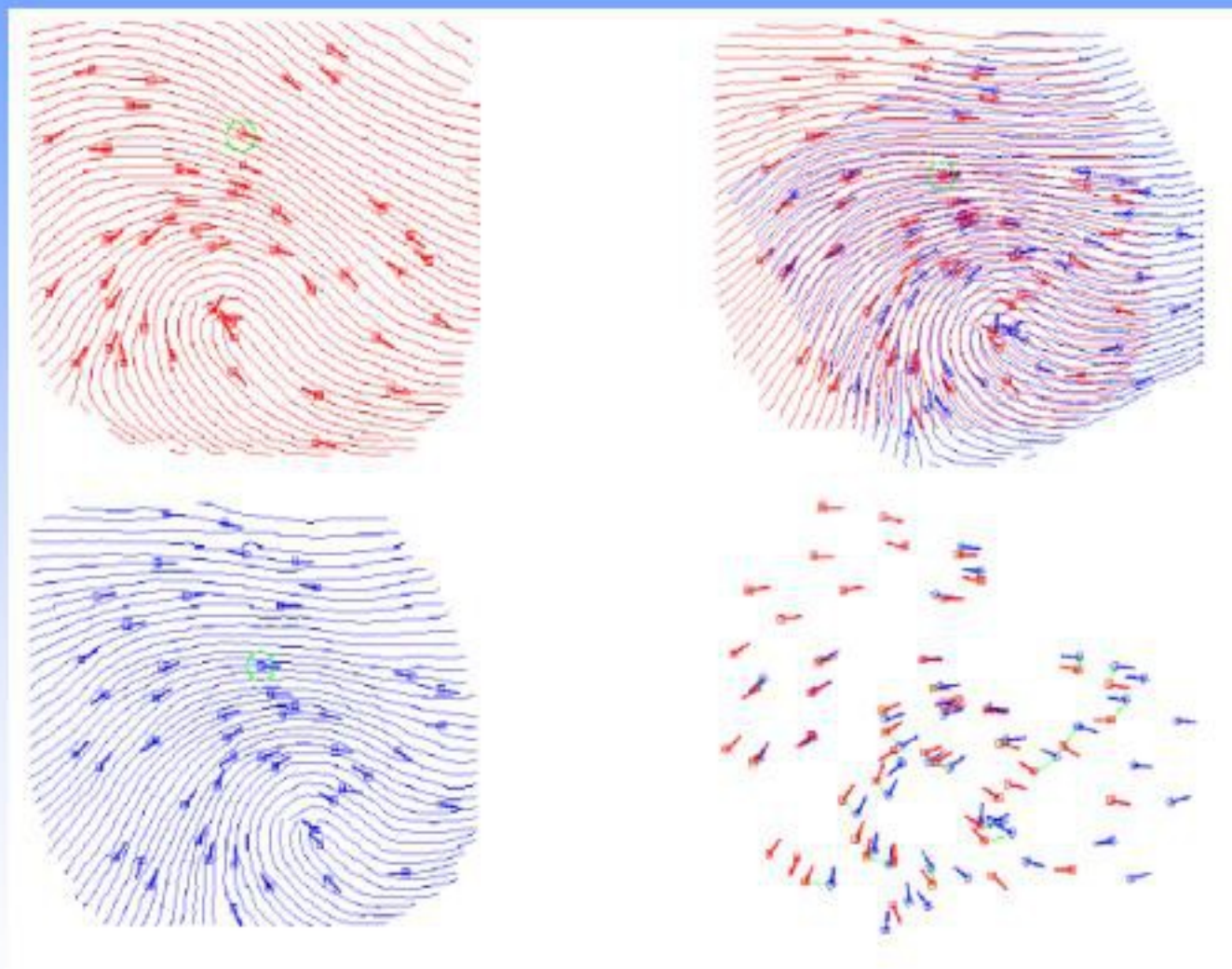


Whorl (W)



Double Loop (W)

# Minutiae Matching Result



# Ejemplos

Dominio	Aplicación	Patrón Entrada	Clases
Data mining	Búsqueda patrones significativos	Puntos espacio multidimensional	Clusters compactos y bien separados
Clasificación de documentos	Búsqueda en Internet	Texto	Categorías (dep.,negocios)
Biométrica	Identificación de personas	Iris, huellas, caras	Usuarios autorizados
Sensado remoto	Pronóstico de cosechas	Imágenes multiespectrales	Categorías de tierras.
Reconocimiento de voz	Acceso a información sin operador	Señal de voz	Palabras habladas

# **Identificación de los conceptos y componentes básicos de un SRP**

---

# Objetivo de SRP

- Asignar un objeto o fenómeno físico (patrón, en general) a una clase o categoría.



**Reconocimiento de Patrones:** regla de decisión automática que transforma medidas en asignaciones a clases.

# Conceptos Generales

- En el R.P. no se clasifican directamente las entidades u objetos sino su descripción.
- Esta descripción se crea a partir de un modelo de la entidad construida a partir de unas características y algún tipo de representación.
- La representación también es dependiente del problema y puede ser un vector, una cadena, un grafo, etc... Esta representación es utilizada por un clasificador para asignar la entidad a una clase.

# Conceptos Generales

- **Modelo:** representación de un patrón.
- **Características o atributos:** medidas que componen las representaciones.
- **Espacio de representación o de características:** conjunto de todas las representaciones posibles para un cierto problema, universo de operación del SRP.



# Características (features)

- Los objetos se describen por características :
- **Cuantitativas:**
  - Continuos: presión, longitud
  - Discretos : número de ciudadanos, tantos
- **Cualitativas**
  - Ordinal : grado de educación
  - Nominal: profesión, marca de auto.
- Se necesita una metodología para pasar de características cualitativas a cuantitativos en general son métodos heurísticos subjetivos , Ej: medida inteligencia, conocimiento, belleza, sentido del humor.

# Ej: características

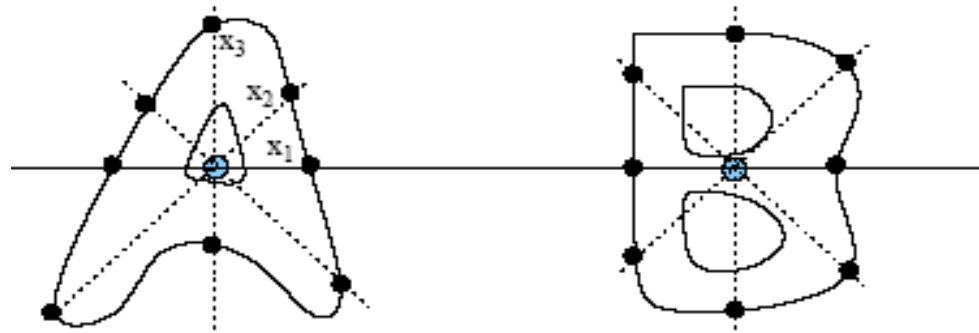
Parámetros involucrados :

$c = 26$  - número de caracteres

$d = 8$  - número de medidas

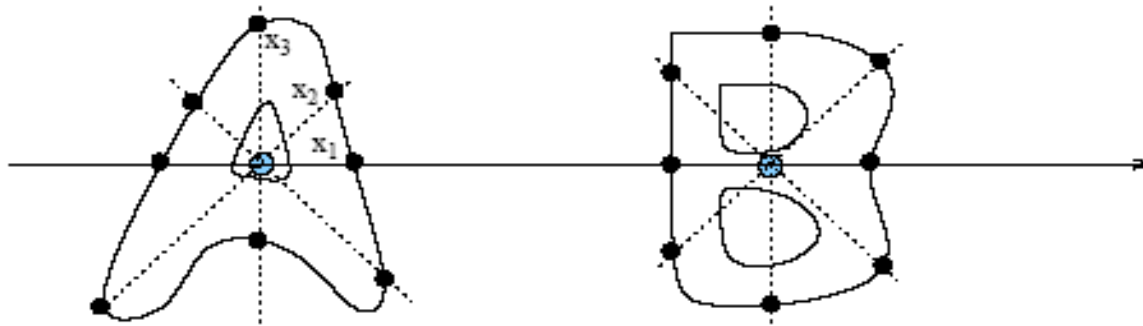
$x_i$  con  $i = 1 \dots d$  : distancia entre el centro de gravedad y el punto de intersección más lejano en la semirrecta formada

por  $\frac{(i-1)\pi}{4}$  con el eje  $0x \rightarrow$



# Ejemplo: Reconocimiento de caracteres

- $\mathbf{x} = (x_1, \dots, x_8)$  con  $x_i$  distancia extremos semirrectas y centro gravedad.



- “Buena representación”: una en la que las relaciones estructurales entre los componentes son simples y se revelan en forma natural y en la que el modelo verdadero (desconocido) se puede expresar.

# Clases y Etiquetas

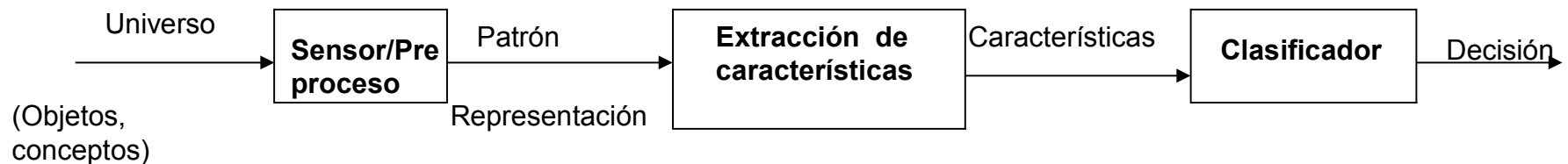
- Para el reconocimiento automático, es importante que patrones que describen objetos de una misma clase, presenten características similares. Patrones que describen objetos de diferentes clases presenten características diferenciadas.
- Asumiremos:
  - $C = \{w_1, \dots, w_c\}$ : conjunto finito de eventos
  - $w_i$  de  $C$ : clase del conjunto de clases
  - $c = \text{card}(C)$ : número de clases
  - Las  $c$  clases  $w_i$  son mutuamente excluyentes y completas

# Clasificación de Patrones

- $\mathbf{x}$ : vector de características: colección de observaciones.
- Clasificación de patrones: proceso de inferir  $w$  de  $\mathbf{x}$ , proceso de asignar nombres a las observaciones  $\mathbf{x}$ .
- Mapeo  $w \rightarrow \mathbf{x}$ , uno a muchos debido a la variabilidad, muchos patrones representan el mismo concepto: Ej  $A A \ A \ A \ \ A$  ,
- El mapeo  $\mathbf{x} \rightarrow w$ , puede o no ser único.
- Variabilidad: descrita por la distribución de las clases en el espacio de características  $\mathbb{R}^d$ .

# Sistemas de Reconocimiento de Patrones

- Modelos perceptuales:
  - adquisición de datos sensoriales/preproceso
  - extracción de características
  - toma de decisiones



**Etapas en un Sistema de Reconocimiento de Patrones**

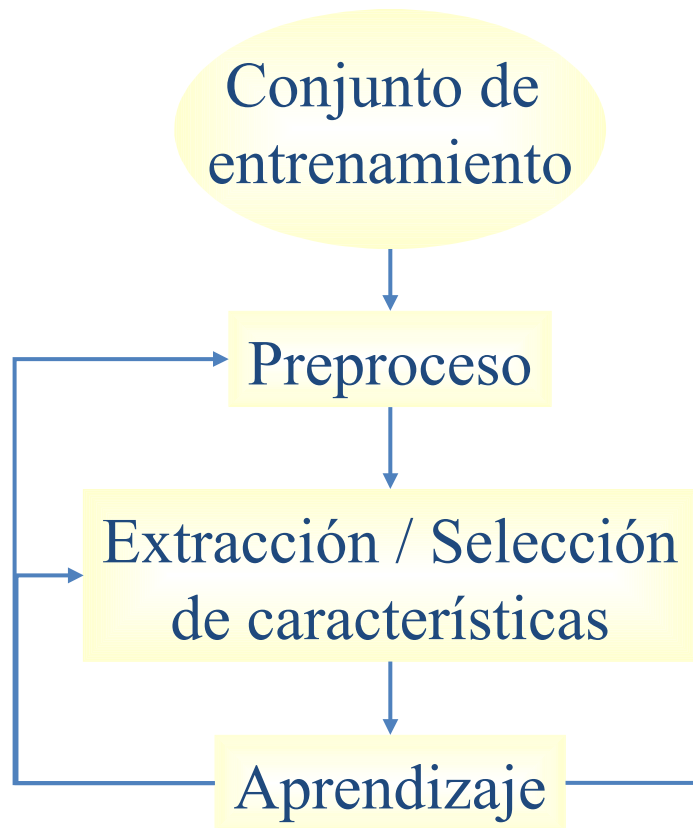
# Diseño de SRP

1. Inferencia del modelo a partir de un conjunto de datos de entrenamiento
2. Desarrollo de reglas de decisión prácticas
3. Simulación y evaluación del rendimiento del sistema.

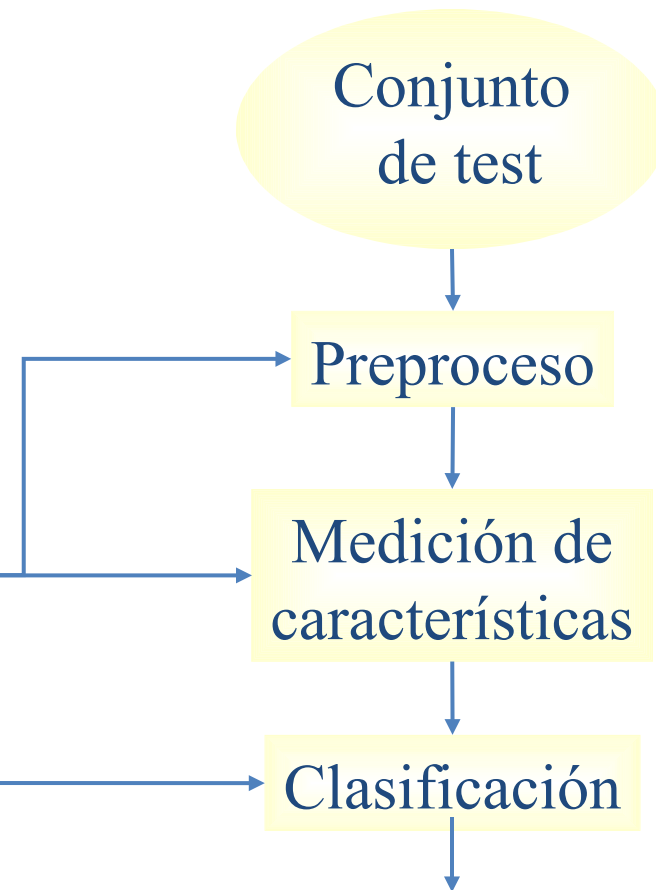
(Machine learning, reconocimiento de patrones estadístico)

# SRP Estadístico

## Fase de Entrenamiento



## Fase de Clasificación





# Sensor

- Función del sensor: Medición, dar representación de los elementos a ser clasificados.
  - Condiciona el rendimiento del sistema
  - Debería adquirir todas las propiedades físicas que permiten discriminar los objetos.
  - Limitaciones: no se dispone conocimiento, mediciones no intrusivas, económicamente no viable, datos históricos (fichas dactilares)
-

# Preproceso

- Modificar la representación inicial para poder resaltar las características relevantes: Filtraje, Realce, Cambio de espacio, etc.
- No suele existir mucho conocimiento que dirija el preproceso.

# Aprendizaje

- El proceso de aprendizaje permite establecer un modelo: establecer los parámetros del modelo o adquirir conocimiento sobre el problema.
- Tipos de aprendizaje:
  - Recopilar conocimiento (deductivo) humano sobre el problema. (Sistemas expertos).
  - Adquirir (inductivamente) el conocimiento a partir de ejemplos específicos. Ej. Inferencia gramatical, estimación de parámetros.

# Aprendizaje

- El aprendizaje se puede realizar en una fase previa al reconocimiento (“diseño del clasificador”) o continuar durante el proceso de reconocimiento (Aprendizaje continuo).
- Un objetivo del Aprendizaje puede ser la determinación del conjunto de descriptores “óptimo”. Este proceso se llama **selección de características**.
- La selección de características se suele llevar a cabo mediante técnicas estadísticas y puede requerir conocimiento profundo de la naturaleza del problema.

# Selección y Extracción de Características:

- Extraer la información que puede permitir la discriminación.
  - Eliminar información redundante e irrelevante.
  - Reducir la dimensionalidad del problema.
-

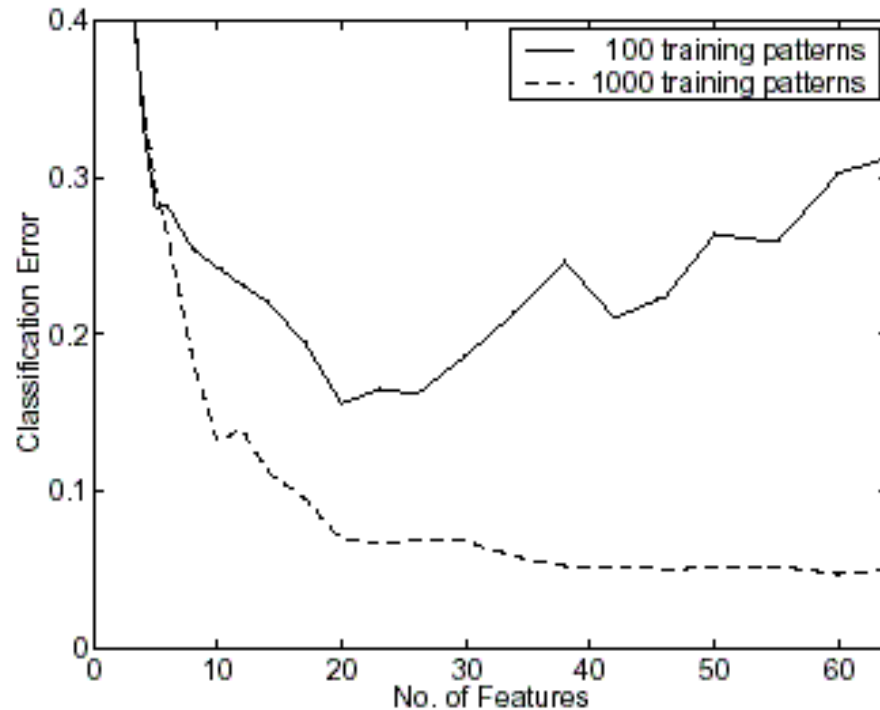
# Dimensionalidad de los datos

- El desempeño de un clasificador depende de la **relación entre el número de muestras, el número de características** y la complejidad del clasificador.
- **Maldición de la Dimensionalidad:** En la práctica se ha observado que el agregar características puede degradar el desempeño si el número de muestras es pequeño en relación al de atributos.
- En clasificadores paramétricos la confiabilidad con la que se estiman los parámetros disminuye al aumentar número de características para un número de muestras dado.

# Dimensionalidad de los datos

- Para mejorar el rendimiento del sistema la otra alternativa es aumentar la complejidad del clasificador
- El rendimiento disminuye cuando aumenta la complejidad del clasificador. – A este hecho se le llama “**el problema de la generalización**”, su explicación es que el clasificador se ajusta tanto a las muestras de entrenamiento que no “captura” bien el comportamiento de las muestras nuevas (**sobre-entrenamiento**)
- Hay una teoría filosófica para la elección de clasificadores sencillos. Ya William de Occam (1284-1347?) afirmaba (“navaja de Occam”) que si dos explicaciones son igual de buenas se debe elegir la simple sobre la complicada.

# Maldición de la dimensionalidad



➤ Buena práctica:  $n/d > 10$



# Reducción de la Dimensionalidad

- Razones para mantener la dimensionalidad tan baja como sea posible:
    - Costo de medida
    - Precisión de la clasificación
  - Un conjunto saliente de características simplifica la representación y el diseño del clasificador. El bajar mucho puede hacer que se pierda poder de discriminación.
-

# Selección de Características

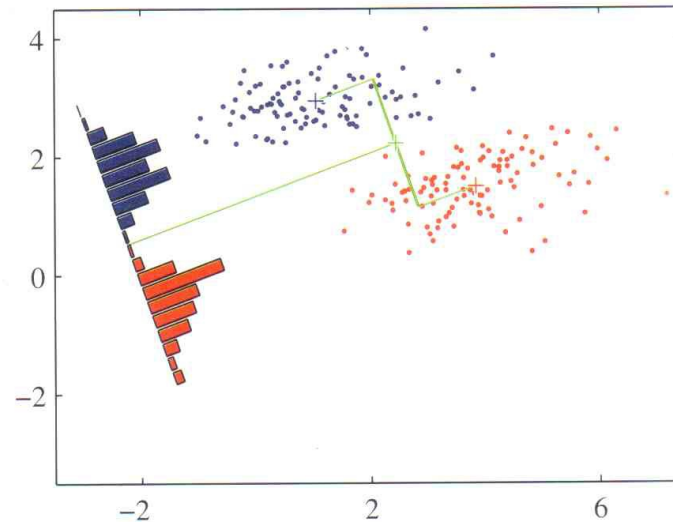
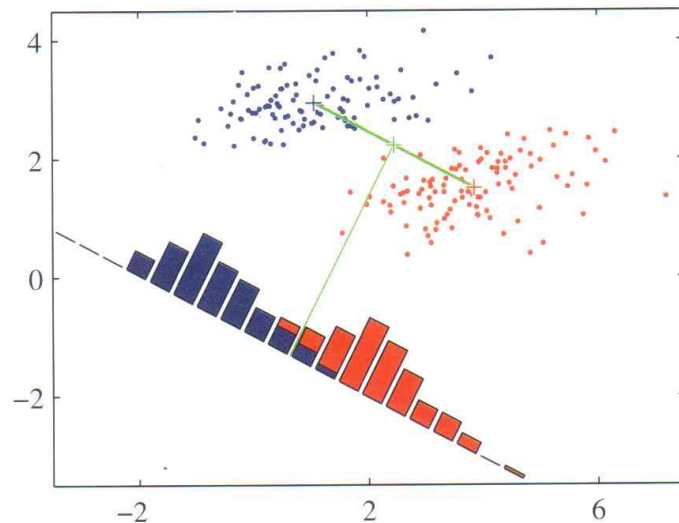
- Tiene por objeto seleccionar las características (sensadas o transformadas en el proceso de extracción) con mayor poder de discriminación.
- **Filtrado (filtering)**: Selecciona las características en forma independiente del clasificador, usando un criterio de “relevancia” .
- **Encapsulado (wrapping)**: Selecciona los subconjuntos de características en función del desempeño de un clasificador.
- **Intrínseco (embedding)**: Realizan la selección en el proceso de aprendizaje devuelve un subconjunto de características y el clasificador entrenado. Evalúo costo de agregar o quitar característica pero no reentreno.

# Métodos de Selección de características

Método	Propiedades	Comentarios
Búsqueda exhaustiva	Evalúa todas las combinaciones de $p$ tomadas de $d$	Garantiza encontrar el subconjunto óptimo
Mejores características individuales	Selecciona las $p$ mejores características individuales	Computacionalmente simple pero no garantiza un subconjunto óptimo
Selección secuencial hacia delante (SFS)	Selecciona la mejor característica y agrega una por vez tal que combinada maximiza función criterio.	Una vez que se agrega una no se puede quitar, Computacionalmente eficiente.
Selección secuencial hacia atrás (SBS)	Empieza con todas y quita una a la vez	Una vez que se quita una no se puede volver al conjunto optimo.

# Extracción de características

- Métodos que permiten determinar un conjunto de dimensionalidad menor en el espacio original ( $d$  menor  $p$ ).  
Crean nuevas características combinación de las características sensadas.



# Métodos de Extracción de características

Método	Propiedades	Comentarios
Análisis de componentes principales (PCA)	Mapas lineales, rápidos, basados en vectores propios.	Tradicional, bueno para datos Gaussianos.
Análisis de Discriminante Lineales	Mapas lineales supervisados, rápidos, basados en vectores prop.	Mejor que PCA para clasificación.
Análisis de Componentes Independientes (ICA)	Mapas lineales, iterativo, no Gaussiano	Se usa para separar mezclas de fuentes con distribución no Gaussiana.
PCA no lineal	Criterio no Gaussiano, usualmente iterativo	Enfoque redes neuronales.
Self-Organizing Map (SOM)	No lineales, iterativos	Basados en redes neuronales, adecuado para baja dimensionalidad

# **Clasificador funciones discriminantes y regiones de clasificación**

---

# Clasificador

- Objetivo: dividir el espacio de características en regiones de decisión asociadas a las clases.
- La clasificación de un patrón consiste en localizar a que “Región” pertenece.
- Las fronteras entre las Regiones de decisión se llaman Fronteras de Decisión.
- La elección de un clasificador depende del problema.
- Cuestiones a resolver son: - ¿Como hacer mínimo el error de clasificación? - ¿Cual debe ser su complejidad?

# Clasificador

- Asigna los “objetos” percibidos (a los que no se le conoce la clase de pertenencia) a la clase adecuada.
- Dado un patrón  $\mathbf{x} \in \mathbb{R}^d$ ,  $\mathbf{x} = (x_1, x_2, \dots, x_d)$
- Objetivo determinar a cuál de las  $c$  clases de  $C = \{w_1, w_2, \dots, w_c\}$  pertenece dicho patrón

$$D : \mathbb{R}^d \rightarrow C, \quad D(\mathbf{x}) = w_i \quad i = 1, \dots, c$$

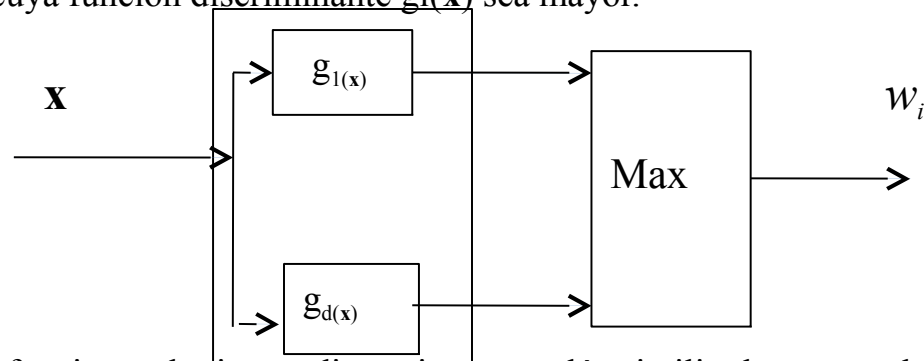


# Funciones Discriminantes

- Podemos expresar el clasificador  $D$  en términos de  $c$  funciones discriminantes,  $g_{i(x)}$ :

$$D(\mathbf{x}) = \omega_i \Leftrightarrow g_i(\mathbf{x}) > g_j(\mathbf{x}) \quad \forall j \neq i \quad i, j = 1, \dots, c$$

- El clasificador asigna  $\mathbf{x}$  a la clase cuya función discriminante  $g_i(\mathbf{x})$  sea mayor.



- $g_i(\mathbf{x})$ : Probabilidades a posteriori, funciones de riesgo, distancia que evalúa similitud entre  $\mathbf{x}$  y las clases.

# Región de Decisión

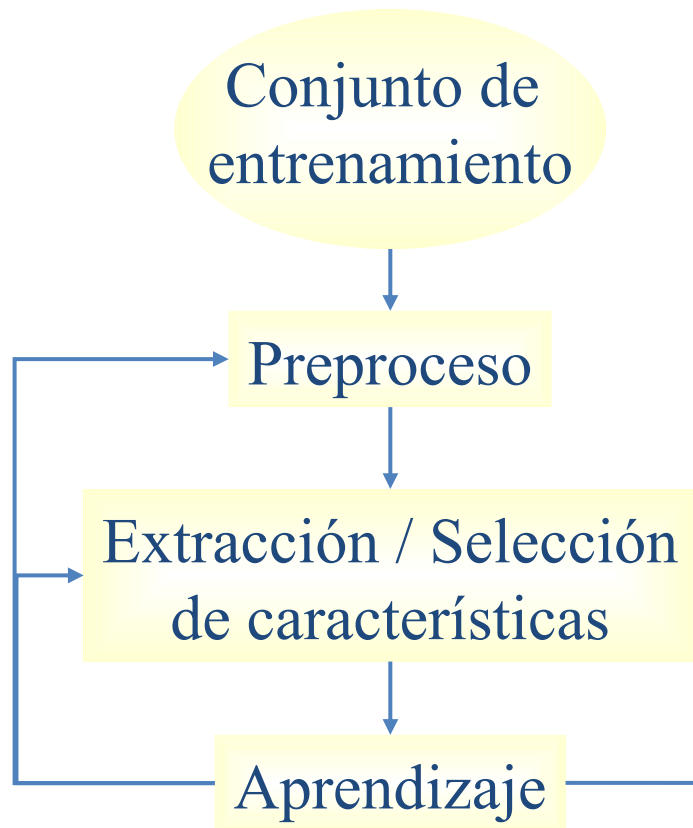
- **Fronteras de decisión:** Existen zonas del espacio de características donde  $g_i(\mathbf{x}) = g_j(\mathbf{x})$ , igual probabilidad de pertenecer a distintas clases.
- La **Región de decisión**  $i$  esta conformada por el conjunto de puntos tales que la función  $g_i$  es máxima para esos puntos. A todos los puntos de esa región se le asigna la clase  $w_i$ .
- Las regiones de decisión son determinadas por el clasificador o equivalentemente, por el conjunto de las funciones de decisión.

# **Reconocimiento de patrones estadístico**

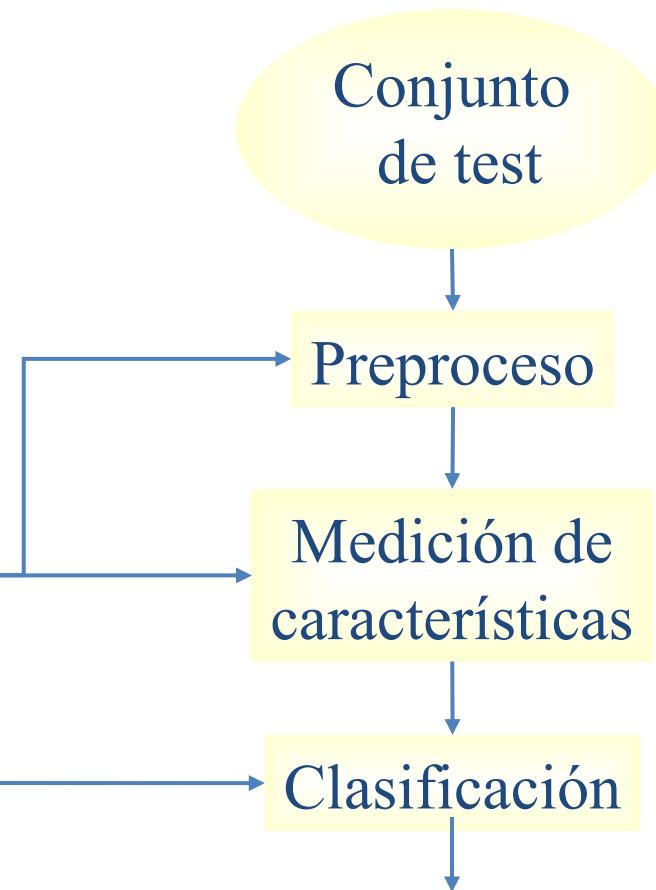
---

# SRP Estadístico

## Fase de Entrenamiento



## Fase de Clasificación



# SRP Estadístico

- Patrones de la misma clase están sujetos a variaciones las cuales no necesariamente son causadas por un proceso randómico.
  - Es útil tratarlos como si lo fueran.
  - Estadística es una herramienta útil para describir variaciones.
  - Objetivo: Encontrar mapeo óptimo entre el espacio de medida y el espacio de decisión. **Problema de optimización estadística.**
-

# Enfoque estadístico

- Patrón:  $[\mathbf{x}, w]$
- $\mathbf{x} = [x_1, x_2, \dots, x_d]$  : vector de características
- $w$  : estado de la naturaleza, significado, clase, variable discreta.
- $\{w_1, w_2, \dots, w_c\}$  : conjunto finito de estados
- Diseño del clasificador requiere conocimiento de la situación estadística específica :
  - Conocimiento de las distribuciones relevantes: probabilidad condicional de clase  $p(\mathbf{x}/w_i)$ .
  - *Conocimiento de las penetraciones.* Cada clase  $w_i$  tiene una probabilidad a priori  $P(w_i)$ . Conocimiento que se tiene cuando no se conoce la observación.

# Enfoque estadístico

- Usando Bayes se puede estimar la Probabilidad a posteriori  $P(w_{i/\mathbf{x}})$ - Probabilidad de la clase una vez que se observa  $\mathbf{x}$ .

$$P(w_i / \mathbf{x}) = \frac{p(\mathbf{x} / w_i)P(w_i)}{p(\mathbf{x})} \quad p(\mathbf{x}) = \sum_i^c p(\mathbf{x} / w_i)P(w_i)$$

- ¿Como determino la regla de decisión a partir de  $P(w_i/\mathbf{x})$ ?

# Regla de mínimo error de Bayes

- Supongo que el costo de no equivocarme es 0, y del equivocarme 1 para cualquier clase.
- Con la evidencia a la vista decido la más probable.

Decido asignar  $\mathbf{x}$  a  $w_i$  si  $P(w_i/\mathbf{x}) = \max_{1 \leq j \leq c} P(w_j/\mathbf{x})$

$$P(error) = \min [P(w_1/\mathbf{x}), P(w_2/\mathbf{x})]$$



# Costo de decisión

- Existe un costo asociado a cada decisión
- Ej: Autorización de firmas – 2 clases:
  - Firma genuina
  - Firma falsa
- Costos distintos:
  - Rechazar una firma genuina
  - Admitir una firma falsa.
- Diseño la regla de decisión para minimizar el costo total: **Regla de Decisión del costo mínimo de Bayes.**

# Clasificación

- En la mayoría de las aplicaciones prácticas, no se conocen  $P(w_i)$ ,  $p(\mathbf{x}/w_i)$  ni  $P(w_i/\mathbf{x})$ . En tal caso, deben estimarse a partir de un conjunto de entrenamiento.
- Se presentan dos alternativas:
  - Estimar directamente  $P(w_i/\mathbf{x})$  o
  - A partir de  $P(w_i)$  y estimación de  $p(\mathbf{x}/w_i)$

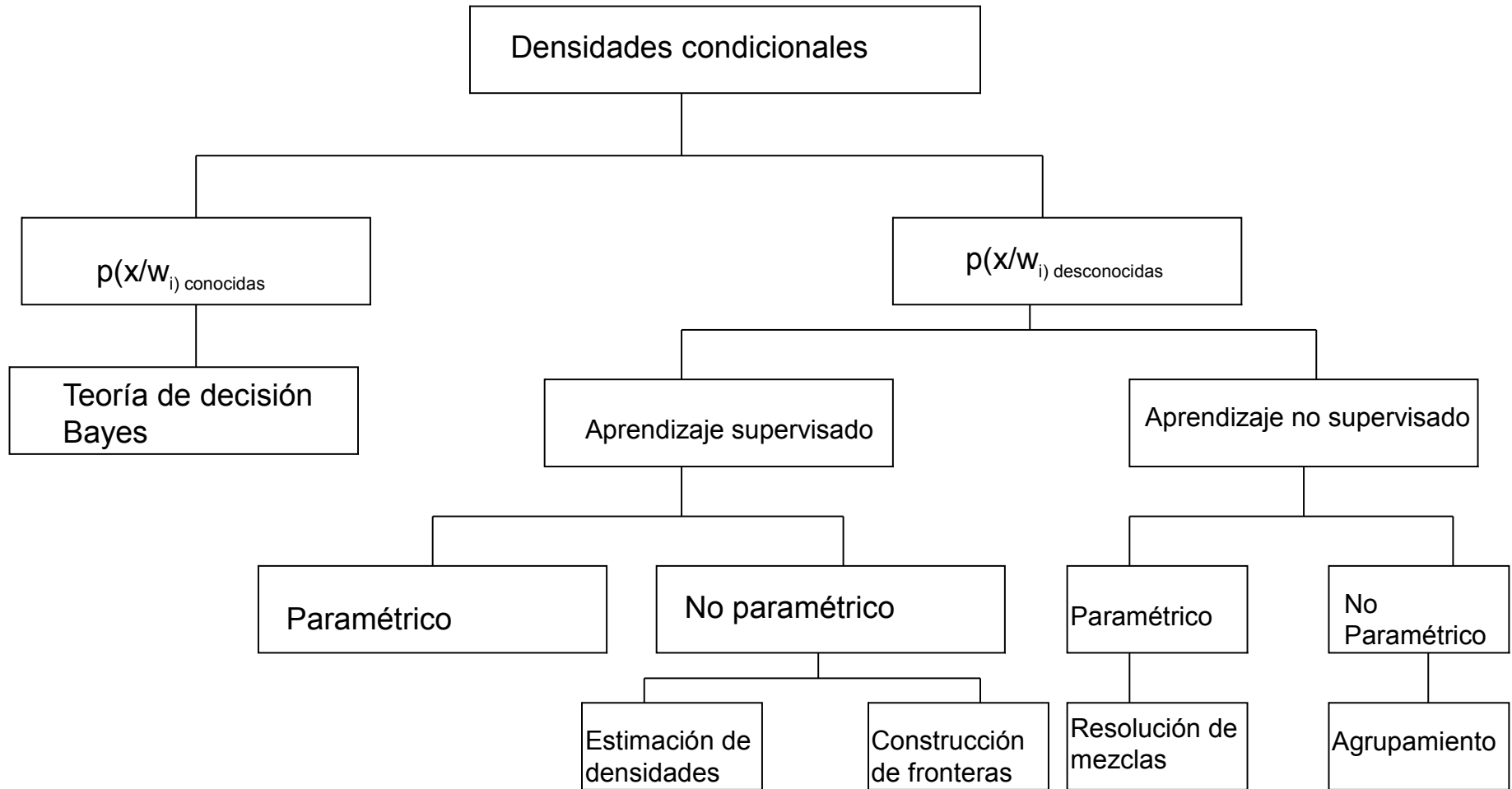
$P(w_i)$  determinables por el conocimiento previo del problema en muchos casos se supone clases equiprobables.

$p(\mathbf{x}/w_i)$  se puede estimar por métodos paramétricos o no paramétricos.

# Reconstrucción de leyes de probabilidad

- Problema inverso: a partir de un conjunto de muestras tomadas en forma más o menos randómica queremos recuperar leyes de probabilidad del modelo estadístico desconocido.
  - Más complicado que la interpolación de funciones continuas a partir de las muestras.
    - Diferencias en la dimensionalidad
    - Muestreo randómico
    - Muestreo disperso: Conjuntos de aprendizaje pequeños
-

# Reconocimiento de Patrones Estadístico



# Aprendizaje

## Aprendizaje supervisado

- Se dispone de un conjunto de muestras etiquetadas.

Clasificación por  
criterios de vecindad

## Aprendizaje no supervisado

- Se dispone de un conjunto de muestras sin etiquetar.
- El número de clases puede ser también desconocido.

Agrupamiento

# Aproximaciones

## Aproximación paramétrica

- Conocimiento a priori de la forma de las distribuciones de probabilidad de cada clase sobre el espacio de representación.
- Las fronteras de decisión están definidas por dichas distribuciones.

## Aproximación no paramétrica

- El único conocimiento a priori se refiere al inducido a partir de un conjunto de muestras (conjunto de entrenamiento) de las que se conoce su clase.
- Las fronteras de decisión están definidas por dichas muestras.

# Aprendizaje supervisado

- **Aproximación paramétrica**

- Conozco o supongo que las clases se distribuyen con un modelo paramétrico por ej: Gaussiano. Debo a partir de los datos (conjunto de aprendizaje) determinar el valor medio y la matriz de covarianza.

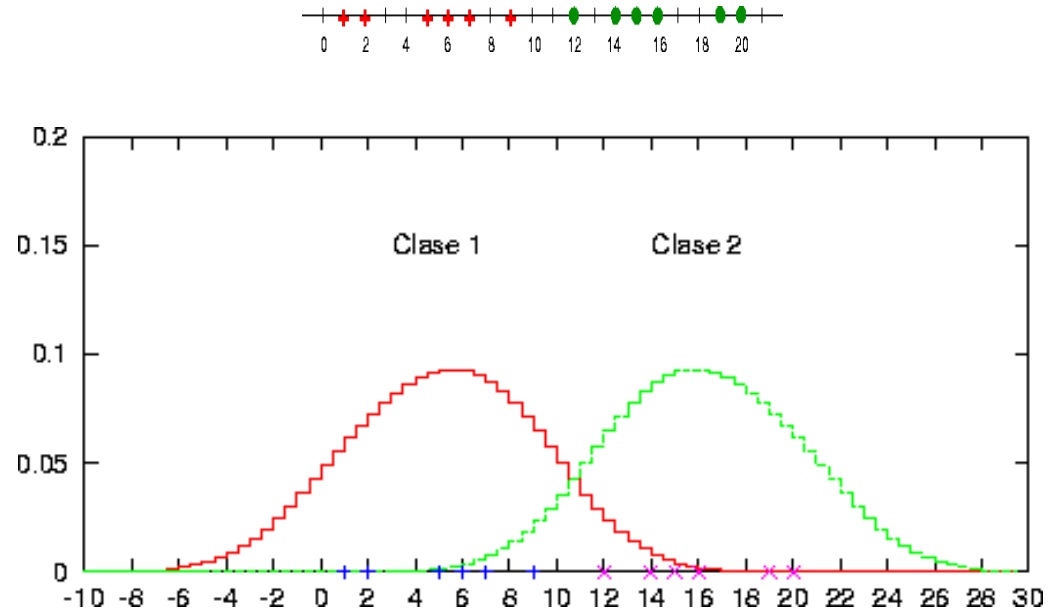
- **Aproximación no paramétrica**

- Estimo el valor de la función densidad o directamente la probabilidad a posteriori a partir de la información proporcionada por el conjunto de prototipos sin imponer un modelo a priori.
  - Estimación por Núcleos o ventanas de Parzen
-

# Estimación de densidades - Parzen

- Sea  $k_i$  numero de prototipos de la clase  $w_i$  en un volumen  $V$  centrado en  $\mathbf{x}$ .
- $n_i$  numero total de prototipos de la clase  $w_i$

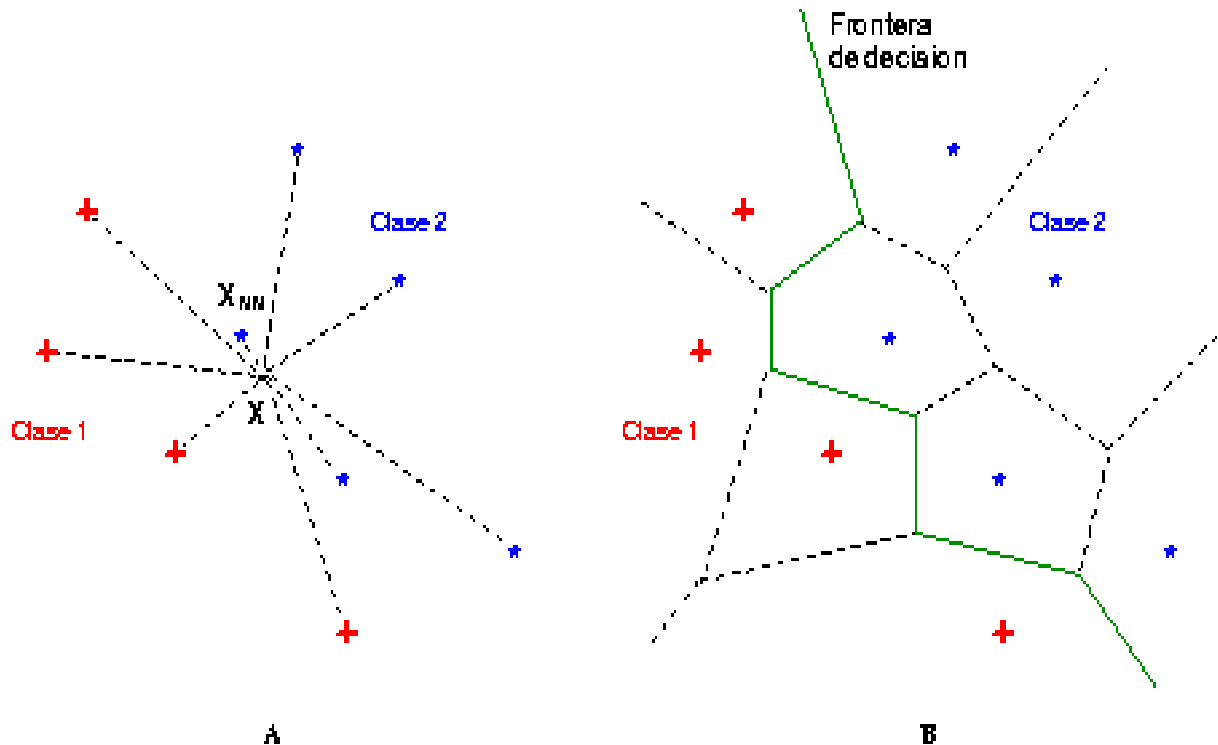
$$p(\mathbf{x} / w_i) \approx \frac{k_i / n_i}{V}$$





# Regla del vecino más cercano 1-NN

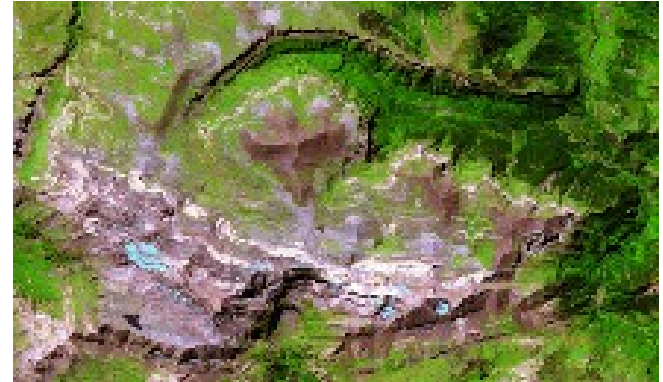
- Selecciono  $w_j$  si: 
$$d(\mathbf{x}, \mathbf{x}_{NN}) = \min_{i=1..c} \{\delta(\mathbf{x}, \mathbf{x}_i)\}$$



Interpretación: Divide al espacio en n regiones de Voronoi

# Aprendizaje no supervisado y análisis de agrupamiento

- Se tiene un conjunto de patrones de entrenamiento  $x$  para los cuales no conocemos sus etiquetas.
- Ej: sensores remotos de imágenes satelitales de terrenos o galaxias en los que sería muy costoso o imposible recoger la clase cierta de imágenes.
- Dado un conjunto de entrenamiento suficientemente grande se puede inferir la función de probabilidad conjunta  $p(x)$ .



# Distribución conjunta multimodal

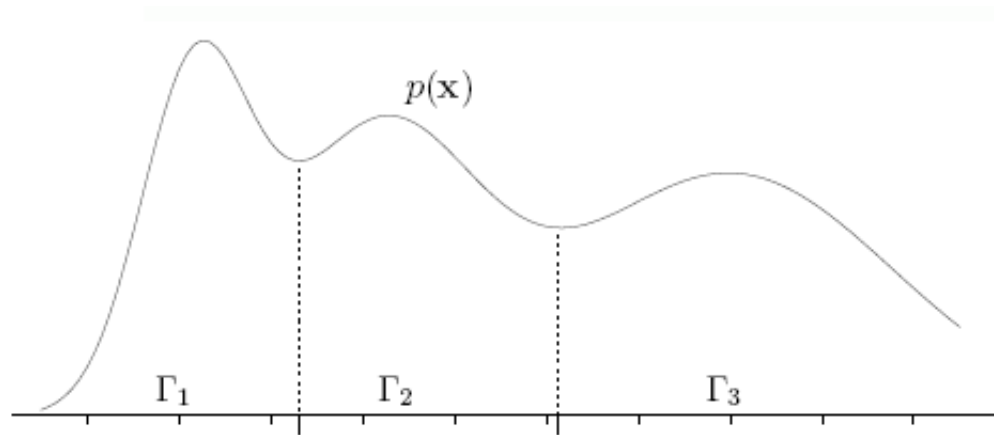
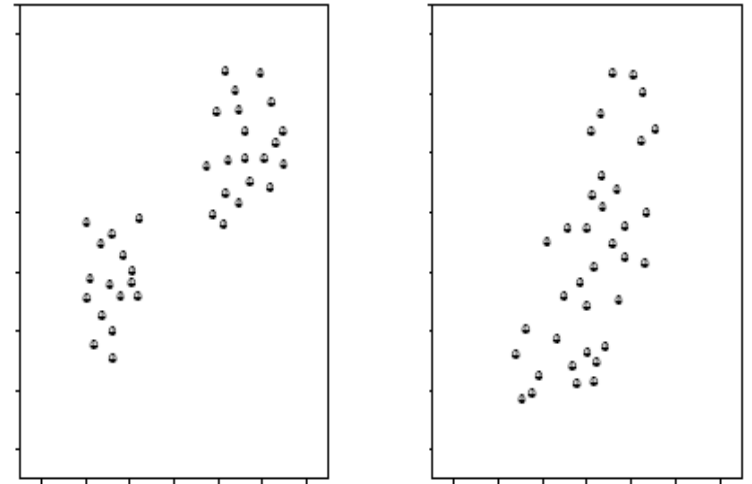


Figure 1: Regiones asociadas a cada clase

- Si  $p(x)$  multimodal cada uno de los modos debería corresponder con la distribución condicional de las clases.
- Identificar los modos permite partir el espacio de observación en regiones disjuntas asociadas a cada una de las clases presentes.

# Datos estructurados



- Técnicas de agrupamiento:
  - Detectar y agrupar racimos de puntos
  - Analizar y extraer la estructura presente en el conjunto de muestras de entrenamiento.
- **Conjunto de datos bien estructurado:** contiene regiones de alta densidad separadas por regiones vacías o con poca densidad

# Medida de similitud y criterios de agrupamiento

- $x$  pertenece a un agrupamiento si esta cerca del mismo o es parecido.
- **Medida de proximidad o similitud:** Medir afinidad de un punto a un agrupamiento:
  - Similitud con otros puntos en el agrupamiento
  - Similitud con modelo del agrupamiento
- Medidas mas usadas distancias en particular la distancia Euclideana.

# Algoritmo de k-medias

➤ El conjunto de datos  $X$  contiene  $k$  agrupamientos  $X_i$

que pueden representarse adecuadamente con su valor medio  $\mu_i$ .

**Medida de similitud:** distancia Euclídeana a  $\mu_i$

**Criterio de agrupamiento:** suma total de la distancia cuadrática de cada punto al vector medio de su agrupamiento.

➤ **Objetivo del algoritmo:** Encontrar entre todas las particiones de  $X$  en  $k$  conjuntos  $X_i ; i = 1, 2, \dots, k$  aquella que minimiza el criterio de agrupamiento.

# El algoritmo k-mean

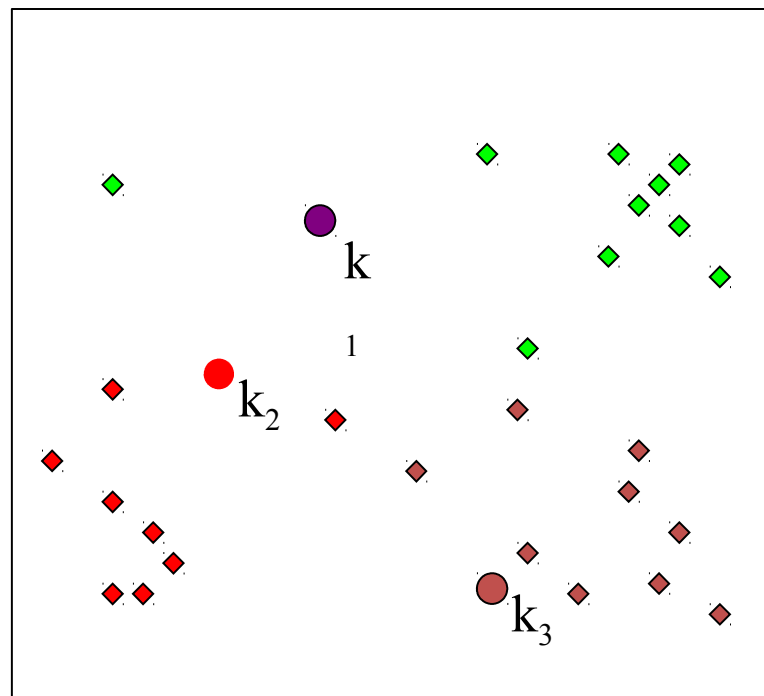
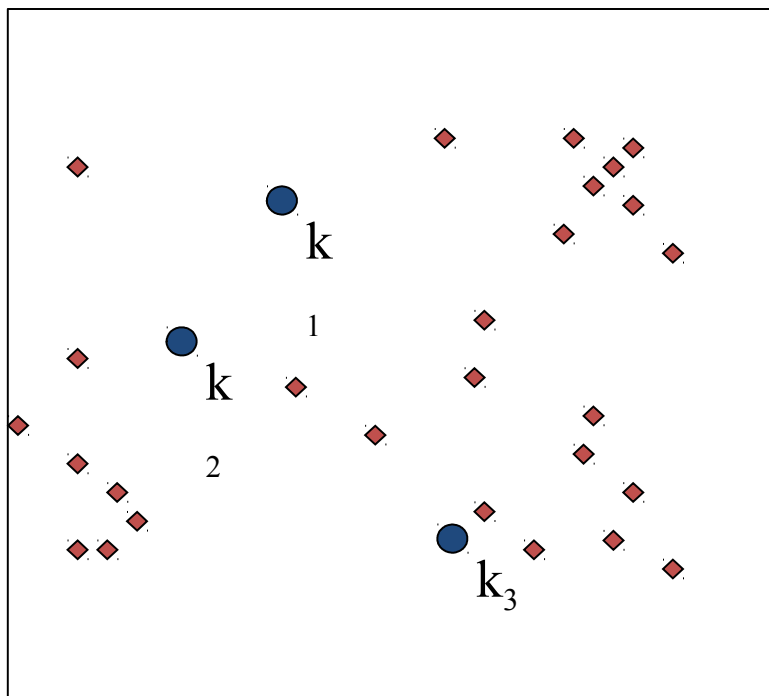
De lo anterior se deducen los pasos del algoritmo

**Paso 1:** Elegir una partición inicial en  $k$  grupos al azar

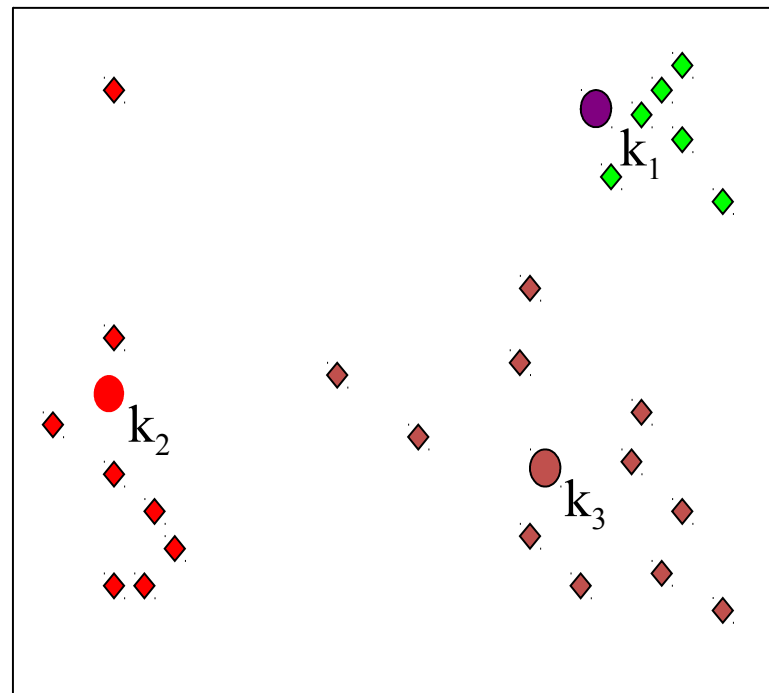
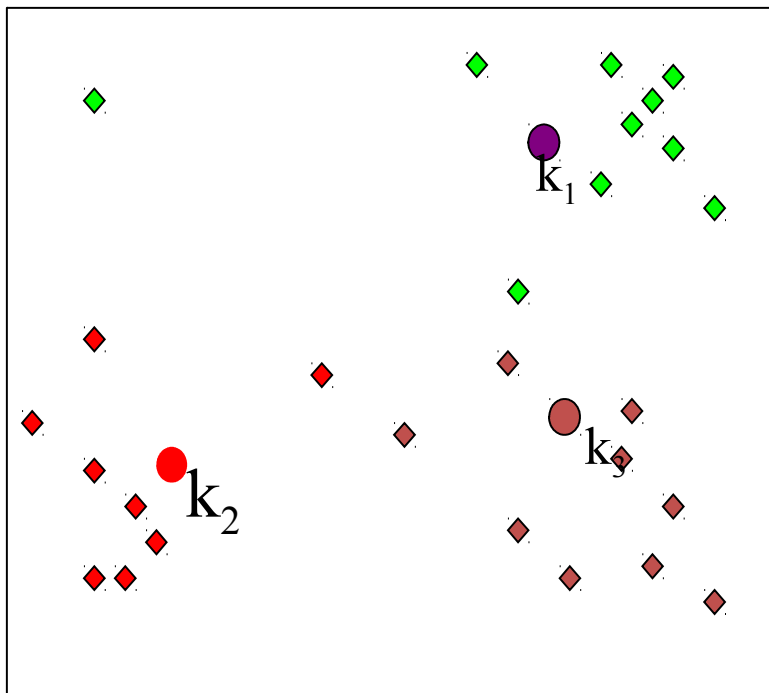
**Paso 2:** Calcular las medias  $\mu_i$  de los *clusters*

**Paso 3:** Seleccionar secuencialmente un punto  $\mathbf{x}$  del conjunto y, si corresponde, reasignarlo al *cluster* que minimiza  $d(\mathbf{x}, \mu_i)$

**Paso 4:** Si no hay reasignaciones en todo el conjunto terminar; sino volver al **Paso 2**.







# Métodos de Clasificación

Método	Propiedades	Comentarios
Correspondencia de modelos (Template Matching)	Asigna patrón a la clase del modelo más parecido	Modelo y métrica suministrada por el usuario. Dependiente de la métrica
Vecino más cercano	Asigna el patrón a la clase de la muestra más cercana	No necesita entrenamiento, desempeño robusto, lento dependiente de la métrica
Clasificador de Parzen	Regla de Bayes para estimación de densidades con ventanas e Parzen	Asintóticamente óptima, dependiente de la métrica, lento
Árboles de decisión binarios	Encuentra un conjunto de umbrales para una secuencia de atributos.	Entrenamiento iterativo, sensible al sobreentrenamiento
Perceptrón	Optimización iterativa de un clasificador lineal	Sensible al entrenamiento de los parámetros.

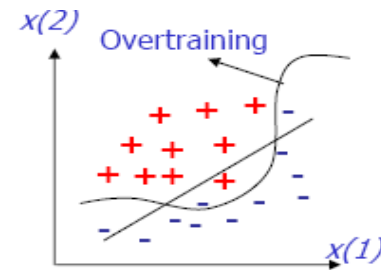
# **Evaluación de desempeño**

# Estimación de error y confianza

- Aplico el clasificador a un conjunto de test de patrones cuya clase es conocida  $Z_{ts}$
- Estimo el error contando discrepancias entre clase verdadera y etiqueta asignada por el clasificador
- $Error(D) = N_{error}/N_{ts}$  tasa de error aparente
- Necesitamos un número grande de muestras para verificar con confianza relativamente razonable.

# Conjunto de entrenamiento y de test

- Queremos usar la mayor cantidad de datos posibles para el entrenamiento y para la evaluación del desempeño del clasificador.



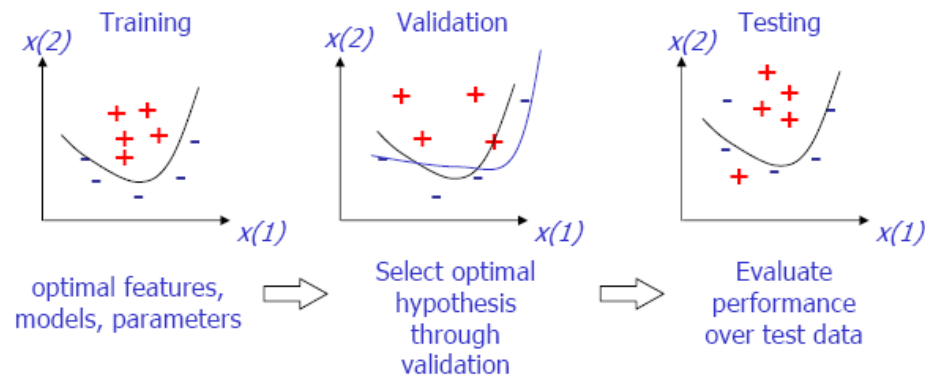
- Si usamos todos los datos para el entrenamiento y el mismo conjunto para la evaluación podemos **SOBREENTRENAR** el clasificador. Lo que puede hacer que falle con datos no vistos.
- **Reclasificación:** procedimiento de evaluación de desempeño usando el conjunto de aprendizaje.
- **Generalización:** Evaluación con un conjunto de test independiente.

# Utilización de los datos: entrenamiento vs testing

- **Resustitución (R-método):** Utilizo todo el conjunto de muestras para diseñar y testear. El error se estima en forma optimista.
- **Hold-out (H- metodo):** Separo  $Z$  en mitades una la utilizo para entrenar y la otra para estimar desempeño. Se pueden intercambiar los conjuntos y estimar el error como el promedio de los desempeños.
- **Validación cruzada:** Divido  $Z$  en  $k$  subconjuntos de tamaño  $n/k$ . Usamos un conjunto para testear desempeño del clasificador entrenado con los restantes  $k-1$ . Este procedimiento se repite  $k$  veces eligiendo conjunto diferentes.  $P_D$  se estima promediando las  $k$  estimaciones. Cuando  $k=N$  el método se denomina uno afuera (U-método)

# Utilización de los datos: entrenamiento vs testing

- **Bootstrap:** se utiliza para corregir la estimación optimista del R-método. Se hace generando  $L$  conjuntos de cardinalidad  $n$  a partir del conjunto  $Z$  con remplazo. Luego se promedia error de clasificación de los conjuntos.
- **Hold out:** Utilizo 3 conjuntos: entrenamiento, validación y test. Se continua el entrenamiento hasta que no se logra mejora con el conjunto de validación.



# Matrices de Confusión

- Para determinar como se distribuyen los errores en las clases se construyen las matrices de confusión usando el conjunto de test  $Z_{ts}$ .
- La entrada  $a_{ij}$  de cada matriz indica el número de elementos de  $Z_{ts}$  cuya clase cierta es  $w_i$  y que se le asignó clase  $w_j$



# Matrices de Confusión- Clasificación de huellas dactilares

Mean filter performance				
	Arch	Left loop	Right loop	Whorl
Arch	73 (71.57%)	21 (2.41%)	12 (1.57%)	10 (0.99%)
Left loop	7 (6.86%)	816 (93.69%)	20 (2.62%)	69 (6.83%)
Right loop	21 (20.59%)	21 (2.41%)	706 (92.41%)	90 (8.91%)
Whorl	1 (0.98%)	13 (1.49%)	26 (3.40%)	841 (83.27%)
Total	102 (100%)	871 (100%)	764 (100%)	1010(100%)

Anisotropic diffusion performance				
	Arch	Left loop	Right loop	Whorl
Arch	108 (84.38%)	39 (4.16%)	8 (0.94%)	3 (0.34%)
Left loop	9 (7.03%)	865 (92.32%)	15 (1.75%)	20 (2.28%)
Right loop	11 (8.59%)	16 (1.71%)	811 (94.85%)	22 (2.51%)
Whorl	0 (0%)	17 (1.81%)	21 (2.46%)	831 (94.86%)
Total	128 (100%)	937 (100%)	855 (100%)	876 (100%)

# Comparación de desempeño

Mean filter performance	
Correct classification	2436 (87.00%)
Wrong classification	311 (11.11%)
Not classified	53 (1.89%)
Total	2800 (100%)

Anisotropic diffusion performance	
Correct classification	2615 (93.39%)
Wrong classification	181 (6.46%)
Not classified	4 (0.14%)
Total	2800 (100%)

# Evaluación de Desempeño

- En una aplicación de control de acceso o de verificación (confrontación) existen dos clases:

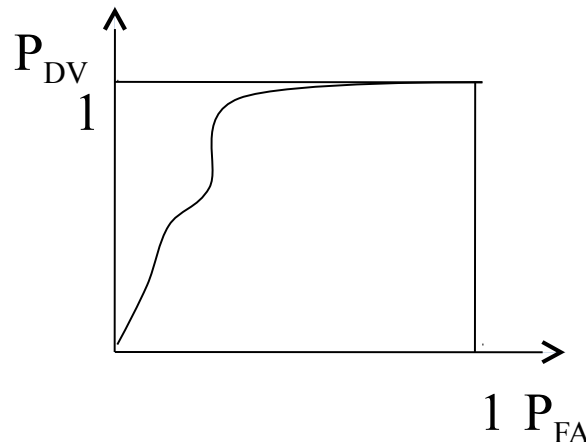
$w_1$  la identidad es auténtica o

$w_2$  la identidad es falsa

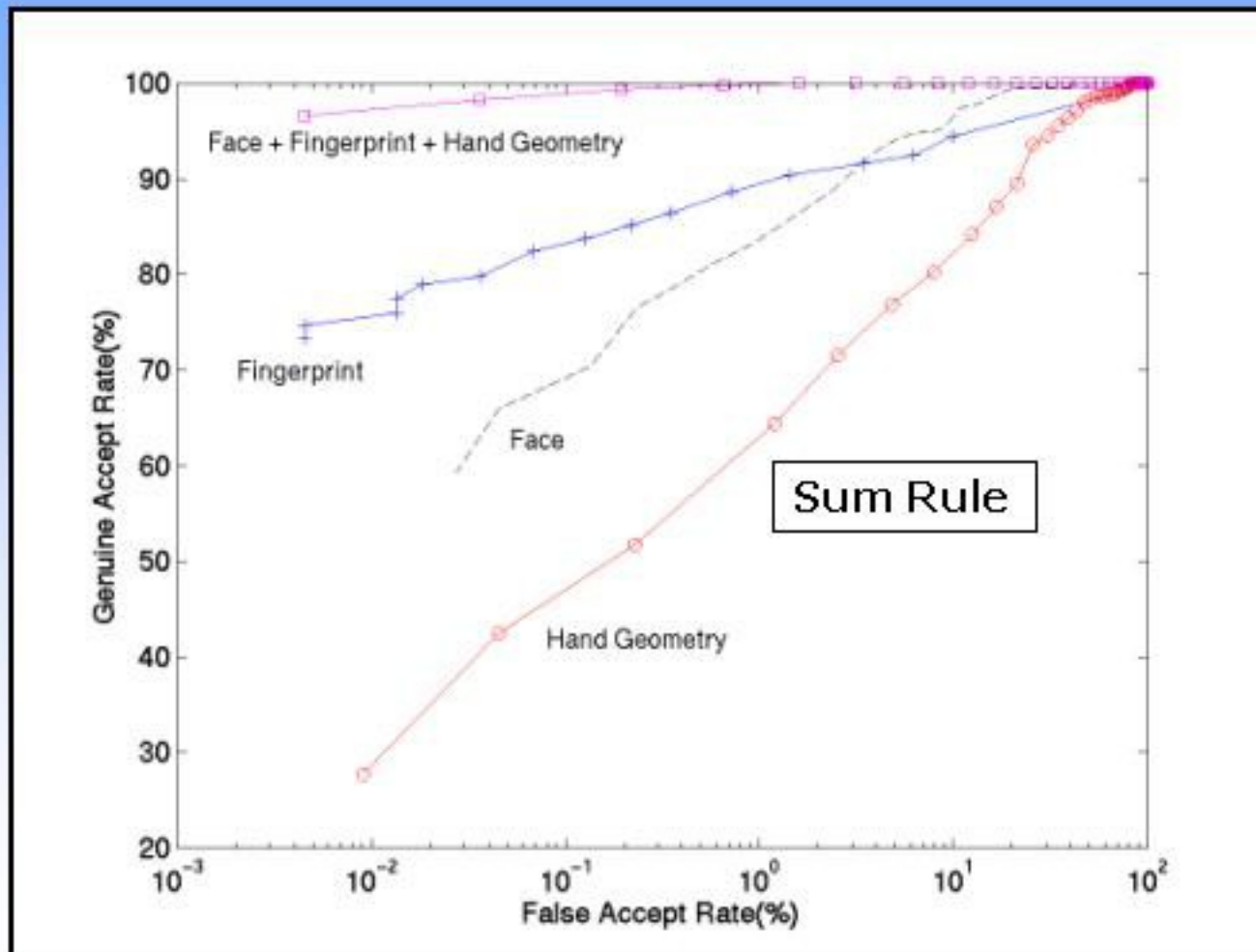
- En este contexto podemos cometer dos errores de clasificación que tienen dos costos muy distintos, minimizar el error promedio no es lo adecuado-

# Receiver Operating Characteristic

- Asigna a cada decisión errónea un costo y se busca **minimizar el costo total**.
- Otra opción es analizar el comportamiento de las falsas aceptaciones contra los falsos rechazos al variar el comportamiento del clasificador y determinar el punto de trabajo utilizando la **ROC (receiver Operating Characteristic)**, por ejemplo de forma que las falsas aceptaciones no superen un valor objetivo.



# Multibiometric System Performance

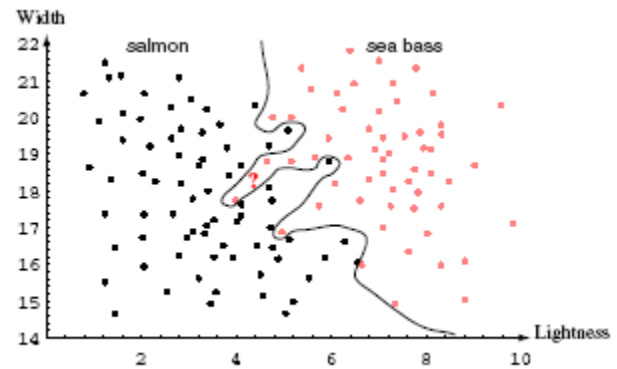
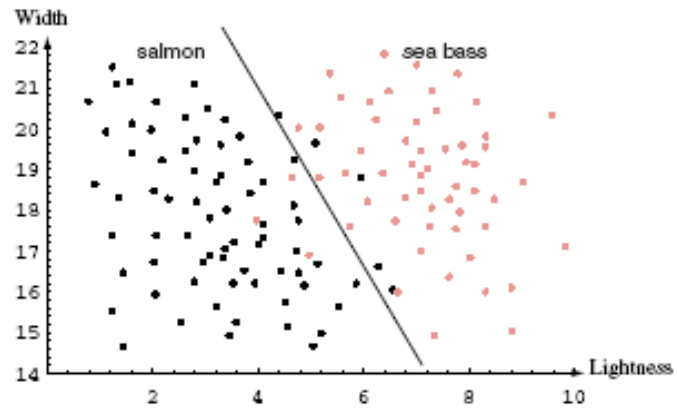
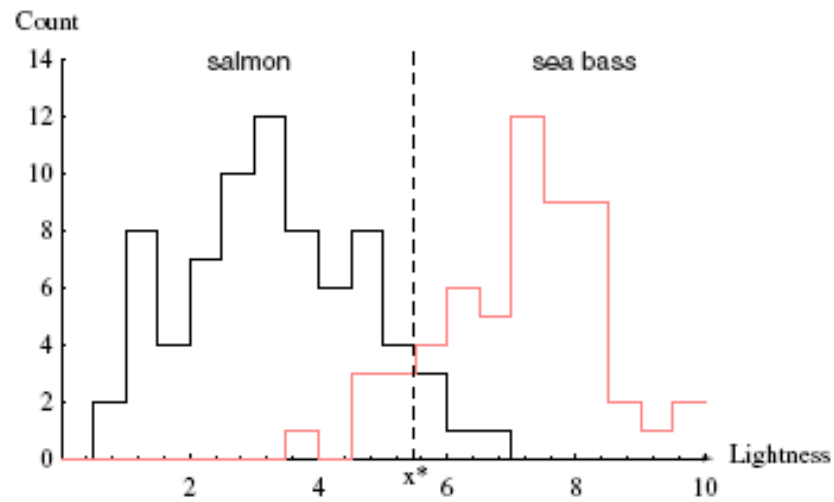


# Ejemplo de Problema Real

- Una planta procesadora de pescado quiere automatizar el proceso de clasificación de pescado con respecto a la especie (salmón o ródalo).
- El sistema automatizado consiste en:
  - Una cinta transportadora para los productos recibidos
  - Dos cintas transportadoras para los productos clasificados
  - Un brazo robótico para tomar y colocar objetos
  - Un sistema de visión con una cámara
  - Una computadora para clasificar las imágenes y controlar el robot

# Ejemplo de Problema Real

- **Sensado:** El sistema de visión captura una imagen en cuanto un nuevo pescado entra en el área de clasificación.
- **Preprocesado** – Algoritmos de Procesamiento de Imágenes
  - Ajuste de niveles de intensidad
  - Segmentación para separar el pescado del fondo de la imagen
- **Extracción de características** – Supongamos que sabemos que en media, el ródalo es más largo que el salmón
  - A partir de la imagen segmentada estimamos la longitud del pescado
- **Clasificación**
  - Seleccionar un conjunto de muestras de ambas especies.
  - Calcular la distribución de longitudes para ambas clases.
  - Determinar la frontera de decisión (umbral) que minimiza el error de clasificación.
  - Estimamos la probabilidad de error y se obtiene un mal resultado del orden del 40%
- ¿Qué hacemos ahora?





# Ejemplo de Problema Real

- **Mejora del Desempeño:** Para obtener un error inferior al 5%, probamos con nuevas características:
  - Anchura, área, posición de los ojos respecto a la cara,...
  - Finalmente encontramos una “buena” característica: Intensidad media media de las escamas .
  - Combinamos: “longitud” e “intensidad media de las escamas” para mejorar Longitud la separabilidad de las clases
  - Buscamos un clasificador que proporcione una frontera de decisión lineal (clasificador lineal) y obtenemos un 4.3% de error.

# Ejemplo de Problema Real

## ➤ Costo y Error de Clasificación

- El clasificador que se diseñó hace mínimo el error de clasificación.
- ¿Es éste el mejor criterio para procesar pescado ?
- El costo de clasificar erróneamente salmón como ródalo es que el consumidor encontrará una pieza sabrosa de salmón cuando compra ródalo.
- El costo de clasificar erróneamente ródalo como salmón es que el consumidor encontrará una pieza ródalo comprada al precio de salmón
- Deberíamos ajustar la frontera de decisión para minimizar una función de costo que incluya diferentes costos de confusión.

# Resumen

- Los objetivos del Reconocimiento de Patrones está relacionados con la elección del algoritmo más apropiado para el problema a resolver.
- Esto requiere conocimiento a priori (distribución de los datos, probabilidades a priori, complejidad del problema, física del fenómeno que generó los datos, etc).
- En ausencia de conocimiento a priori no hay ningún clasificador mejor que otro. Sin embargo con información a priori, algunos clasificadores funcionan mejor con determinados tipos de problemas.
- El reto es entonces identificar el clasificador o la combinación adecuada para el problema a resolver.

# Proyectos de Investigación recientes:

- Detección de melanomas
  - Detección de focos epilépticos
  - Detección de pólipos en colonoscopia virtual
  - Clasificación de tráfico en redes de datos
  - Detección de fraudes en consumos de energía
  - Reconocimiento de Caras
  - Reconocimiento de Melodías
-