

Bases estadísticas del reconocimiento de patrones

César Martínez
cmartinez_AT_fich.unl.edu.ar

Inteligencia Computacional
FICH-UNL



Tarea muuuuy simple: ¿Cuántas llaves hay?



Percepción humana

Tarea simple: ¿Cuántas llaves hay?



Tarea no tan simple: ¿Cuántas llaves hay?



Percepción humana

Tarea compleja: ¿Cuántas llaves hay?



Aprendizaje maquina

Problema: ¿Cómo enseñarle a una máquina a que emule las tareas humanas de detección, percepción, identificación, reconocimiento, etc.?



Objetivo: Programar máquinas para

-

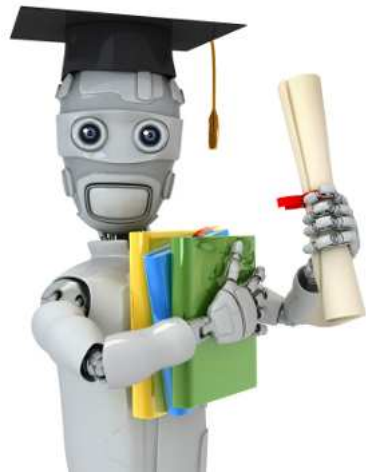
Aprendizaje maquina

Objetivo: Programar máquinas para

- aprender a realizar una *tarea*,
- mejorando su *desempeño*
- basado en la *experiencia*.

Restricciones adicionales:

- mínima intervención humana posible,
- con reducido conjunto de ejemplos,
- etc...



Conceptos básicos de Aprendizaje Maquina

Contexto de la disciplina

Podemos considerar dos ramas de la Inteligencia Artificial (IA):

- IA clásica:

- Modelado del proceso de **razonamiento** humano mediante técnicas de la *Lógica*.
- Aprendizaje deductivo.

- Reconocimiento de Formas:

- Modelado del proceso de **percepción** humano mediante técnicas de la *Teoría de la Decisión Estadística* y de la *Teoría de los Lenguajes Formales*.
- Aprendizaje inductivo.

Contexto de la disciplina

Podemos considerar dos ramas de la Inteligencia Artificial (IA):

- IA clásica:
 - Modelado del proceso de **razonamiento** humano mediante técnicas de la *Lógica*.
 - Aprendizaje deductivo.
- Reconocimiento de Formas:
 - Modelado del proceso de **percepción** humano mediante técnicas de la *Teoría de la Decisión Estadística* y de la *Teoría de los Lenguajes Formales*.
 - Aprendizaje inductivo.

Contexto de la disciplina

Vista como una disciplina de la IA...

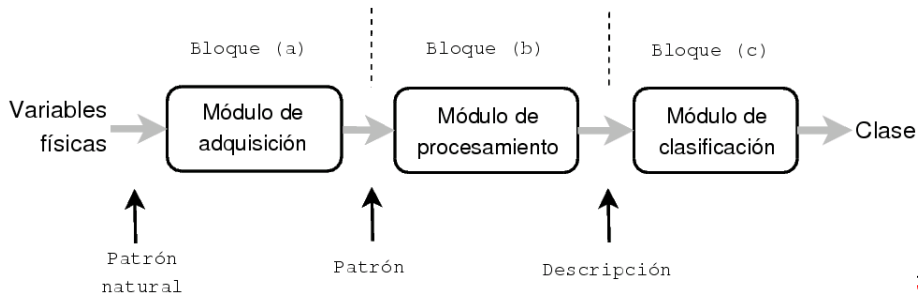
- *Adquisición y representación del conocimiento*: conformación y almacenamiento de conjuntos de patrones o prototipos.
- *Aprendizaje*: algoritmos de aprendizaje inductivo a partir de un conjunto de entrenamiento.
- *Clasificación*: etiquetado de patrones nuevos utilizando el conjunto de clases disponible.
- *Evaluación*: mecanismos para evaluar la bondad, confianza o error del sistema.

Definiciones

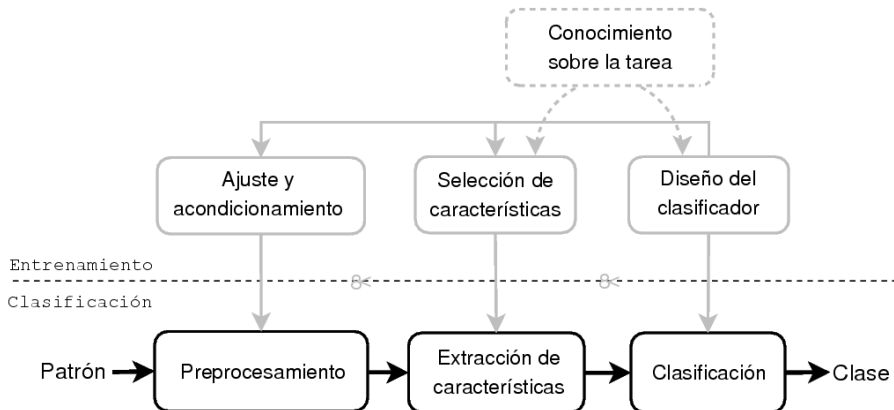
- Patrón:
 - Objeto de interés que es identificable del resto.
 - Posiblemente difusos, no bien definidos, no visibles o tangibles.
 - Ej: una huella digital, la voz de una persona, una cara, etc.
- Reconocimiento de patrones (RP):
 - Estudio de los procesos de percepción y razonamiento humanos:
 - capacidad de distinguir y aislar los patrones,
 - reunirlos en grupos,
 - asignarles un nombre identificador a cada grupo.
 - Objetivo del RP: crear sistemas informáticos que imiten el comportamiento descripto.

Paradigma de trabajo conceptual

- (a) Adquisición: transducción del mundo real a la representación digital.
- (b) Procesamiento digital: acondicionamiento y representación alternativa.
- (c) Clasificación: decisión sobre la clase.



Paradigma de trabajo funcional



Aproximaciones

- Aproximación geométrica o *estadística*:

- Basada en la Teoría Estadística de la Decisión.
- Representación de patrones como *vectores numéricos*.
- Representación de clases mediante *patrones prototipo*.
- Tipos de clasificadores: gaussianos, basados en distancia, etc.

- Aproximación estructural o *sintáctica*:

- Basada en la Teoría de Lenguajes Formales.
- Representación de patrones como *cadena de símbolos*.
- Utilización de *reglas sintácticas* para especificar los patrones válidos de una clase.
- Tipos de clasificadores: autómatas, gramáticas, HMM, etc.

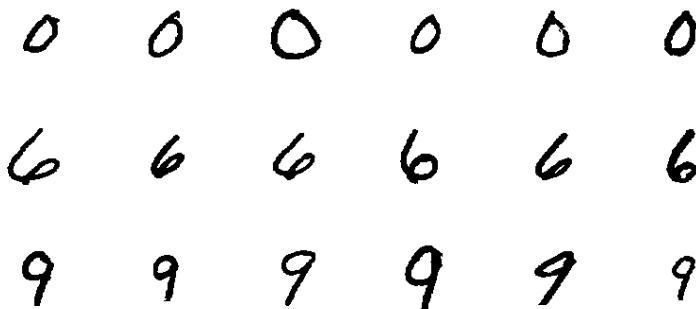
Aproximaciones

- Aproximación geométrica o *estadística*:
 - Basada en la Teoría Estadística de la Decisión.
 - Representación de patrones como *vectores numéricos*.
 - Representación de clases mediante *patrones prototipo*.
 - Tipos de clasificadores: gaussianos, basados en distancia, etc.

- Aproximación estructural o *sintáctica*:
 - Basada en la Teoría de Lenguajes Formales.
 - Representación de patrones como *cadena de símbolos*.
 - Utilización de *reglas sintácticas* para especificar los patrones válidos de una clase.
 - Tipos de clasificadores: autómatas, gramáticas, HMM, etc.

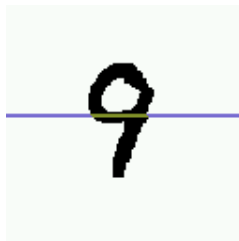
Ejemplo de aproximación geométrica

OCR de dígitos manuscritos:



Ejemplo de aproximación geométrica

OCR de dígitos manuscritos:

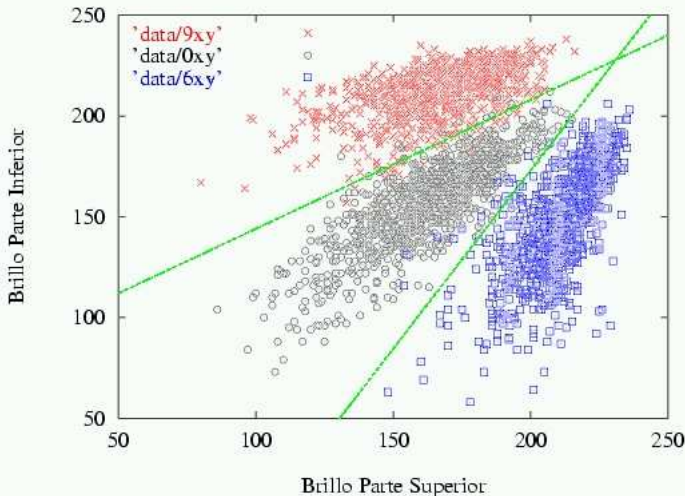


Característica simple: cálculo de brillo de partes superior e inferior

Ejemplo de aproximación geométrica

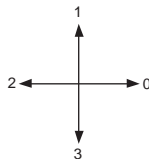
Clasificador geométrico con funciones lineales:

0
6
9

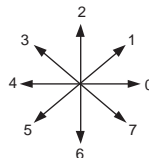


Ejemplo de aproximación sintáctica

Descripción de dígitos mediante cadenas de contorno:



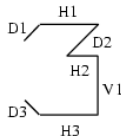
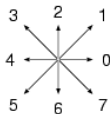
000...03332323...22111



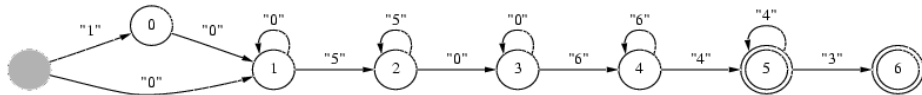
000...066655...1122244...222

Ejemplo de aproximación sintáctica

Clasificador sintáctico basado en autómatas:



S	-->	D1	H1	D2	H2	V1	H3	D3
D1	->	"1"		<nil>				
H1	->	"0"		"0"	H1			
D2	->	"5"		"5"	D2			
H2	->	"0"		"0"	H2			
V1	->	"6"		"6"	V1			
H3	->	"4"		"4"	H3			
D3	->	"3"		<nil>				



Conceptos de clasificación estadística

Definiciones

- Adquisición y preproceso de datos: vector numérico que representa al patrón natural.

Formalmente, un patrón es una variable aleatoria n -dimensional

$\mathbf{y} = [y_1 \ y_2 \ \dots \ y_n]^T$, con $y_i \in \mathbb{R}$ para $i = 1, 2, \dots, n$, que representa un punto en el espacio de patrones $P \in \mathbb{R}^n$.

Definiciones

- Adquisición y preproceso de datos: vector numérico que representa al patrón natural.

Formalmente, un patrón es una variable aleatoria n -dimensional

$\mathbf{y} = [y_1 \ y_2 \ \dots \ y_n]^T$, con $y_i \in \mathbb{R}$ para $i = 1, 2, \dots, n$, que representa un punto en el espacio de patrones $P \in \mathbb{R}^n$.

- Extracción de características: información *relevante* para la clasificación.

Formalmente, dado un conjunto de patrones n -dimensionales \mathbf{y} se trata de obtener un nuevo conjunto de representaciones

$\mathbf{x} = [x_1 \ x_2 \ \dots \ x_d]^T$, donde $d \leq n$.

- Cambio en el espacio de representación: transformaciones lineales que maximizan la varianza ($d = n$).
 - Reducción de dimensionalidad de los datos ($d < n$).
- Espacio de características: $E \in \mathbb{R}^d$.

Definiciones

- Extracción de características: propiedades deseables
 - Precisión: representaciones diferentes para objetos diferentes.
 - Unicidad o determinismo: representación única para cada objeto.
 - Continuidad en el espacio: inmunidad al ruido y capacidad de generalización.
- Clases informacionales: salidas del sistema
 - Número de clases: c
 - Conjunto de (etiquetas de) clases: $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$
 - Conjunto extendido: $\Omega^* = \{\omega_1, \omega_2, \dots, \omega_c, \omega_0\}$, donde ω_0 es la *clase de rechazo*.

Definiciones

- Extracción de características: propiedades deseables
 - Precisión: representaciones diferentes para objetos diferentes.
 - Unicidad o determinismo: representación única para cada objeto.
 - Continuidad en el espacio: inmunidad al ruido y capacidad de generalización.
- Clases informacionales: salidas del sistema
 - Número de clases: c
 - Conjunto de (etiquetas de) clases: $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$
 - Conjunto extendido: $\Omega^* = \{\omega_1, \omega_2, \dots, \omega_c, \omega_0\}$, donde ω_0 es la *clase de rechazo*.

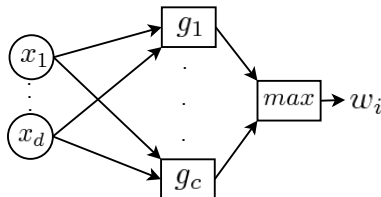
Definiciones

- Clasificador estadístico: máquina formada por c (funciones) discriminantes

$$g_i : E \rightarrow \mathbb{R}, 1 \leq i \leq c$$

tal que dado un patrón $\mathbf{x} \in E$,

\mathbf{x} se asigna a la clase ω_i si $g_i(\mathbf{x}) > g_j(\mathbf{x}) \forall j \neq i$



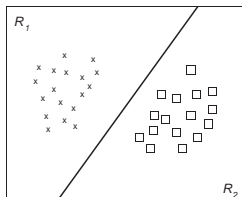
Regiones y fronteras de decisión

- Regiones de decisión: un clasificador divide el espacio en c regiones de decisión R_1, R_2, \dots, R_c , tal que

$$R_i = \{\mathbf{x} \in E : g_i(\mathbf{x}) > g_j(\mathbf{x}) \ \forall j \neq i\}$$

- Fronteras de decisión: superficies del espacio que separan regiones de decisión contiguas.

Hipersuperficies definidas por: $g_i(\mathbf{x}) - g_j(\mathbf{x}) = 0, \ i \neq j, \ 1 \leq i, j \leq c$



Clasificadores estadísticos básicos

Las FD son combinaciones lineales o cuadráticas de las componentes del vector de características.

- Clasificador lineal:

$$g(\mathbf{x}) = \sum_{i=1}^d w_i x_i + w_0 = \mathbf{w}^t \mathbf{x} + w_0$$

Parámetros: $d + 1$. Fronteras: hiperplanos.

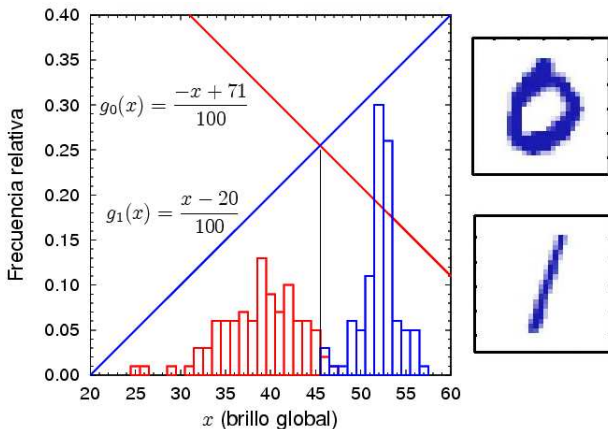
- Clasificador cuadrático:

$$g(\mathbf{x}) = \sum_{i=1}^d \sum_{j=1}^d w_{ij} x_i x_j + \sum_{i=1}^d w_i x_i + w_0 = \mathbf{x}^t W \mathbf{x} + \mathbf{w}^t \mathbf{x} + w_0$$

Parámetros: $\frac{1}{2}d(d + 1) + d + 1$. Fronteras: hipercuádricas.

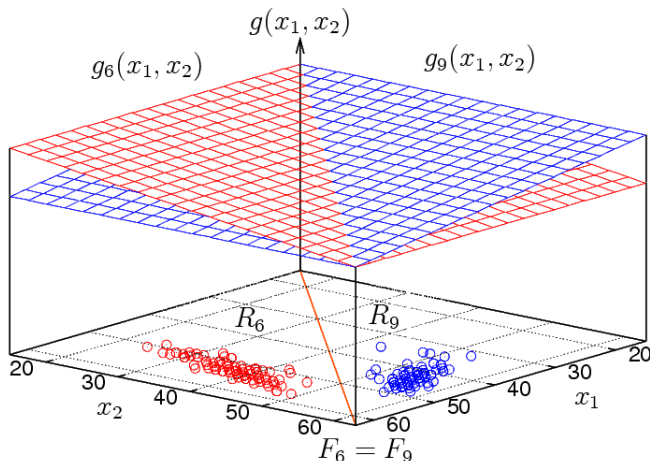
Ejemplo: OCR de dígitos manuscritos

- Unidimensional: brillo global.



Ejemplo: OCR de dígitos manuscritos

- Bidimensional: brillo de mitad superior (x_1) e inferior (x_2).



Definiciones de teoría de la decisión

- **Probabilidad a priori** $P(\omega_i)$ de una clase ω_i : probabilidad de que una muestra arbitraria pertenezca a ω_i .

“Probabilidad de observar la etiqueta c sin saber qué muestra es”

Puede verse como la proporción de muestras de ω_i respecto al total (conocimiento que se tiene antes de hacer los experimentos).

Condiciones:

- $0 \leq P(\omega_i) \leq 1$, para $i = 1, \dots, c$.
- $\sum_{i=1}^c P(\omega_i) = 1$.

Clasificador trivial de dos clases basado en probabilidades a priori:

- Decidir por ω_1 si $P(\omega_1) > P(\omega_2)$
- Decidir por ω_2 si $P(\omega_1) < P(\omega_2)$

Definiciones de teoría de la decisión

- **Densidad condicional** $P(\mathbf{x}|\omega_i)$ de una clase ω_i : función de densidad de probabilidad que caracteriza la distribución estadística de las muestras de ω_i .

“Probabilidad de observar la muestra \mathbf{x} sabiendo que la etiqueta es c ”

El vector de características \mathbf{x} se considera una variable aleatoria d -dimensional de función de densidad $P(\mathbf{x}|\omega_i)$ cuando las muestras pertenecen a ω_i .

Condiciones:

- $P(\mathbf{x}|\omega_i) \geq 0$, para $i = 1, \dots, c$.
- $\int_E P(\mathbf{x}|\omega_i) d\mathbf{x} = 1$.

Definiciones de teoría de la decisión

- La **probabilidad conjunta** $P(\mathbf{x}, \omega_i)$ muestra-clase se define como

$$P(\mathbf{x}, \omega_i) = P(\omega_i)P(\mathbf{x}|\omega_i)$$

“Probabilidad de observar la muestra \mathbf{x} con la etiqueta c ”

- La **densidad incondicional** $P(\mathbf{x})$ de las muestras se define como

$$P(\mathbf{x}) = \sum_{j=1}^c P(\mathbf{x}, \omega_j) = \sum_{j=1}^c P(\mathbf{x}|\omega_j)P(\omega_j)$$

“Probabilidad de observar la muestra \mathbf{x} sin saber cuál es su etiqueta”

La densidad incondicional caracteriza la distribución estadística de las muestras con independencia de las clases a las que pertenecen.

Se cumple que: $P(\mathbf{x}) \geq 0$, y $\int_E P(\mathbf{x}) d\mathbf{x} = 1$.

Definiciones de teoría de la decisión

- La **probabilidad conjunta** $P(\mathbf{x}, \omega_i)$ muestra-clase se define como

$$P(\mathbf{x}, \omega_i) = P(\omega_i)P(\mathbf{x}|\omega_i)$$

“Probabilidad de observar la muestra \mathbf{x} con la etiqueta c ”

- La **densidad incondicional** $P(\mathbf{x})$ de las muestras se define como

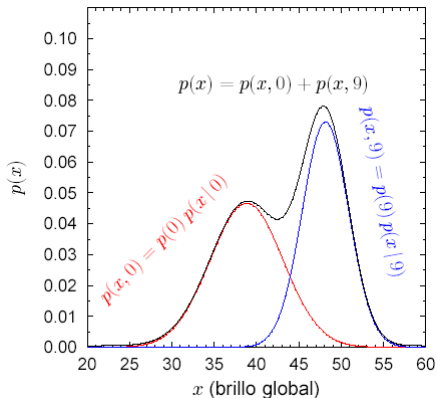
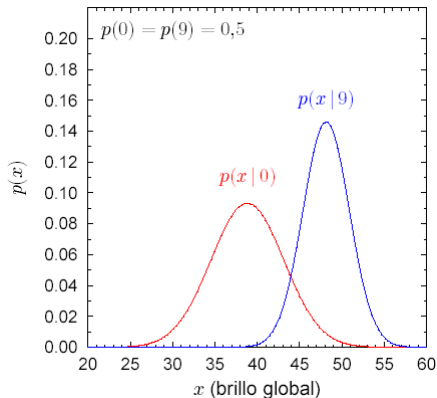
$$P(\mathbf{x}) = \sum_{j=1}^c P(\mathbf{x}, \omega_j) = \sum_{j=1}^c P(\mathbf{x}|\omega_j)P(\omega_j)$$

“Probabilidad de observar la muestra \mathbf{x} sin saber cuál es su etiqueta”

La densidad incondicional caracteriza la distribución estadística de las muestras con independencia de las clases a las que pertenecen.

Se cumple que: $P(\mathbf{x}) \geq 0$, y $\int_E P(\mathbf{x}) d\mathbf{x} = 1$.

Definiciones de teoría de la decisión



Significado de: $P(\omega = 0) = 0,5$; $P(x = 45|\omega = 0) = 0,033$;
 $P(x = 45, \omega = 0) = 0,016$; $P(x = 45) = 0,054$.

Definiciones de teoría de la decisión

Supongamos que tenemos conocidas $P(\omega_i)$, $P(\mathbf{x}|\omega_i)$ y una medida de \mathbf{x} .
¿Cómo influye este conocimiento sobre la decisión del clasificador?

- **Probabilidad a posteriori** $P(\omega_i|\mathbf{x})$ de una clase ω_i : probabilidad de que una muestra particular \mathbf{x} pertenezca a ω_i .

"Probabilidad de observar la etiqueta c sabiendo que la muestra es \mathbf{x} "

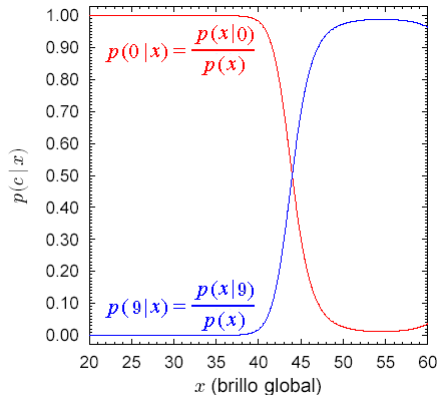
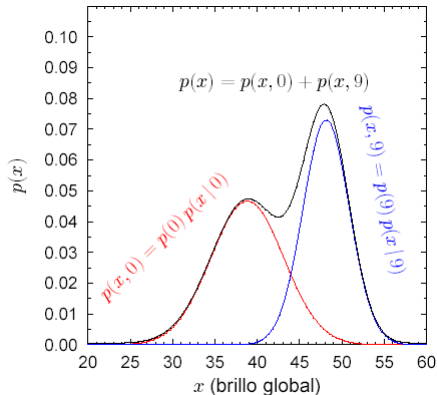
Se calcula mediante la regla de Bayes:

$$P(\omega_i|\mathbf{x}) = \frac{P(\mathbf{x}|\omega_i)P(\omega_i)}{P(\mathbf{x})} = \frac{P(\mathbf{x}|\omega_i)P(\omega_i)}{\sum_{j=1}^c P(\mathbf{x}|\omega_j)P(\omega_j)}$$

Se cumple que: $0 \leq P(\omega_i|\mathbf{x}) \leq 1$, y $\sum_{i=1}^c P(\omega_i|\mathbf{x}) = 1$.

Puede verse como una actualización de $P(\omega_i)$ después de observar \mathbf{x} .

Definiciones de teoría de la decisión



Idea para un clasificador...

Clasificador de Bayes

- Regla de clasificación de Bayes: asignar a \mathbf{x} la clase con mayor probabilidad a posteriori.

$$\hat{\omega} = \arg \max_{\omega_i: 1 \leq i \leq c} P(\omega_i | \mathbf{x})$$

- Clasificador de Bayes:** utiliza las probabilidades a posteriori como funciones discriminantes. Es de mínimo error basado en FD monótonas crecientes $f(\cdot)$.

$$g_i(\mathbf{x}) = P(\omega_i | \mathbf{x}) = \frac{P(\mathbf{x} | \omega_i) P(\omega_i)}{P(\mathbf{x})}$$

$$\equiv P(\mathbf{x} | \omega_i) P(\omega_i)$$

$$\equiv \log P(\mathbf{x} | \omega_i) + \log P(\omega_i)$$

Equivalencia entre clasificadores:

$$(g_1, \dots, g_c) \equiv (f(g_1), \dots, f(g_c)) \equiv (g'_1, \dots, g'_c) \text{ si } R_c = R'_c \forall c.$$

Clasificador de Bayes

- Error de Bayes a posteriori: probabilidad de error puntual dado por

$$P(\text{error}|\mathbf{x}) = 1 - \max_{1 \leq i \leq c} P(\omega_i|\mathbf{x})$$

- Probabilidad media de error del clasificador:

$$P(\text{error}) = \int_E P(\text{error}|\mathbf{x})P(\mathbf{x})d\mathbf{x}$$

- Problemas con el cálculo analítico del error:
 - $P(\omega_i)$ y $P(\mathbf{x}|\omega_i)$ desconocidas.
 - Regiones de integración complejas.
- Solución: *estimación* del error.

Clasificador de Bayes

- Error de Bayes a posteriori: probabilidad de error puntual dado por

$$P(\text{error}|\mathbf{x}) = 1 - \max_{1 \leq i \leq c} P(\omega_i|\mathbf{x})$$

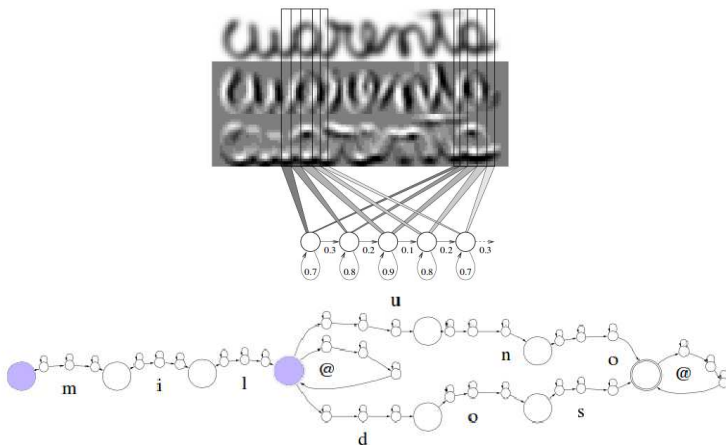
- Probabilidad media de error del clasificador:

$$P(\text{error}) = \int_E P(\text{error}|\mathbf{x})P(\mathbf{x})d\mathbf{x}$$

- Problemas con el cálculo analítico del error:
 - $P(\omega_i)$ y $P(\mathbf{x}|\omega_i)$ desconocidas.
 - Regiones de integración complejas.
- Solución: *estimación* del error.

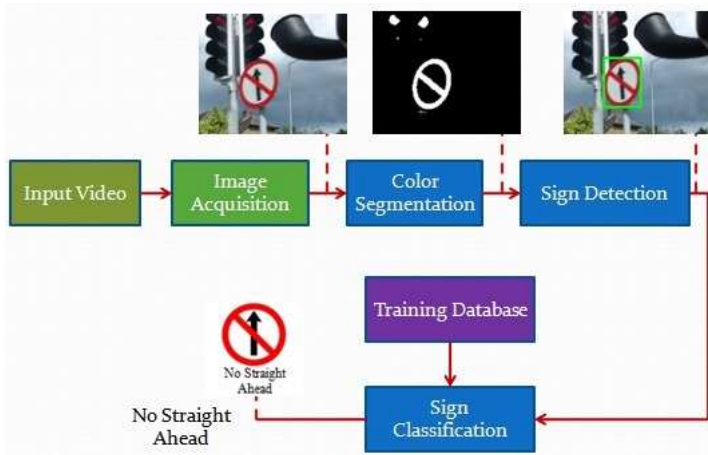
Aplicaciones del RP

OCR manuscrito:



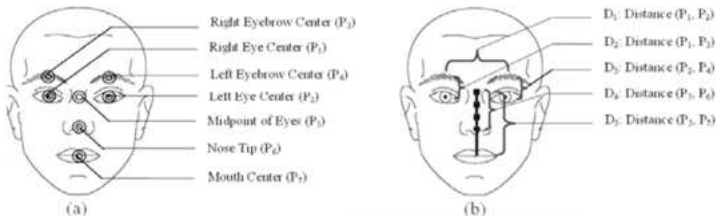
Aplicaciones del RP

Reconocimiento de señales de tráfico:



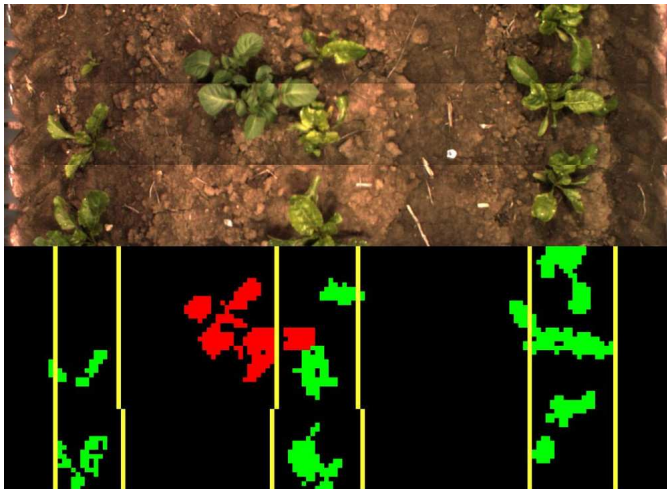
Aplicaciones del RP

Reconicimiento facial, expresiones y emociones:



Aplicaciones del RP

Detección de malezas en aplicaciones agrícolas:



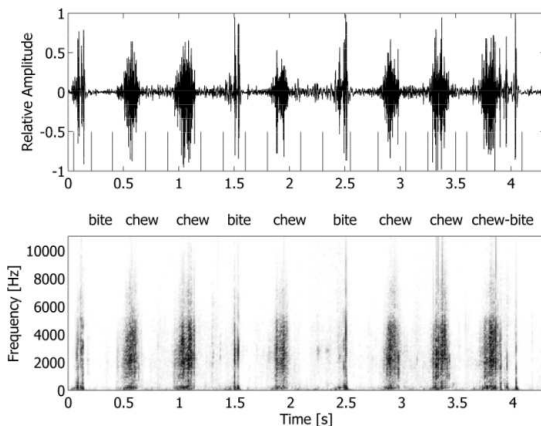
Aplicaciones del RP

Reconocimiento de sonidos masticatorios en rumiantes:



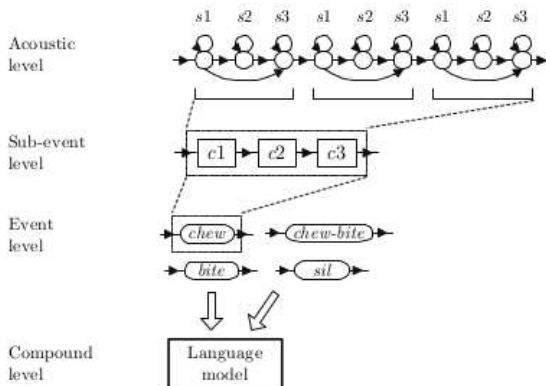
Aplicaciones del RP

Reconocimiento de sonidos masticatorios en rumiantes:



Aplicaciones del RP

Reconocimiento de sonidos masticatorios en rumiantes:



Aplicaciones del RP

Reconocimiento de Imágenes

OCR - Análisis de documentos - Rec. de firmas - Identif. de patentes -
Detección de defectos en piezas industriales (control de calidad) - Rec.
de gestos faciales

Reconocimiento de Habla y Lenguaje

Rec. de palabras aisladas y discurso continuo - Identif. del locutor -
Traducción automática - Comprensión

Aplicaciones Biomédicas

Segmentación de tejido en imágenes médicas - Bioidentificación
(huellas dactilares, palma, iris, otros)

Economía

Minería de datos - Detección de patrones de fraude

Aplicaciones en astronomía, agricultura, protección civil, etc.

Fin

Bibliografía:

- Duda R., Hart P, Stork D., Pattern Classification, Second Edition. Wiley-Interscience, 2000. Capítulos 1, 2.1 a 2.6, 3.1 a 3.2.
- Lista de la cátedra