

# Reconocimiento de Escritura Continua Off-Line

## Reconocimiento de Escritura Manuscrita Off-Line

Alejandro H. Toselli

Departamento de Sistemas Informáticos y Computación  
Universidad Politécnica de Valencia



18 de febrero de 2008

## 1 Reconocimiento de Escritura Off-Line

- Introducción
- Visión Global del Sistema
- Preproceso de Texto Manuscrito - Página
- Eliminación del Fondo y Filtrado del Ruido
- Corrección de “Skew”
- Detección y Segmentado de Líneas
- Preproceso de Texto Manuscrito - Línea
- Corrección y Normalización de Estilos - Ilustración
- Corrección de “Slope”
- Corrección de “Slant”
- Normalización del Tamaño
- Extracción de Características
- Extracción de Características Utilizadas
- Reducción de la Dimensionalidad
- Entrenamiento de HMMs - Generalidades

# Temario

## 1 Reconocimiento de Escritura Off-Line

### ● Introducción

- Visión Global del Sistema
- Preproceso de Texto Manuscrito - Página
- Eliminación del Fondo y Filtrado del Ruido
- Corrección de “Skew”
- Detección y Segmentado de Líneas
- Preproceso de Texto Manuscrito - Línea
- Corrección y Normalización de Estilos - Ilustración
- Corrección de “Slope”
- Corrección de “Slant”
- Normalización del Tamaño
- Extracción de Características
- Extracción de Características Utilizadas
- Reducción de la Dimensionalidad
- Entrenamiento de HMMs - Generalidades

# Introducción

## Objetivo

*El reconocimiento de texto manuscrito off-line, aplicando tecnología actual de reconocimiento del habla.*

## El reconocimiento de texto manuscrito *off-line*

- *Tarea de gran dificultad.*
- *Presenta aspectos comunes con Reconocimiento del Habla.*

# Introducción

## Objetivo

*El reconocimiento de texto manuscrito off-line, aplicando tecnología actual de reconocimiento del habla.*

## El reconocimiento de texto manuscrito *off-line*

- *Tarea de gran dificultad.*
- *Presenta aspectos comunes con Reconocimiento del Habla.*

# Introducción

## La tecnología actual de reconocimiento del habla

- *Basada en los Modelos ocultos de Markov (HMMs).*
- *Técnicas basadas en la “no segmentación explícita” de las imágenes.*
- *Modelado de los niveles de percepción: morfológico, léxico y sintáctico.*
- *Integración de estos niveles en la fase de reconocimiento.*
- *Utilización de MEFs: HMMs, gramáticas o autómatas.*

# Introducción

## La tecnología actual de reconocimiento del habla

- *Basada en los Modelos ocultos de Markov (HMMs).*
- *Técnicas basadas en la “no segmentación explícita” de las imágenes.*
- *Modelado de los niveles de percepción: morfológico, léxico y sintáctico.*
- *Integración de estos niveles en la fase de reconocimiento.*
- *Utilización de MEFs: HMMs, gramáticas o autómatas.*

# Introducción

## La tecnología actual de reconocimiento del habla

- *Basada en los Modelos ocultos de Markov (HMMs).*
- *Técnicas basadas en la “no segmentación explícita” de las imágenes.*
- *Modelado de los niveles de percepción: morfológico, léxico y sintáctico.*
- *Integración de estos niveles en la fase de reconocimiento.*
- *Utilización de MEFs: HMMs, gramáticas o autómatas.*



# Introducción

## La tecnología actual de reconocimiento del habla

- *Basada en los Modelos ocultos de Markov (HMMs).*
- *Técnicas basadas en la “no segmentación explícita” de las imágenes.*
- *Modelado de los niveles de percepción: morfológico, léxico y sintáctico.*
- *Integración de estos niveles en la fase de reconocimiento.*
- *Utilización de MEFs: HMMs, gramáticas o autómatas.*

# Introducción

## La tecnología actual de reconocimiento del habla

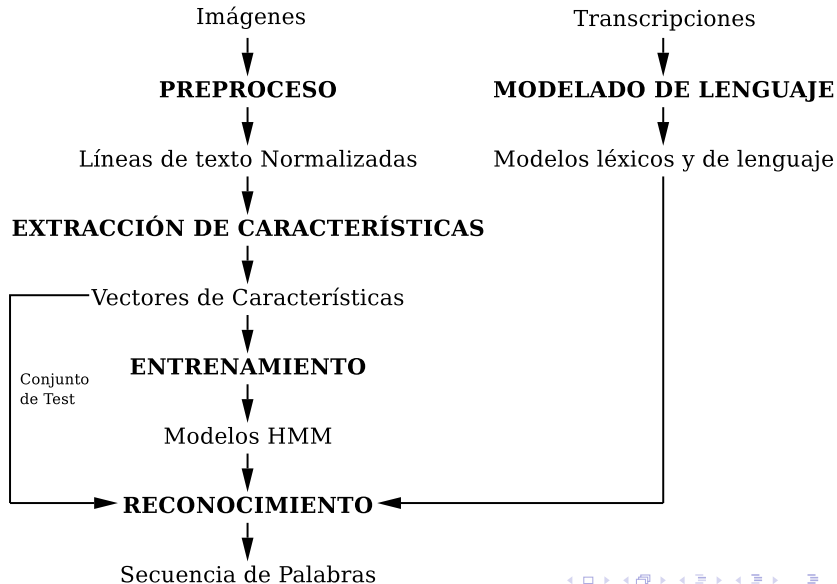
- *Basada en los Modelos ocultos de Markov (HMMs).*
- *Técnicas basadas en la “no segmentación explícita” de las imágenes.*
- *Modelado de los niveles de percepción: morfológico, léxico y sintáctico.*
- *Integración de estos niveles en la fase de reconocimiento.*
- *Utilización de MEFs: HMMs, gramáticas o autómatas.*

# Temario

## 1 Reconocimiento de Escritura Off-Line

- Introducción
- **Visión Global del Sistema**
- Preproceso de Texto Manuscrito - Página
- Eliminación del Fondo y Filtrado del Ruido
- Corrección de “Skew”
- Detección y Segmentado de Líneas
- Preproceso de Texto Manuscrito - Línea
- Corrección y Normalización de Estilos - Ilustración
- Corrección de “Slope”
- Corrección de “Slant”
- Normalización del Tamaño
- Extracción de Características
- Extracción de Características Utilizadas
- Reducción de la Dimensionalidad
- Entrenamiento de HMMs - Generalidades

# Visión Global del Sistema



# Temario

## 1 Reconocimiento de Escritura Off-Line

- Introducción
- Visión Global del Sistema
- **Preproceso de Texto Manuscrito - Página**
- Eliminación del Fondo y Filtrado del Ruido
- Corrección de “Skew”
- Detección y Segmentado de Líneas
- Preproceso de Texto Manuscrito - Línea
- Corrección y Normalización de Estilos - Ilustración
- Corrección de “Slope”
- Corrección de “Slant”
- Normalización del Tamaño
- Extracción de Características
- Extracción de Características Utilizadas
- Reducción de la Dimensionalidad
- Entrenamiento de HMMs - Generalidades

# Preproceso de Texto Manuscrito - Página

Operaciones de preproceso que se aplican sobre la imagen de una página con texto manuscrito:

**Filtrado de Ruido:** Intrínseco de la imagen y/o que pueda aparecer en el proceso de digitalización.

**Eliminación del Fondo:** Remoción de lo que no es texto manuscrito.

**Umbralización:** Transformación de una imagen en escala de grises a una imagen binaria.

**“Skew” o desencuadre:** Corrección del ángulo global de la página respecto de la horizontal.

**Extrac. de líneas:** Detección y segmentación de líneas de texto que se encuentran en la imagen de la página.

# Preproceso de Texto Manuscrito - Página

Operaciones de preproceso que se aplican sobre la imagen de una página con texto manuscrito:

**Filtrado de Ruido:** Intrínseco de la imagen y/o que pueda aparecer en el proceso de digitalización.

**Eliminación del Fondo:** Remoción de lo que no es texto manuscrito.

**Umbralización:** Transformación de una imagen en escala de grises a una imagen binaria.

**“Skew” o desencuadre:** Corrección del ángulo global de la página respecto de la horizontal.

**Extrac. de líneas:** Detección y segmentación de líneas de texto que se encuentran en la imagen de la página.

# Preproceso de Texto Manuscrito - Página

Operaciones de preproceso que se aplican sobre la imagen de una página con texto manuscrito:

**Filtrado de Ruido:** Intrínseco de la imagen y/o que pueda aparecer en el proceso de digitalización.

**Eliminación del Fondo:** Remoción de lo que no es texto manuscrito.

**Umbralización:** Transformación de una imagen en escala de grises a una imagen binaria.

**“Skew” o desencuadre:** Corrección del ángulo global de la página respecto de la horizontal.

**Extrac. de líneas:** Detección y segmentación de líneas de texto que se encuentran en la imagen de la página.



# Preproceso de Texto Manuscrito - Página

Operaciones de preproceso que se aplican sobre la imagen de una página con texto manuscrito:

**Filtrado de Ruido:** Intrínseco de la imagen y/o que pueda aparecer en el proceso de digitalización.

**Eliminación del Fondo:** Remoción de lo que no es texto manuscrito.

**Umbralización:** Transformación de una imagen en escala de grises a una imagen binaria.

**“Skew” o desencuadre:** Corrección del ángulo global de la página respecto de la horizontal.

**Extrac. de líneas:** Detección y segmentación de líneas de texto que se encuentran en la imagen de la página.

# Preproceso de Texto Manuscrito - Página

Operaciones de preproceso que se aplican sobre la imagen de una página con texto manuscrito:

**Filtrado de Ruido:** Intrínseco de la imagen y/o que pueda aparecer en el proceso de digitalización.

**Eliminación del Fondo:** Remoción de lo que no es texto manuscrito.

**Umbralización:** Transformación de una imagen en escala de grises a una imagen binaria.

**“Skew” o desencuadre:** Corrección del ángulo global de la página respecto de la horizontal.

**Extrac. de líneas:** Detección y segmentación de líneas de texto que se encuentran en la imagen de la página.

# Temario

## 1 Reconocimiento de Escritura Off-Line

- Introducción
- Visión Global del Sistema
- Preproceso de Texto Manuscrito - Página
- **Eliminación del Fondo y Filtrado del Ruido**
- Corrección de “Skew”
- Detección y Segmentado de Líneas
- Preproceso de Texto Manuscrito - Línea
- Corrección y Normalización de Estilos - Ilustración
- Corrección de “Slope”
- Corrección de “Slant”
- Normalización del Tamaño
- Extracción de Características
- Extracción de Características Utilizadas
- Reducción de la Dimensionalidad
- Entrenamiento de HMMs - Generalidades

# Eliminación del Fondo y Filtrado del Ruido

dejar el catolicismo social, que parece amenazar, y no se  
tema, si una invasión de la ignorancia vendrá, como  
Atlántico, como Attila a quemar sobre las ruinas más bellas.



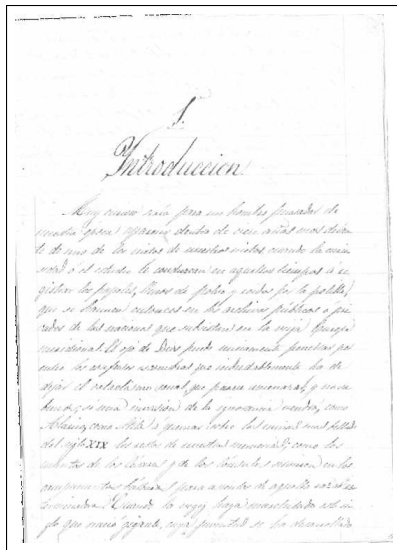
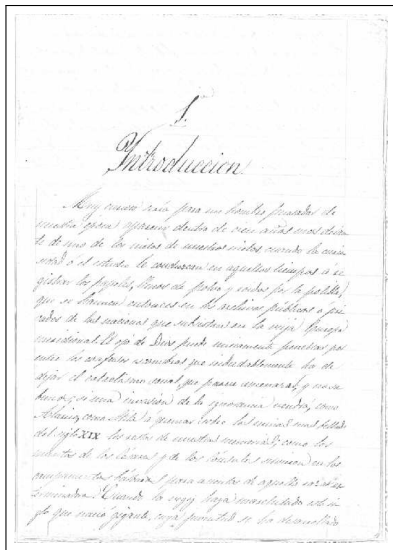
dejar el catolicismo social, que parece amenazar, y no se  
tema, si una invasión de la ignorancia vendrá, como  
Atlántico, como Attila a quemar sobre las ruinas más bellas.

# Temario

## 1 Reconocimiento de Escritura Off-Line

- Introducción
- Visión Global del Sistema
- Preproceso de Texto Manuscrito - Página
- Eliminación del Fondo y Filtrado del Ruido
- **Corrección de "Skew"**
- Detección y Segmentado de Líneas
- Preproceso de Texto Manuscrito - Línea
- Corrección y Normalización de Estilos - Ilustración
- Corrección de "Slope"
- Corrección de "Slant"
- Normalización del Tamaño
- Extracción de Características
- Extracción de Características Utilizadas
- Reducción de la Dimensionalidad
- Entrenamiento de HMMs - Generalidades

# Corrección de "Skew"

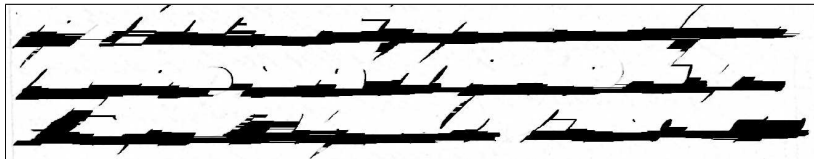


# Temario

## 1 Reconocimiento de Escritura Off-Line

- Introducción
- Visión Global del Sistema
- Preproceso de Texto Manuscrito - Página
- Eliminación del Fondo y Filtrado del Ruido
- Corrección de “Skew”
- **Detección y Segmentado de Líneas**
- Preproceso de Texto Manuscrito - Línea
- Corrección y Normalización de Estilos - Ilustración
- Corrección de “Slope”
- Corrección de “Slant”
- Normalización del Tamaño
- Extracción de Características
- Extracción de Características Utilizadas
- Reducción de la Dimensionalidad
- Entrenamiento de HMMs - Generalidades

# Detección y Segmentado de Líneas



dejar el catolicismo social, que parece amenazar, y no se  
teme, si una invasión de la ignorancia vendrá, como  
Abramo, como Mela a quemar sobre las cenizas más bellas.

dejar el catolicismo social, que parece amenazar, y no se

teme, si una invasión de la ignorancia vendrá, como

Abramo, como Mela a quemar sobre las cenizas más bellas.



# Temario

## 1 Reconocimiento de Escritura Off-Line

- Introducción
- Visión Global del Sistema
- Preproceso de Texto Manuscrito - Página
- Eliminación del Fondo y Filtrado del Ruido
- Corrección de “Skew”
- Detección y Segmentado de Líneas
- **Preproceso de Texto Manuscrito - Línea**
- Corrección y Normalización de Estilos - Ilustración
- Corrección de “Slope”
- Corrección de “Slant”
- Normalización del Tamaño
- Extracción de Características
- Extracción de Características Utilizadas
- Reducción de la Dimensionalidad
- Entrenamiento de HMMs - Generalidades

# Preproceso de Texto Manuscrito - Línea

Operaciones de preproceso que se aplican sobre los atributos de estilo de la escritura:

**Pendiente o “Slope”:** es el ángulo de la línea base del texto manuscrito respecto a la horizontal.

**Inclinación o “Slant”:** es el ángulo del trazo respecto a la vertical.

**Altura del Trazo de Escritura:** La cual varía en función de la tarea y de los autores. Tiene importancia la altura relativa entre letras ascendentes (ej.: b, l, t, etc.), descendentes (ej.: p, q, j, etc) y normales.

**Anchura de las Letras:** Al igual que la altura, varía en función de la tarea y de los autores.

**Grosor del Trazo:** el uso de diferentes elementos de escritura puede dar lugar a diferentes tipos y grosores de trazo.

# Preproceso de Texto Manuscrito - Línea

Operaciones de preproceso que se aplican sobre los atributos de estilo de la escritura:

**Pendiente o “Slope”:** es el ángulo de la línea base del texto manuscrito respecto a la horizontal.

**Inclinación o “Slant”:** es el ángulo del trazo respecto a la vertical.

**Altura del Trazo de Escritura:** La cual varía en función de la tarea y de los autores. Tiene importancia la altura relativa entre letras ascendentes (ej.: b, l, t, etc.), descendentes (ej.: p, q, j, etc) y normales.

**Anchura de las Letras:** Al igual que la altura, varía en función de la tarea y de los autores.

**Grosor del Trazo:** el uso de diferentes elementos de escritura puede dar lugar a diferentes tipos y grosores de trazo.

# Preproceso de Texto Manuscrito - Línea

Operaciones de preproceso que se aplican sobre los atributos de estilo de la escritura:

**Pendiente o “Slope”:** es el ángulo de la línea base del texto manuscrito respecto a la horizontal.

**Inclinación o “Slant”:** es el ángulo del trazo respecto a la vertical.

**Altura del Trazo de Escritura:** La cual varía en función de la tarea y de los autores. Tiene importancia la altura relativa entre letras ascendentes (ej.: b, l, t, etc.), descendentes (ej.: p, q, j, etc) y normales.

**Anchura de las Letras:** Al igual que la altura, varía en función de la tarea y de los autores.

**Grosor del Trazo:** el uso de diferentes elementos de escritura puede dar lugar a diferentes tipos y grosores de trazo.

# Preproceso de Texto Manuscrito - Línea

Operaciones de preproceso que se aplican sobre los atributos de estilo de la escritura:

**Pendiente o “Slope”:** es el ángulo de la línea base del texto manuscrito respecto a la horizontal.

**Inclinación o “Slant”:** es el ángulo del trazo respecto a la vertical.

**Altura del Trazo de Escritura:** La cual varía en función de la tarea y de los autores. Tiene importancia la altura relativa entre letras ascendentes (ej.: b, l, t, etc.), descendentes (ej.: p, q, j, etc) y normales.

**Anchura de las Letras:** Al igual que la altura, varía en función de la tarea y de los autores.

**Grosor del Trazo:** el uso de diferentes elementos de escritura puede dar lugar a diferentes tipos y grosores de trazo.

# Preproceso de Texto Manuscrito - Línea

Operaciones de preproceso que se aplican sobre los atributos de estilo de la escritura:

**Pendiente o “Slope”:** es el ángulo de la línea base del texto manuscrito respecto a la horizontal.

**Inclinación o “Slant”:** es el ángulo del trazo respecto a la vertical.

**Altura del Trazo de Escritura:** La cual varía en función de la tarea y de los autores. Tiene importancia la altura relativa entre letras ascendentes (ej.: b, l, t, etc.), descendentes (ej.: p, q, j, etc) y normales.

**Anchura de las Letras:** Al igual que la altura, varía en función de la tarea y de los autores.

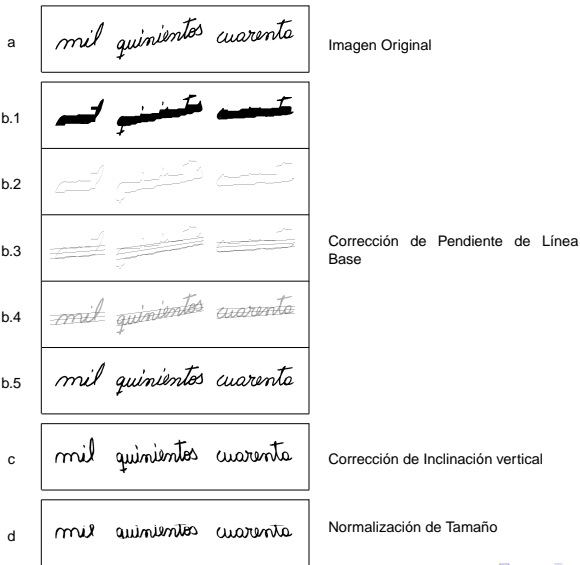
**Grosor del Trazo:** el uso de diferentes elementos de escritura puede dar lugar a diferentes tipos y grosores de trazo.

# Temario

## 1 Reconocimiento de Escritura Off-Line

- Introducción
- Visión Global del Sistema
- Preproceso de Texto Manuscrito - Página
- Eliminación del Fondo y Filtrado del Ruido
- Corrección de "Skew"
- Detección y Segmentado de Líneas
- Preproceso de Texto Manuscrito - Línea
- **Corrección y Normalización de Estilos - Ilustración**
- Corrección de "Slope"
- Corrección de "Slant"
- Normalización del Tamaño
- Extracción de Características
- Extracción de Características Utilizadas
- Reducción de la Dimensionalidad
- Entrenamiento de HMMs - Generalidades

# Corrección y Normalización de Estilos - Ilustración



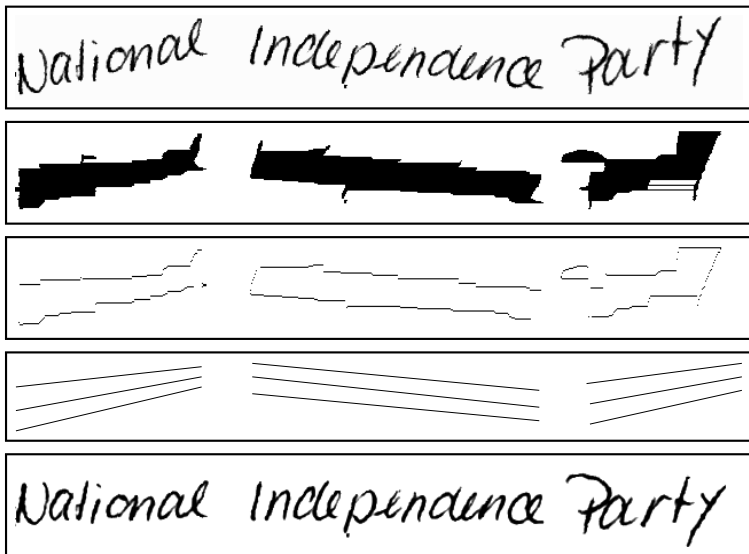


# Temario

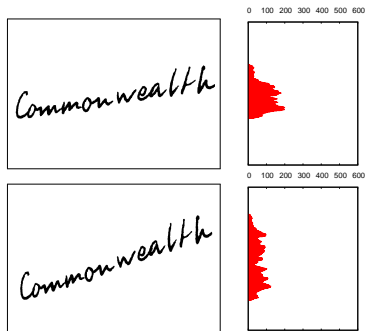
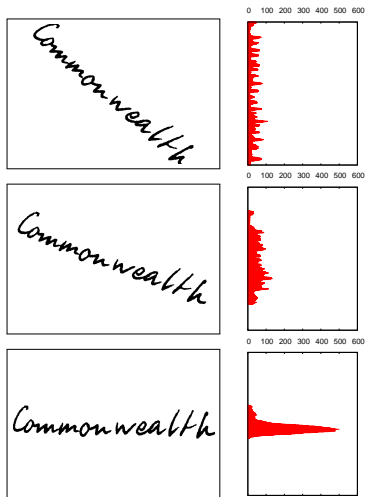
## 1 Reconocimiento de Escritura Off-Line

- Introducción
- Visión Global del Sistema
- Preproceso de Texto Manuscrito - Página
- Eliminación del Fondo y Filtrado del Ruido
- Corrección de "Skew"
- Detección y Segmentado de Líneas
- Preproceso de Texto Manuscrito - Línea
- Corrección y Normalización de Estilos - Ilustración
- **Corrección de "Slope"**
- Corrección de "Slant"
- Normalización del Tamaño
- Extracción de Características
- Extracción de Características Utilizadas
- Reducción de la Dimensionalidad
- Entrenamiento de HMMs - Generalidades

# Corrección de "Slope": Método 1



# Corrección de "Slope": Método 2



$$\hat{\alpha} = \arg \max_{\alpha \in [45:135]} F(P_h(y, \alpha))$$

$$P_h(y, \alpha) = \sum_{x=1}^{rows} I_{\alpha}(x, y)$$

$$F(P_h(y, \alpha)) = \sqrt{\frac{\sum_{y=1}^{rows} (\mu - P_h(y, \alpha))^2}{rows}}$$

# Temario

## 1 Reconocimiento de Escritura Off-Line

- Introducción
- Visión Global del Sistema
- Preproceso de Texto Manuscrito - Página
- Eliminación del Fondo y Filtrado del Ruido
- Corrección de "Skew"
- Detección y Segmentado de Líneas
- Preproceso de Texto Manuscrito - Línea
- Corrección y Normalización de Estilos - Ilustración
- Corrección de "Slope"
- **Corrección de "Slant"**
- Normalización del Tamaño
- Extracción de Características
- Extracción de Características Utilizadas
- Reducción de la Dimensionalidad
- Entrenamiento de HMMs - Generalidades

# Corrección de "Slant" - Por detección de bordes

- Convolución de la imagen con filtros de "Sobel":

$$K_{S_h} = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \quad K_{S_v} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$$

- Cómputo del ángulo en cada punto de la imagen:

$$\theta_{i,j} = \arctan \left( \frac{G_{y,i,j}}{G_{x,i,j}} \right)$$

- Construcción de histograma de frecuencia de ángulos:

$$H(\alpha) = f(\theta_{i,j}).$$

- Cómo ángulo de "slant" se toma la media del histograma:

$$H'(\alpha) = H(\alpha) \cdot F(\alpha), \text{ donde:}$$

$$F(\alpha) = \left( 1 - \frac{|90 - \alpha|}{90} \right)$$

# Corrección de "Slant" - Por detección de bordes

- Convolución de la imagen con filtros de "Sobel":

$$K_{S_h} = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \quad K_{S_v} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$$

- Cómputo del ángulo en cada punto de la imagen:

$$\theta_{i,j} = \arctan \left( \frac{G_{y,i,j}}{G_{x,i,j}} \right)$$

- Construcción de histograma de frecuencia de ángulos:

$$H(\alpha) = f(\theta_{i,j}).$$

- Cómo ángulo de "slant" se toma la media del histograma:

$$H'(\alpha) = H(\alpha) \cdot F(\alpha), \text{ donde:}$$

$$F(\alpha) = \left( 1 - \frac{|90 - \alpha|}{90} \right)$$

# Corrección de "Slant" - Por detección de bordes

- Convolución de la imagen con filtros de "Sobel":

$$K_{S_h} = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \quad K_{S_v} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$$

- Cómputo del ángulo en cada punto de la imagen:

$$\theta_{i,j} = \arctan \left( \frac{G_{y,i,j}}{G_{x,i,j}} \right)$$

- Construcción de histograma de frecuencia de ángulos:

$$H(\alpha) = f(\theta_{i,j}).$$

- Cómo ángulo de "slant" se toma la media del histograma:

$$H'(\alpha) = H(\alpha) \cdot F(\alpha), \text{ donde:}$$

$$F(\alpha) = \left( 1 - \frac{|90 - \alpha|}{90} \right)$$

# Corrección de "Slant" - Por detección de bordes

- Convolución de la imagen con filtros de "Sobel":

$$K_{S_h} = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \quad K_{S_v} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$$

- Cómputo del ángulo en cada punto de la imagen:

$$\theta_{i,j} = \arctan \left( \frac{G_{y,i,j}}{G_{x,i,j}} \right)$$

- Construcción de histograma de frecuencia de ángulos:

$$H(\alpha) = f(\theta_{i,j}).$$

- Cómo ángulo de "slant" se toma la media del histograma:

$$H'(\alpha) = H(\alpha) \cdot F(\alpha), \text{ donde:}$$

$$F(\alpha) = \left( 1 - \frac{|90 - \alpha|}{90} \right)$$



# Corrección de "Slant" - Por Proyección Vertical(1)

- Aplicamos transformadas de desplazamiento a la imagen  $I(x, y)$  con diferentes ángulos  $\alpha \in [45 : 135]$ .
- Obtenemos las proyecciones verticales:  $P_v(x, \alpha) = \sum_{y=1}^{rows} I_\alpha(x, y)$
- Cómo ángulo de "slant" se toma aquel que maximice:

$$\hat{\alpha} = \arg \max_{\alpha \in [45:135]} F(P_v(x, \alpha))$$

donde  $F$  es una de las siguientes funciones objetivo:

- Desviación Típica:  $F(P_v(x, \alpha)) = \sqrt{\sum_{x=1}^{cols} \frac{(\mu - P_v(x, \alpha))^2}{cols}}$
- Longitud del Perfil:  

$$F(P_v(x, \alpha)) = \sum_{x=1}^{cols-1} \sqrt{1 + (P_v(x, \alpha) - P_v(x+1, \alpha))^2}$$
- IDIAP:  $F(P_v(x, \alpha)) = \sum_{x: C_\alpha(x)=1} P_v(x, \alpha)^2$  con  $C_\alpha(x) = \frac{P_v(x, \alpha)}{y_{sup}(x) - y_{inf}(x)}$

# Corrección de "Slant" - Por Proyección Vertical(1)

- Aplicamos transformadas de desplazamiento a la imagen  $I(x, y)$  con diferentes ángulos  $\alpha \in [45 : 135]$ .
- Obtenemos las proyecciones verticales:  $P_v(x, \alpha) = \sum_{y=1}^{rows} I_\alpha(x, y)$
- Cómo ángulo de "slant" se toma aquel que maximice:

$$\hat{\alpha} = \arg \max_{\alpha \in [45:135]} F(P_v(x, \alpha))$$

donde F es una de las siguientes funciones objetivo:

- Desviación Típica:  $F(P_v(x, \alpha)) = \sqrt{\sum_{x=1}^{cols} \frac{(\mu - P_v(x, \alpha))^2}{cols}}$
- Longitud del Perfil:  
 $F(P_v(x, \alpha)) = \sum_{x=1}^{cols-1} \sqrt{1 + (P_v(x, \alpha) - P_v(x+1, \alpha))^2}$
- IDIAP:  $F(P_v(x, \alpha)) = \sum_{x: C_\alpha(x)=1} P_v(x, \alpha)^2$  con  $C_\alpha(x) = \frac{P_v(x, \alpha)}{y_{sup}(x) - y_{inf}(x)}$

# Corrección de "Slant" - Por Proyección Vertical(1)

- Aplicamos transformadas de desplazamiento a la imagen  $I(x, y)$  con diferentes ángulos  $\alpha \in [45 : 135]$ .
- Obtenemos las proyecciones verticales:  $P_v(x, \alpha) = \sum_{y=1}^{rows} I_\alpha(x, y)$
- Cómo ángulo de "slant" se toma aquel que maximice:

$$\hat{\alpha} = \arg \max_{\alpha \in [45:135]} F(P_v(x, \alpha))$$

donde F es una de las siguientes funciones objetivo:

- Desviación Típica:  $F(P_v(x, \alpha)) = \sqrt{\sum_{x=1}^{cols} \frac{(\mu - P_v(x, \alpha))^2}{cols}}$
- Longitud del Perfil:  
 $F(P_v(x, \alpha)) = \sum_{x=1}^{cols-1} \sqrt{1 + (P_v(x, \alpha) - P_v(x+1, \alpha))^2}$
- IDIAP:  $F(P_v(x, \alpha)) = \sum_{x: C_\alpha(x)=1} P_v(x, \alpha)^2$  con  $C_\alpha(x) = \frac{P_v(x, \alpha)}{y_{sup}(x) - y_{inf}(x)}$

# Corrección de "Slant" - Por Proyección Vertical(1)

- Aplicamos transformadas de desplazamiento a la imagen  $I(x, y)$  con diferentes ángulos  $\alpha \in [45 : 135]$ .
- Obtenemos las proyecciones verticales:  $P_v(x, \alpha) = \sum_{y=1}^{rows} I_\alpha(x, y)$
- Cómo ángulo de "slant" se toma aquel que maximice:

$$\hat{\alpha} = \arg \max_{\alpha \in [45:135]} F(P_v(x, \alpha))$$

donde F es una de las siguientes funciones objetivo:

- Desviación Típica:  $F(P_v(x, \alpha)) = \sqrt{\sum_{x=1}^{cols} \frac{(\mu - P_v(x, \alpha))^2}{cols}}$
- Longitud del Perfil:  

$$F(P_v(x, \alpha)) = \sum_{x=1}^{cols-1} \sqrt{1 + (P_v(x, \alpha) - P_v(x+1, \alpha))^2}$$
- IDIAP:  $F(P_v(x, \alpha)) = \sum_{x: C_\alpha(x)=1} P_v(x, \alpha)^2$  con  $C_\alpha(x) = \frac{P_v(x, \alpha)}{y_{sup}(x) - y_{inf}(x)}$

# Corrección de "Slant" - Por Proyección Vertical(1)

- Aplicamos transformadas de desplazamiento a la imagen  $I(x, y)$  con diferentes ángulos  $\alpha \in [45 : 135]$ .
- Obtenemos las proyecciones verticales:  $P_v(x, \alpha) = \sum_{y=1}^{rows} I_\alpha(x, y)$
- Cómo ángulo de "slant" se toma aquel que maximice:

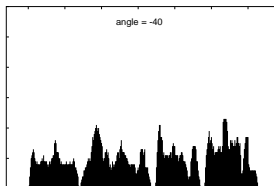
$$\hat{\alpha} = \arg \max_{\alpha \in [45:135]} F(P_v(x, \alpha))$$

donde F es una de las siguientes funciones objetivo:

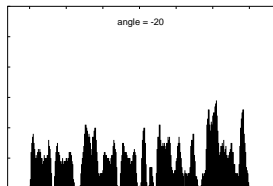
- Desviación Típica:  $F(P_v(x, \alpha)) = \sqrt{\sum_{x=1}^{cols} \frac{(\mu - P_v(x, \alpha))^2}{cols}}$
- Longitud del Perfil:  

$$F(P_v(x, \alpha)) = \sum_{x=1}^{cols-1} \sqrt{1 + (P_v(x, \alpha) - P_v(x+1, \alpha))^2}$$
- IDIAP:  $F(P_v(x, \alpha)) = \sum_{x: C_\alpha(x)=1} P_v(x, \alpha)^2$  con  $C_\alpha(x) = \frac{P_v(x, \alpha)}{y_{sup}(x) - y_{inf}(x)}$

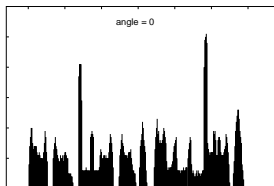
# Corrección de "Slant" - Por Proyección Vertical(2)



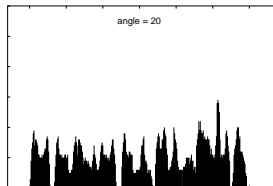
*ochocientos*



*ochocientos*



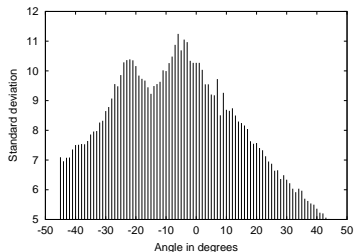
*ochocientos*



*ochocientos*

# Corrección de "Slant" - Por Proyección Vertical(3)

Cuando  $F(P_V(x, \alpha)) = STD$  y el perfil vertical presenta una forma multimodal:



El ángulo suavizado resultante  $\bar{\alpha}$ , se calcula como el centro de masa de los ángulos  $\alpha$  cuya desviación estándar sea mayor o igual que  $\hat{\alpha}(1 - \rho)$ :

$$\bar{\alpha} = \frac{\sum_{\alpha: \alpha \geq \hat{\alpha}(1-\rho)} \alpha F(P_V(x, \alpha))}{\sum_{\alpha: \alpha \geq \hat{\alpha}(1-\rho)} F(P_V(x, \alpha))} \quad \rho \in [0, 1]$$

# Temario

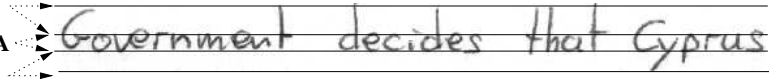
## 1 Reconocimiento de Escritura Off-Line

- Introducción
- Visión Global del Sistema
- Preproceso de Texto Manuscrito - Página
- Eliminación del Fondo y Filtrado del Ruido
- Corrección de “Skew”
- Detección y Segmentado de Líneas
- Preproceso de Texto Manuscrito - Línea
- Corrección y Normalización de Estilos - Ilustración
- Corrección de “Slope”
- Corrección de “Slant”
- **Normalización del Tamaño**
- Extracción de Características
- Extracción de Características Utilizadas
- Reducción de la Dimensionalidad
- Entrenamiento de HMMs - Generalidades



# Normalización del Tamaño

AA  
MBA  
DA



Government decides that Cyprus

1

is to be made at a

2

is to be made at a

3

is to be made at a

4

is to be made at a

5

is to be made at a

6

is to be made at a

# Temario

## 1 Reconocimiento de Escritura Off-Line

- Introducción
- Visión Global del Sistema
- Preproceso de Texto Manuscrito - Página
- Eliminación del Fondo y Filtrado del Ruido
- Corrección de “Skew”
- Detección y Segmentado de Líneas
- Preproceso de Texto Manuscrito - Línea
- Corrección y Normalización de Estilos - Ilustración
- Corrección de “Slope”
- Corrección de “Slant”
- Normalización del Tamaño
- **Extracción de Características**
- Extracción de Características Utilizadas
- Reducción de la Dimensionalidad
- Entrenamiento de HMMs - Generalidades

# Extracción de Características

- Establece un nuevo espacio de representación de la señal de entrada (imagen).
- Debe ser:
  - compacta.
  - minimice la variabilidad intra-clase.
  - maximice la variabilidad inter-clase.
  - adecuada al tipo de reconocedor/clasificador utilizado.
  - fácil de obtener, a un bajo coste computacional.
- La evaluación de las mismas se hace indirectamente, a partir del error de reconocimiento/clasificación.
- Existen métodos formales para selección de características, aunque no son ampliamente utilizados.
- Que a partir de ellas se pueda reconstruir la señal de entrada o una aproximación de ella.

# Extracción de Características

- Establece un nuevo espacio de representación de la señal de entrada (imagen).
- Debe ser:
  - compacta.
  - minimice la variabilidad intra-clase.
  - maximice la variabilidad inter-clase.
  - adecuada al tipo de reconocedor/clasificador utilizado.
  - fácil de obtener, a un bajo coste computacional.
- La evaluación de las mismas se hace indirectamente, a partir del error de reconocimiento/clasificación.
- Existen métodos formales para selección de características, aunque no son ampliamente utilizados.
- Que a partir de ellas se pueda reconstruir la señal de entrada o una aproximación de ella.

# Extracción de Características

- Establece un nuevo espacio de representación de la señal de entrada (imagen).
- Debe ser:
  - compacta.
  - minimice la variabilidad intra-clase.
  - maximice la variabilidad inter-clase.
  - adecuada al tipo de reconocedor/clasificador utilizado.
  - fácil de obtener, a un bajo coste computacional.
- La evaluación de las mismas se hace indirectamente, a partir del error de reconocimiento/clasificación.
- Existen métodos formales para selección de características, aunque no son ampliamente utilizados.
- Que a partir de ellas se pueda reconstruir la señal de entrada o una aproximación de ella.

# Extracción de Características

- Establece un nuevo espacio de representación de la señal de entrada (imagen).
- Debe ser:
  - compacta.
  - minimice la variabilidad intra-clase.
  - maximice la variabilidad inter-clase.
  - adecuada al tipo de reconocedor/clasificador utilizado.
  - fácil de obtener, a un bajo coste computacional.
- La evaluación de las mismas se hace indirectamente, a partir del error de reconocimiento/clasificación.
- Existen métodos formales para selección de características, aunque no son ampliamente utilizados.
- Que a partir de ellas se pueda reconstruir la señal de entrada o una aproximación de ella.

# Extracción de Características

- Establece un nuevo espacio de representación de la señal de entrada (imagen).
- Debe ser:
  - compacta.
  - minimice la variabilidad intra-clase.
  - maximice la variabilidad inter-clase.
  - adecuada al tipo de reconocedor/clasificador utilizado.
  - fácil de obtener, a un bajo coste computacional.
- La evaluación de las mismas se hace indirectamente, a partir del error de reconocimiento/clasificación.
- Existen métodos formales para selección de características, aunque no son ampliamente utilizados.
- Que a partir de ellas se pueda reconstruir la señal de entrada o una aproximación de ella.

# Extracción de Características - Taxonomía(1)

- Global o Local.
- Extraídas a partir de distribuciones estadísticas de los puntos:
  - Vent: Sencillas, bajo coste computacional, dimensionalidad baja.
  - Desv: No permiten la reconstrucción de la señal original.
  - Ej.: Proyecciones verticales/horizontales, medias, desviaciones, etc.
- Extraídas de transformadas globales o expansiones en series:
  - Vent: Permiten reconstruir la la señal original.
  - Desv: Alto coste computacional por lo general.
  - Ej.: Transformadas de Fourier, Transformadas Wavelets Discretas, Filtros Gabor 2D, Momentos Geométricos.



# Extracción de Características - Taxonomía(1)

- Global o Local.
- Extraídas a partir de distribuciones estadísticas de los puntos:
  - Vent:** Sencillas, bajo coste computacional, dimensionalidad baja.
  - Desv:** No permiten la reconstrucción de la señal original.
  - Ej.:** Proyecciones verticales/horizontales, medias, desviaciones, etc.
- Extraídas de transformadas globales o expansiones en series:
  - Vent:** Permiten reconstruir la la señal original.
  - Desv:** Alto coste computacional por lo general.
  - Ej.:** Transformadas de Fourier, Transformadas Wavelets Discretas, Filtros Gabor 2D, Momentos Geométricos.

# Extracción de Características - Taxonomía(1)

- Global o Local.
- Extraídas a partir de distribuciones estadísticas de los puntos:
  - Vent:** Sencillas, bajo coste computacional, dimensionalidad baja.
  - Desv:** No permiten la reconstrucción de la señal original.
  - Ej.:** Proyecciones verticales/horizontales, medias, desviaciones, etc.
- Extraídas de transformadas globales o expansiones en series:
  - Vent:** Permiten reconstruir la la señal original.
  - Desv:** Alto coste computacional por lo general.
  - Ej.:** Transformadas de Fourier, Transformadas Wavelets Discretas, Filtros Gabor 2D, Momentos Geométricos.

# Extracción de Características - Taxonomía(2)

- Extraídas de características topológicas y geométricas:

**Vent:** Robusto frente a estilos de escritura y distorsiones.

**Desv:** ?.

**Ej.:** Cadenas de contorno, medidas y representaciones de propiedades geométricas, uso de primitivas para representar estructuras topológicas y grafos para representar la relación de ellas entre sí.

# Extracción de Características (HMMs)

Adecuadas para ser utilizadas con HMMs:

- Secuencia de vectores reales de dimensión fija.
- Los vectores se extraen en el sentido y dirección de escritura (se recorre la imagen de izquierda a derecha).
- Basados en características geométricas.
- Permiten la reconstrucción aproximada de la imagen original.

# Temario

## 1 Reconocimiento de Escritura Off-Line

- Introducción
- Visión Global del Sistema
- Preproceso de Texto Manuscrito - Página
- Eliminación del Fondo y Filtrado del Ruido
- Corrección de “Skew”
- Detección y Segmentado de Líneas
- Preproceso de Texto Manuscrito - Línea
- Corrección y Normalización de Estilos - Ilustración
- Corrección de “Slope”
- Corrección de “Slant”
- Normalización del Tamaño
- Extracción de Características
- **Extracción de Características Utilizadas**
- Reducción de la Dimensionalidad
- Entrenamiento de HMMs - Generalidades

# Comparativas

## Extracción de Características del PRHLT vs FKI:

	<b>PRHLT</b>	<b>FKI</b>
Origen	Bazzi&Schwartz <sup>†</sup>	— <sup>‡</sup>
Tipo	Geométrica	Geométrica
Nivel	Local	Global
Se Extraen de	Ventanas Locales	Columna de Píxeles
Dimensión	$\geq 60$	9
Caract. Normalizadas	no	sí
Num. Vect/Imagen	$\propto$ Dimensión	Ancho Imagen

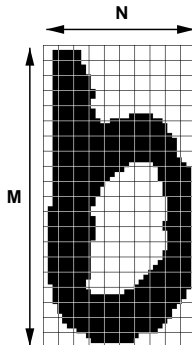
<sup>†</sup> Issam Bazzi, Richard Schwartz y John Makhoul. "An Omnifont Open-Vocabulary OCR System for English and Arabic". *IEEE Trans. Pattern Analysis and Machine Intelligence*. 21(6),495-504 (junio 1999).

<sup>‡</sup> Marti, U. and Bunke, H. 2002. *Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition systems*. In *Hidden Markov Models: Applications in Computer Vision World Scientific Series In Machine Perception And Artificial Intelligence Series*, vol. 45. World Scientific Publishing Co., River Edge, NJ, 65-90.

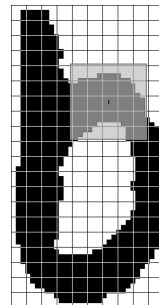
# Método del PRHLT - Ventana de Muestreo



Imagen Original

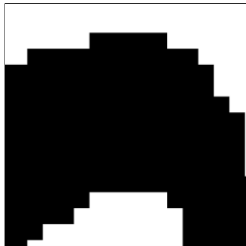


Rejilla

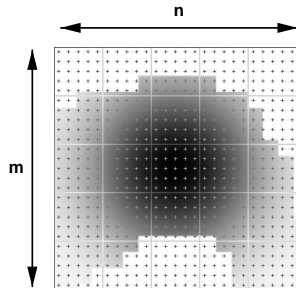


Ventana de Muestreo

# Método del PRHLT - Nivel de Gris



Porción de trazo encerrado en la ventana de muestreo



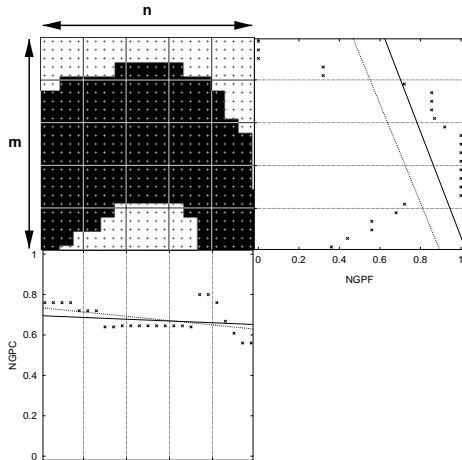
Puntos de muestreo definidos en la ventana y trazo afectado por el filtro Gaussiano

La intensidad  $\hat{l}(i, j)$  afectada por el filtro Gaussiano es:

$$\hat{l}(i, j) = l(i, j) \exp \left[ -\frac{1}{2} \left( \frac{j - (n/2)^2}{(n/4)^2} + \frac{i - (m/2)^2}{(m/4)^2} \right) \right]$$



# Método del PRHLT - Derivadas



Número Promedio de Píxeles por Columna:

$$\text{NGPC} \rightarrow g_j = \frac{\sum_{i=1}^m l(i, j)}{m}$$

Recta a Aproximar:

$$J = \sum_{j=1}^n w_j (g_j - (j+b))^2$$

Filtro:

$$w_j = \exp \left( -\frac{1}{2} \frac{(j - n/2)^2}{(n/4)^2} \right)$$

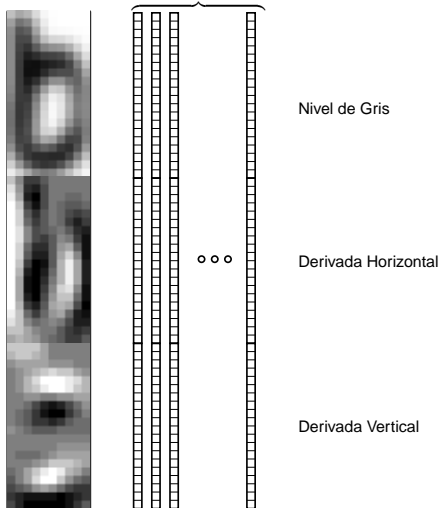
Condición:

$$\frac{\partial J}{\partial a} = 0 \quad \text{y} \quad \frac{\partial J}{\partial b} = 0$$

$$\text{Derivada: } a = \frac{\left( \sum_{j=1}^n w_j g_j \right) \left( \sum_{j=1}^n w_j j \right) - \left( \sum_{j=1}^n w_j \right) \left( \sum_{j=1}^n w_j g_j j \right)}{\left( \sum_{j=1}^n w_j j \right)^2 - \left( \sum_{j=1}^n w_j \right) \left( \sum_{j=1}^n w_j j^2 \right)}$$

# Método del PRHLT - Representación

Secuencia de Vectores  
de Características



# Método del Grupo FKI del IAM (Suiza)

Las 9 características utilizadas para la columna  $j$  son:

$c_1(j)$ : Masa de la columna.

$c_2(j)$ : Centro de gravedad de la columna.

$c_3(j)$ : Momento de segundo orden de la columna.

$c_4(j)$ : Posición del píxel negro más alto en la columna.

$c_5(j)$ : Posición del píxel negro mas bajo en la columna.

$c_6(j)$ : Derivada del contorno superior.

$c_7(j)$ : Derivada del contorno inferior.

$c_8(j)$ : Número de transiciones negro-blanco en la columna.

$c_9(j)$ : Número de píxeles negros entre el contorno superior y el inferior.

# Método del Grupo FKI del IAM (Suiza)

Las 9 características utilizadas para la columna  $j$  son:

$c_1(j)$ : Masa de la columna.

$c_2(j)$ : Centro de gravedad de la columna.

$c_3(j)$ : Momento de segundo orden de la columna.

$c_4(j)$ : Posición del píxel negro más alto en la columna.

$c_5(j)$ : Posición del píxel negro mas bajo en la columna.

$c_6(j)$ : Derivada del contorno superior.

$c_7(j)$ : Derivada del contorno inferior.

$c_8(j)$ : Número de transiciones negro-blanco en la columna.

$c_9(j)$ : Número de píxeles negros entre el contorno superior y el inferior.

# Método del Grupo FKI del IAM (Suiza)

Las 9 características utilizadas para la columna  $j$  son:

$c_1(j)$ : Masa de la columna.

$c_2(j)$ : Centro de gravedad de la columna.

$c_3(j)$ : Momento de segundo orden de la columna.

$c_4(j)$ : Posición del píxel negro más alto en la columna.

$c_5(j)$ : Posición del píxel negro mas bajo en la columna.

$c_6(j)$ : Derivada del contorno superior.

$c_7(j)$ : Derivada del contorno inferior.

$c_8(j)$ : Número de transiciones negro-blanco en la columna.

$c_9(j)$ : Número de píxeles negros entre el contorno superior y el inferior.

# Método del Grupo FKI del IAM (Suiza)

Las 9 características utilizadas para la columna  $j$  son:

$c_1(j)$ : Masa de la columna.

$c_2(j)$ : Centro de gravedad de la columna.

$c_3(j)$ : Momento de segundo orden de la columna.

$c_4(j)$ : Posición del píxel negro más alto en la columna.

$c_5(j)$ : Posición del píxel negro mas bajo en la columna.

$c_6(j)$ : Derivada del contorno superior.

$c_7(j)$ : Derivada del contorno inferior.

$c_8(j)$ : Número de transiciones negro-blanco en la columna.

$c_9(j)$ : Número de píxeles negros entre el contorno superior y el inferior.

# Método del Grupo FKI del IAM (Suiza)

Las 9 características utilizadas para la columna  $j$  son:

$c_1(j)$ : Masa de la columna.

$c_2(j)$ : Centro de gravedad de la columna.

$c_3(j)$ : Momento de segundo orden de la columna.

$c_4(j)$ : Posición del píxel negro más alto en la columna.

$c_5(j)$ : Posición del píxel negro mas bajo en la columna.

$c_6(j)$ : Derivada del contorno superior.

$c_7(j)$ : Derivada del contorno inferior.

$c_8(j)$ : Número de transiciones negro-blanco en la columna.

$c_9(j)$ : Número de píxeles negros entre el contorno superior y el inferior.

# Método del Grupo FKI del IAM (Suiza)

Las 9 características utilizadas para la columna  $j$  son:

$c_1(j)$ : Masa de la columna.

$c_2(j)$ : Centro de gravedad de la columna.

$c_3(j)$ : Momento de segundo orden de la columna.

$c_4(j)$ : Posición del píxel negro más alto en la columna.

$c_5(j)$ : Posición del píxel negro mas bajo en la columna.

$c_6(j)$ : Derivada del contorno superior.

$c_7(j)$ : Derivada del contorno inferior.

$c_8(j)$ : Número de transiciones negro-blanco en la columna.

$c_9(j)$ : Número de píxeles negros entre el contorno superior y el inferior.



# Método del Grupo FKI del IAM (Suiza)

Las 9 características utilizadas para la columna  $j$  son:

$c_1(j)$ : Masa de la columna.

$c_2(j)$ : Centro de gravedad de la columna.

$c_3(j)$ : Momento de segundo orden de la columna.

$c_4(j)$ : Posición del píxel negro más alto en la columna.

$c_5(j)$ : Posición del píxel negro mas bajo en la columna.

$c_6(j)$ : Derivada del contorno superior.

$c_7(j)$ : Derivada del contorno inferior.

$c_8(j)$ : Número de transiciones negro-blanco en la columna.

$c_9(j)$ : Número de píxeles negros entre el contorno superior y el inferior.

# Método del Grupo FKI del IAM (Suiza)

Las 9 características utilizadas para la columna  $j$  son:

$c_1(j)$ : Masa de la columna.

$c_2(j)$ : Centro de gravedad de la columna.

$c_3(j)$ : Momento de segundo orden de la columna.

$c_4(j)$ : Posición del píxel negro más alto en la columna.

$c_5(j)$ : Posición del píxel negro mas bajo en la columna.

$c_6(j)$ : Derivada del contorno superior.

$c_7(j)$ : Derivada del contorno inferior.

$c_8(j)$ : Número de transiciones negro-blanco en la columna.

$c_9(j)$ : Número de píxeles negros entre el contorno superior y el inferior.

# Método del Grupo FKI del IAM (Suiza)

Las 9 características utilizadas para la columna  $j$  son:

$c_1(j)$ : Masa de la columna.

$c_2(j)$ : Centro de gravedad de la columna.

$c_3(j)$ : Momento de segundo orden de la columna.

$c_4(j)$ : Posición del píxel negro más alto en la columna.

$c_5(j)$ : Posición del píxel negro mas bajo en la columna.

$c_6(j)$ : Derivada del contorno superior.

$c_7(j)$ : Derivada del contorno inferior.

$c_8(j)$ : Número de transiciones negro-blanco en la columna.

$c_9(j)$ : Número de píxeles negros entre el contorno superior y el inferior.

# Método del Grupo FKI del IAM (Suiza)

Expresadas en forma matemática:

**ngr:** píxel negro

**blc:** píxel blanco

$$c_1(j) = \sum_{i=1}^m I(i, j)$$

$$c_2(j) = \frac{1}{m} \sum_{i=1}^m i \cdot I(i, j)$$

$$c_3(j) = \frac{1}{m^2} \sum_{i=1}^m i^2 \cdot I(i, j)$$

$$c_4(j) = \min(i \mid I(i, j) = \text{ngr})$$

$$c_5(j) = \max(i \mid I(i, j) = \text{ngr})$$

$$c_6(j) = \frac{c_4(j+1) - c_4(j-1)}{2}$$

$$c_7(j) = \frac{c_5(j+1) - c_5(j-1)}{2}$$

$$c_8(j) = NT_{\text{ngr} \rightarrow \text{blc}}(I(i, j))$$

$$c_9(j) = \sum_{c_4(j) < j < c_5(j)} I(i, j)$$

# Temario

## 1 Reconocimiento de Escritura Off-Line

- Introducción
- Visión Global del Sistema
- Preproceso de Texto Manuscrito - Página
- Eliminación del Fondo y Filtrado del Ruido
- Corrección de “Skew”
- Detección y Segmentado de Líneas
- Preproceso de Texto Manuscrito - Línea
- Corrección y Normalización de Estilos - Ilustración
- Corrección de “Slope”
- Corrección de “Slant”
- Normalización del Tamaño
- Extracción de Características
- Extracción de Características Utilizadas
- **Reducción de la Dimensionalidad**
- Entrenamiento de HMMs - Generalidades

# Métodos de Reducción de Dimensión de Vectores

- El objetivo principal es el de reducir el tiempo de reconocimiento final.
- Se aplica sobre la extracción de características.
- Filtrar ruido.
- Descartar características redundantes.
- Entre los métodos de reducción de dimensionalidad más utilizados se encuentran:
  - PCA: *Principal Analysis Components*
  - LDA: *Linear Discriminant Analysis*

# PCA: Principal Analysis Components

**Propósito:** reducir la dimensionalidad de los datos (conjunto de vectores de alta dimensionalidad). Encontrar las tendencias mayores que describen los datos.

**Método:** se basa en el análisis de la varianza entre componentes (o factores).

**Caract.:** no requiere conocimiento previo sobre las etiquetas de clase de cada vector (no-supervisado).

**Matemática:** es una transformación lineal ortogonal que proyecta los datos de entrada a un nuevo sistema de coordenadas, tal que la que proyección con mayor varianza es representada ahora por la primera coordenada, la proyección con la segunda mayor varianza es representada por la segunda coordenada, y así sucesivamente.

# PCA: Principal Analysis Components

**Propósito:** reducir la dimensionalidad de los datos (conjunto de vectores de alta dimensionalidad). Encontrar las tendencias mayores que describen los datos.

**Método:** se basa en el análisis de la varianza entre componentes (o factores).

**Caract.:** no requiere conocimiento previo sobre las etiquetas de clase de cada vector (no-supervisado).

**Matemática:** es una transformación lineal ortogonal que proyecta los datos de entrada a un nuevo sistema de coordenadas, tal que la que proyección con mayor varianza es representada ahora por la primera coordenada, la proyección con la segunda mayor varianza es representada por la segunda coordenada, y así sucesivamente.



# PCA: Principal Analysis Components

**Propósito:** reducir la dimensionalidad de los datos (conjunto de vectores de alta dimensionalidad). Encontrar las tendencias mayores que describen los datos.

**Método:** se basa en el análisis de la varianza entre componentes (o factores).

**Caract.:** no requiere conocimiento previo sobre las etiquetas de clase de cada vector (no-supervisado).

**Matemática:** es una transformación lineal ortogonal que proyecta los datos de entrada a un nuevo sistema de coordenadas, tal que la que proyección con mayor varianza es representada ahora por la primera coordenada, la proyección con la segunda mayor varianza es representada por la segunda coordenada, y así sucesivamente.

# PCA: Principal Analysis Components

**Propósito:** reducir la dimensionalidad de los datos (conjunto de vectores de alta dimensionalidad). Encontrar las tendencias mayores que describen los datos.

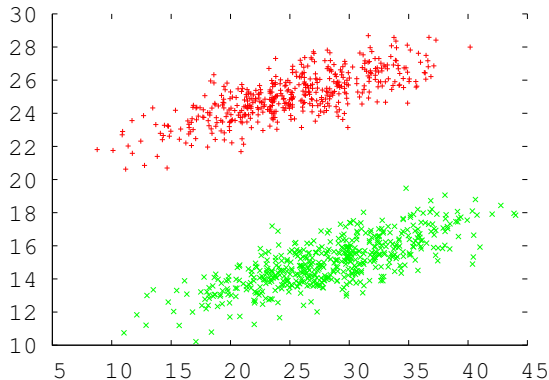
**Método:** se basa en el análisis de la varianza entre componentes (o factores).

**Caract.:** no requiere conocimiento previo sobre las etiquetas de clase de cada vector (no-supervisado).

**Matemática:** es una transformación lineal ortogonal que proyecta los datos de entrada a un nuevo sistema de coordenadas, tal que la proyección con mayor varianza es representada ahora por la primera coordenada, la proyección con la segunda mayor varianza es representada por la segunda coordenada, y así sucesivamente.

# Ejemplo PCA(1)

Ejemplo de 2 clases en el espacio bi-dimensional:



$$\mu_A = (25 \quad 25)$$

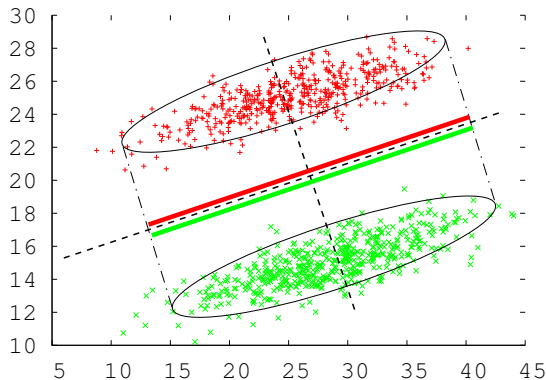
$$\sigma_A = \begin{pmatrix} 32 & 6 \\ 6 & 2 \end{pmatrix}$$

$$\mu_B = (28,5 \quad 15)$$

$$\sigma_B = \begin{pmatrix} 32 & 6 \\ 6 & 2 \end{pmatrix}$$

# Ejemplo PCA(2)

Ejemplo de 2 clases en el espacio bi-dimensional:



$$\mu_A = (25 \quad 25)$$

$$\sigma_A = \begin{pmatrix} 32 & 6 \\ 6 & 2 \end{pmatrix}$$

$$\mu_B = (28,5 \quad 15)$$

$$\sigma_B = \begin{pmatrix} 32 & 6 \\ 6 & 2 \end{pmatrix}$$

# PCA: Derivación por Método de la Covarianza

Sea  $\mathbf{X} = [X_1, X_2, \dots, X_n]$ ,  $X_i \in \mathbb{R}^d$ .

Se busca una matriz ortogonal  $\mathbf{W}$  (lo que implica  $\mathbf{W}^{-1} = \mathbf{W}^T$ ), tal que:

$$\mathbf{Y} = \mathbf{W}(\mathbf{X} - \mu_{\mathbf{X}}) \quad \Sigma_{\mathbf{Y}} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_d \end{bmatrix}$$

donde:  $\mathbf{Y} = [Y_1, Y_2, \dots, Y_n]$ ,  $Y_i \in \mathbb{R}^d$

Sabiendo que:  $\mu_{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n X_i$   $\Sigma_{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_{\mathbf{X}})(X_i - \mu_{\mathbf{X}})^T$ , se llega a:

$$\Sigma_{\mathbf{X}} \mathbf{W} = \mathbf{W} \Sigma_{\mathbf{Y}} \quad \Rightarrow \quad \Sigma_{\mathbf{X}} \mathbf{W} \mathbf{j} = \lambda_j \mathbf{W} \mathbf{j} \quad (\text{vectores propios})$$

# LDA: Linear Discriminant Analysis

**Propósito:** proyectar los datos (vectores de gran dimensionalidad) en un un espacio de menor dimensionalidad preservando al máximo la información de clases de los mismos.

**Método:** basado en el análisis de las varianzas entre-clases e intra-clases de los componentes (o factores).

**Caract.:** requiere conocimiento previo sobre las etiquetas de clase de cada vector (supervisado).

**Matemática:** al igual que en PCA, LDA es una transformación lineal ortogonal que proyecta los datos de entrada a un nuevo sistema de coordenadas, tal que los datos proyectados tengan máxima varianza entre-clases y mínima varianza intra-clases.

# LDA: Linear Discriminant Analysis

**Propósito:** proyectar los datos (vectores de gran dimensionalidad) en un un espacio de menor dimensionalidad preservando al máximo la información de clases de los mismos.

**Método:** basado en el análisis de las varianzas entre-clases e intra-clases de los componentes (o factores).

**Caract.:** requiere conocimiento previo sobre las etiquetas de clase de cada vector (supervisado).

**Matemática:** al igual que en PCA, LDA es una transformación lineal ortogonal que proyecta los datos de entrada a un nuevo sistema de coordenadas, tal que los datos proyectados tengan máxima varianza entre-clases y mínima varianza intra-clases.

# LDA: Linear Discriminant Analysis

**Propósito:** proyectar los datos (vectores de gran dimensionalidad) en un un espacio de menor dimensionalidad preservando al máximo la información de clases de los mismos.

**Método:** basado en el análisis de las varianzas entre-clases e intra-clases de los componentes (o factores).

**Caract.:** requiere conocimiento previo sobre las etiquetas de clase de cada vector (supervisado).

**Matemática:** al igual que en PCA, LDA es una transformación lineal ortogonal que proyecta los datos de entrada a un nuevo sistema de coordenadas, tal que los datos proyectados tengan máxima varianza entre-clases y mínima varianza intra-clases.



# LDA: Linear Discriminant Analysis

**Propósito:** proyectar los datos (vectores de gran dimensionalidad) en un un espacio de menor dimensionalidad preservando al máximo la información de clases de los mismos.

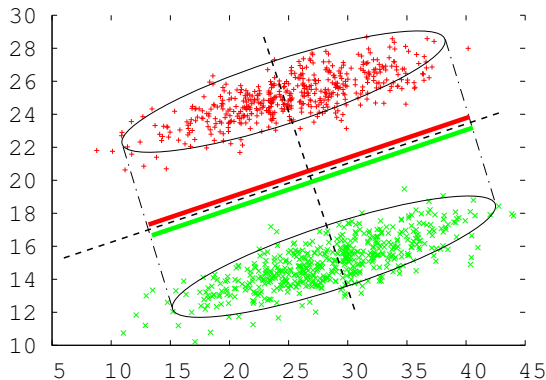
**Método:** basado en el análisis de las varianzas entre-clases e intra-clases de los componentes (o factores).

**Caract.:** requiere conocimiento previo sobre las etiquetas de clase de cada vector (supervisado).

**Matemática:** al igual que en PCA, LDA es una transformación lineal ortogonal que proyecta los datos de entrada a un nuevo sistema de coordenadas, tal que los datos proyectados tengan máxima varianza entre-clases y mínima varianza intra-clases.

# Ejemplo LDA(1)

Ejemplo de 2 clases en el espacio bi-dimensional:



$$\mu_A = (25 \quad 25)$$

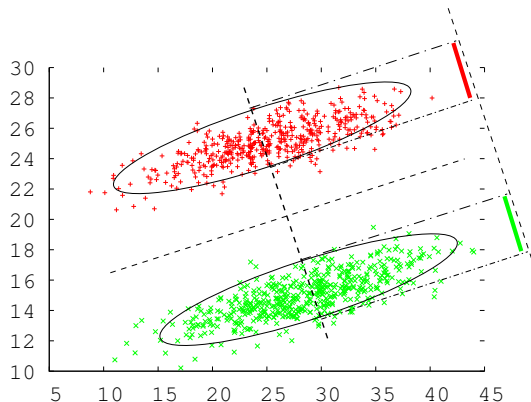
$$\sigma_A = \begin{pmatrix} 32 & 6 \\ 6 & 2 \end{pmatrix}$$

$$\mu_B = (28,5 \quad 15)$$

$$\sigma_B = \begin{pmatrix} 32 & 6 \\ 6 & 2 \end{pmatrix}$$

# Ejemplo LDA(2)

Ejemplo de 2 clases en el espacio bi-dimensional:



$$\mu_A = (25 \quad 25)$$

$$\sigma_A = \begin{pmatrix} 32 & 6 \\ 6 & 2 \end{pmatrix}$$

$$\mu_B = (28,5 \quad 15)$$

$$\sigma_B = \begin{pmatrix} 32 & 6 \\ 6 & 2 \end{pmatrix}$$

# LDA: Formulación Matemática

Sea:

- $n$  : número total de muestras.
- $n_j$  : número total de muestras de la clase  $j$ .
- $N_j$  : conjunto de índices de la clase  $j$ .
- $c$  : número de total de clases.
- $d$  : dimensión de las muestras.
- $\mathbf{x}_i$  : muestra  $i$ -ésima,  $\mathbf{x}_i \in \mathbb{R}^d$ .
- $\mu$  : media total,  $\mu = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$ .
- $\mu_j$  : media de la clase  $j$ ,  $\mu_j = \frac{1}{n_j} \sum_{i \in N_j} \mathbf{x}_i$ .
- $\mathbf{w}_k$  : vector de proyección,  $\mathbf{w}_k \in \mathbb{R}^d$ .

Buscamos una matriz  $\mathbf{W}$ ,  $\mathbf{y} = \mathbf{W}\mathbf{x}$ , que cumpla:  $\mathbf{w}_k = \arg \max_{\mathbf{w}_k} \frac{\mathbf{w}_k^T \Sigma_E \mathbf{w}_k}{\mathbf{w}_k^T \Sigma_I \mathbf{w}_k}$

donde:  $\Sigma_E = \frac{1}{c} \sum_{j=1}^c (\mu_j - \mu)(\mu_j - \mu)^T$      $\Sigma_I = \frac{1}{n} \sum_{j=1}^c \sum_{i \in N_j} (\mu_j - \mathbf{x}_i)(\mu_j - \mathbf{x}_i)^T$

Ésto es equiv. a resolver:

$$(\Sigma_I^{-1} \Sigma_E - \lambda_k I) \mathbf{w}_k = 0 \quad (\text{vectores propios})$$

# Temario

## 1 Reconocimiento de Escritura Off-Line

- Introducción
- Visión Global del Sistema
- Preproceso de Texto Manuscrito - Página
- Eliminación del Fondo y Filtrado del Ruido
- Corrección de “Skew”
- Detección y Segmentado de Líneas
- Preproceso de Texto Manuscrito - Línea
- Corrección y Normalización de Estilos - Ilustración
- Corrección de “Slope”
- Corrección de “Slant”
- Normalización del Tamaño
- Extracción de Características
- Extracción de Características Utilizadas
- Reducción de la Dimensionalidad
- **Entrenamiento de HMMs - Generalidades**

# Entrenamiento de HMMs - Generalidades

Para el entrenamiento de los HMMS:

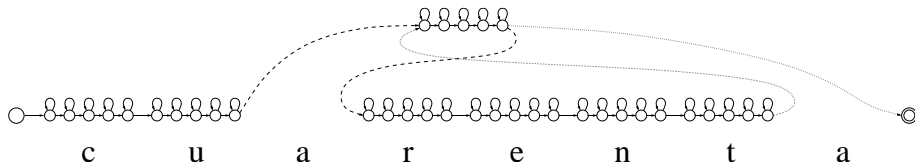
- No se requiere una segmentación explícita de muestras de caracteres para entrenar las clases de caracteres.
- Se utiliza el algoritmo “Forward-Backward” o “Baum-welch” (caso particular del “EM”).
- La topología de HMMs utilizada es la izquierda-derecha.

# Entrenamiento de HMMs sin Segmentación(1)

- *Entrenamiento integrado* conocido como “embedded training Baum-Welch”.
- Parámetros de HMMs entrenados a partir de las transcripciones y secuencias de vectores de caract. de las imágenes de líneas de texto.
- No requiere segmentar previamente la imagen de línea de texto, en caracteres para entrenar su respectivos HMMs.

# Entrenamiento de HMMs sin Segmentación(2)

- Para cada muestra (S.V.C.) se construye “temporalmente” un modelo HMM de frase lineal *izquierda-derecha* formado por concatenación de los modelos HMMs de caracteres, correspondiente a la transcripción.





# Definición de HMM

Formalmente un HMM (continuo)  $\mathcal{M}$  es una máquina de estados finitos definida por la séxtupla  $(Q, I, F, X, a, b)$ , donde:

- $Q$  es un conjunto finito de estados, que incluye el estado inicial  $I \in Q$  y un estado final  $F \in Q$ .
- $X$  es un espacio real  $d$ -dimensional de observaciones:  $X \subseteq \mathbb{R}^d$ .
- $a : (Q - \{F\}) \times (Q - \{I\}) \rightarrow [0, 1]$  es una función de distribución de probabilidad de transición entre estados, tal que:

$$\sum_{q_j \in (Q - \{I\})} a(q_i, q_j) = 1 \quad \forall q_i \in (Q - \{F\})$$

- $b : (Q - \{I, F\}) \times X \rightarrow [0, 1]$  es una función de densidad de probabilidad de emitir un vector  $\vec{x} \in X$  en un estado  $q_i \in Q$ , tal que:

$$\int_{\vec{x} \in X} b(q_i, \vec{x}) d\vec{x} = 1 \quad \forall q_i \in (Q - \{I, F\})$$

# Prob. de Emisión de un HMM(1)

Dado:

- $\mathbf{x} = \langle x_1 x_2 \dots x_T \rangle$  una secuencia de vectores de características, con  $x_i \in \mathbb{R}^d$ .
- $S = \{\mathbf{s} = \langle s_1 s_2 \dots s_T \rangle : s_t = q_i \in (Q - \{I, F\}), 1 \leq i \leq |Q| - 2\}$  un conjunto de secuencias de derivación de estados asociadas a la secuencia  $\mathbf{x}$ .

La probabilidad que la cadena  $\mathbf{x}$  sea generada por el modelo  $\mathcal{M}$  es:

$$P(\mathbf{x}|\mathcal{M}) = \sum_{\mathbf{s} \in S} \left( \prod_{t=1}^T a(s_{t-1}, s_t) b(s_t, x_t) \right) a(s_T, F)$$

donde  $s_0$  es igual al estado inicial  $I$ .

# Prob. de Emisión de un HMM(2)

La función  $b(s_t = q_j, \mathbf{x}_t) = b_j(\mathbf{x}_t)$  es definida por:

$$b_j(\mathbf{x}_t) = \sum_{m=1}^M c_{jm} b_{jm}(\mathbf{x}_t)$$

donde:

$$b_{jm}(\mathbf{x}_t) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_{jm}|}} e^{-\frac{1}{2}(\mathbf{x}_t - \mu_{jm})^t \Sigma_{jm}^{-1} (\mathbf{x}_t - \mu_{jm})}$$

cumpléndose además que  $\sum_{m=1}^M c_{jm} = 1$ .

# Topología del HMM

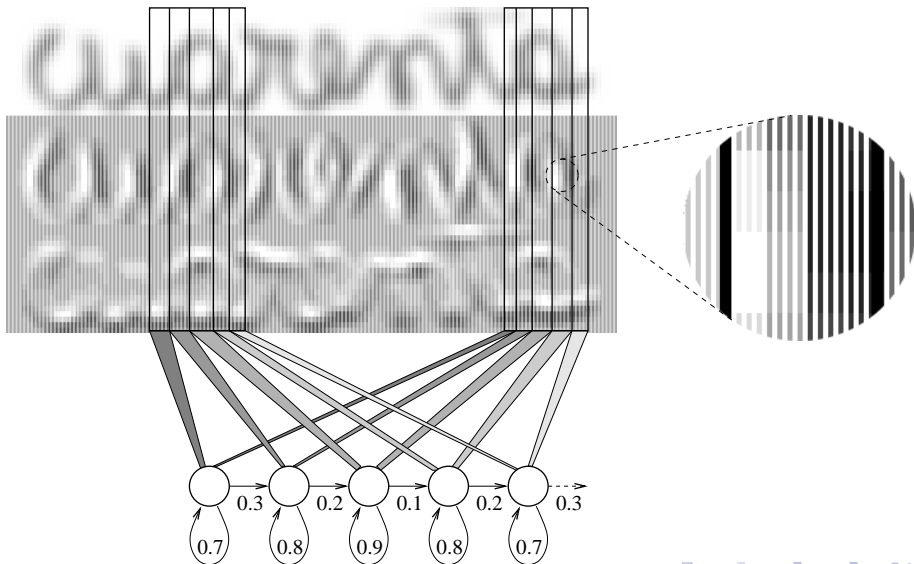
- Topología *izquierda-derecha* con transiciones de los estados a sí mismos y al siguiente.
- Número adecuado de estados para modelar un cierto carácter o conjunto de caracteres, recae sobre la variabilidad horizontal.
- Número adecuado de Gaussianas en una mixtura depende, de la variabilidad vertical que se espera en cada estado.
- Matrices de covarianza diagonales de las Gaussianas (reduce drásticamente el número de parámetros a entrenar).
- Fijamos el mismo número de estados para todos los modelos HMMs de caracteres.
- Fijamos el mismo número de Gaussianas para la mezcla de todos los estados de todos los HMMs de caracteres.

# Topología y Modelado(1)

La topología de los HMMs guarda una estrecha relación con lo que modela.

- Los dos procesos estocásticos implícitos: uno describe las partes constituyentes del objeto (carácter) y otro las variaciones de dichas partes.
- Cada estado modela columnas con perfiles similares, cuyas *pequeñas diferencias* a su vez, se modelan por la mixtura de densidades de Gaussianas dentro de cada estado.

# Topología y Modelado(2)



# Inicialización e Iteración

- EM “Baum-Welch” no siempre garantiza la mejor estimación de parámetros que maximicen la probabilidad de generar las muestras dadas.
- Una inicialización no adecuada de los parámetros puede llevar al EM a converger a un mínimo local.
- La inicialización de los  $\mu$  y  $\Sigma$  de todas las Gaussianas se realiza mediante el cómputo global  $\mu$  y  $\Sigma$  de todas las muestras  $\mathbf{x}$ .
- Entrenamos HMMs con  $N$  Gaussianas por estado, donde  $N = 1, 2, 4, \dots, 64$ .
- Número de iteraciones de re-estimación de parámetros (Baum-Welch) aplicadas, varía entre 3 y 4.
- Los modelos HMMs con  $N$  Gaussianas son creados e inicializados a partir de los anteriores con  $N/2$  Gaussianas.

# Número de Parámetros a Entrenar

Se asume:

- Mismo número de estados para todos los modelos HMMs.
- Mismo número  $L$  de Gaussianas en la mezcla para todos los estados de un modelo HMM.
- Matrices de covarianza  $\Sigma_{jl}$  diagonales.

$$P = M[T + SL(2d + 1)]$$

donde:

*P*: número de parámetros totales a entrenar.

*M*: número de modelos (HMMs).

*S*: número de estados del modelo.

*T*: número de transiciones del modelo.

*L*: número de Gaussianas por estado.

*d*: dimensión de los vectores de las muestras.



# Entrenamiento de HMMs - Forward

Función “Forward” denota la probabilidad de que un proceso de Markov  $\mathcal{M}$  se encuentre en el estado  $q_j$  en el tiempo  $t$ , y haya emitido la secuencia de vectores  $x_1 x_2 \dots x_t$ :

$$p(s_t = q_j, x_1^t | \mathcal{M}) = \alpha_j(t) = \begin{cases} a_{0j} b_j(x_1) & t = 1 \\ \left[ \sum_{i=1}^{N-1} \alpha_i(t-1) a_{ij} \right] b_j(x_t) & 1 < t \leq T \end{cases}$$

donde:

$$\alpha_0(1) = 1$$

$$P(x_1^T | \mathcal{M}) = \sum_{i=1}^{N-1} \alpha_i(T) a_{iN}$$

$N$ : número de estados del modelo  $\mathcal{M}$

$T$ : longitud de la muestra

# Entrenamiento de HMMs - Backward

Función “Backward” denota la probabilidad de que un proceso de Markov  $\mathcal{M}$  vaya a emitir la secuencia de vectores  $x_{t+1} x_{t+2} \dots x_T$ , sabiendo que para el tiempo  $t$  se encuentra en el estado  $q_i$ :

$$p(x_{t+1}^T | s_t = q_i, \mathcal{M}) \beta_i(t) = \begin{cases} a_{iN} & t = T \\ \sum_{j=1}^{N-1} a_{ij} b_j(x_{t+1}) \beta_j(t+1) & 1 \leq t < T \end{cases}$$

donde:

$$\beta_N(T) = 1$$

$$P(x_1^T | \mathcal{M}) = \sum_{j=1}^{N-1} a_{0j} b_j(x_1) \beta_j(1)$$

$N$ : número de estados del modelo  $\mathcal{M}$

$T$ : longitud de la muestra