Trabajo Práctico 3: OCR+SVD

Métodos Numéricos

Laboratorio

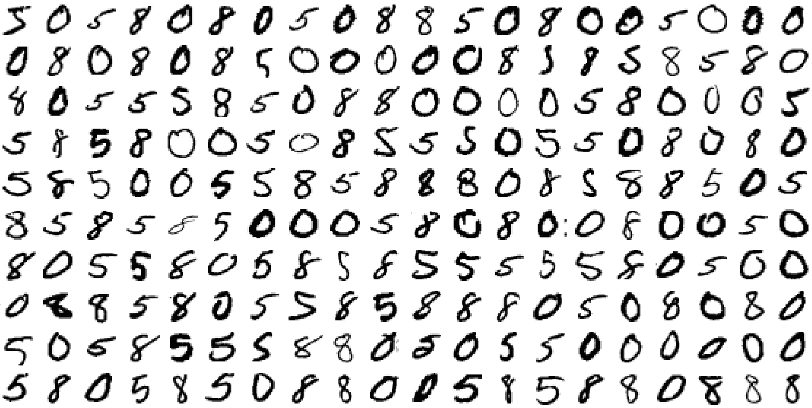
Departmento de Computación
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

1er cuatrimestre 2013

Trabajo Práctico 3

Reconocimiento de dígitos manuscritos usando la descomposición en valores singulares

- Datos: base de datos (etiquetada) de imágenes de dígitos manuscritos
- Objetivo: dado una nueva imagen ¿a qué dígito se corresponde?



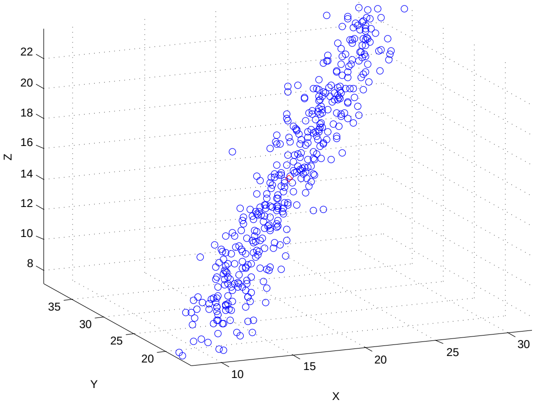
Pequeño ejemplo

Datos en $\mathbb{R}^3$

Sea $x^{(1)}, x^{(2)}, \dots, x^{(n)}$ una secuencia de n datos, con $x^{(i)} \in \mathbb{R}^3$

$$X = \begin{pmatrix} 10.898 & 17.477 & 7.4443 \\ 11.018 & 18.777 & 9.1042 \\ 10.157 & 15.845 & 9.6744 \\ 10.065 & 16.378 & 7.8452 \\ 11.072 & 18.227 & 9.0744 \\ \vdots & \vdots & \vdots \\ 28.889 & 38.937 & 22.877 \end{pmatrix}$$

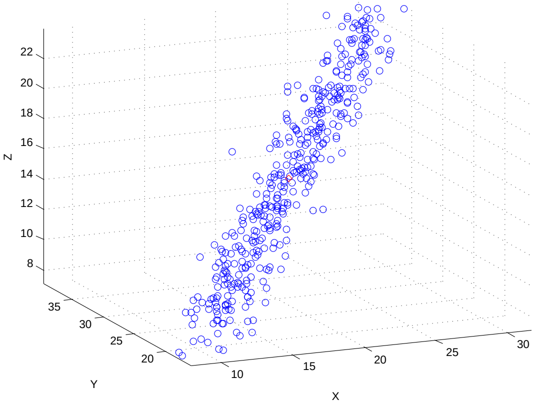
$$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ x_1 & x_2 & x_3 \end{array}$$



Pequeño ejemplo
Datos en $\mathbb{R}^3$

Sea $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ una secuencia de n datos, con $x^{(i)} \in \mathbb{R}^3$

$$X = \begin{pmatrix} x^{(1)\,t} \\ x^{(2)\,t} \\ x^{(3)\,t} \\ x^{(4)\,t} \\ x^{(5)\,t} \\ \vdots \\ x^{(n)\,t} \end{pmatrix}$$

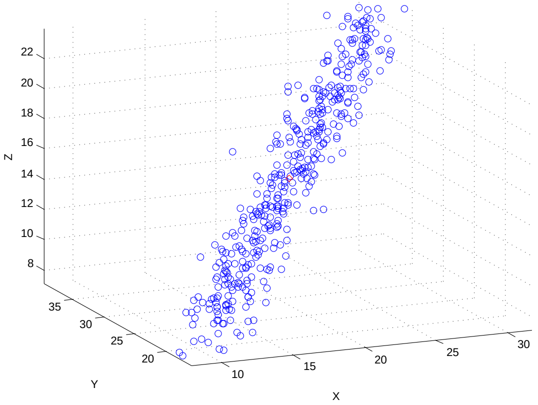


Pequeño ejemplo

Datos en $\mathbb{R}^3$

Sea $x^{(1)}, x^{(2)}, \dots, x^{(n)}$ una secuencia de n datos, con $x^{(i)} \in \mathbb{R}^3$

$$X = \begin{pmatrix} x_1^{(1)\,t} & x_2^{(1)\,t} & x_3^{(1)\,t} \\ x_1^{(2)\,t} & x_2^{(2)\,t} & x_3^{(2)\,t} \\ \vdots & \vdots & \vdots \\ x_1^{(n)\,t} & x_2^{(n)\,t} & x_3^{(n)\,t} \end{pmatrix}$$



Pequeño ejemplo

Datos en $\mathbb{R}^3$

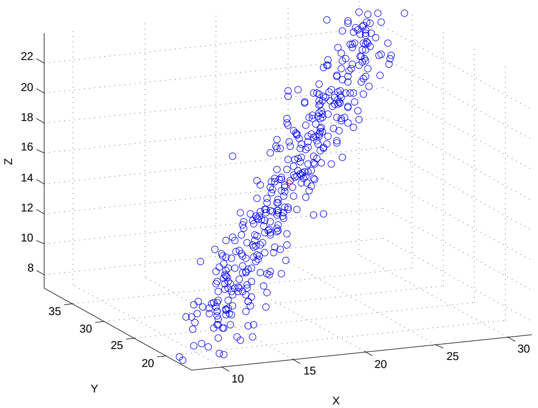
Sea $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ una secuencia de n datos, con $x^{(i)} \in \mathbb{R}^3$

$$X = \begin{pmatrix} 10.898 & 17.477 & 7.4443 \\ 11.018 & 18.777 & 9.1042 \\ 10.157 & 15.845 & 9.6744 \\ 10.065 & 16.378 & 7.8452 \\ 11.072 & 18.227 & 9.0744 \\ \vdots & \vdots & \vdots \\ 28.889 & 38.937 & 22.877 \end{pmatrix}$$

$$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ x_1 & x_2 & x_3 \end{array}$$

$\bar{x}_1 = 19.96, \bar{x}_2 = 27.902, \bar{x}_3 = 15.581$

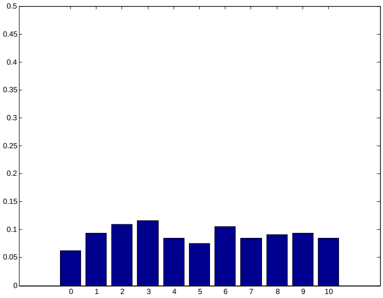
$\sigma_{x_1} = 35.26, \sigma_{x_2} = 5.93, \sigma x_3 = 4.13$

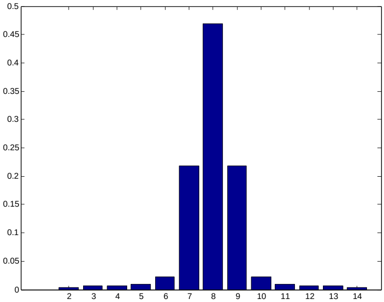


Varianza de una variable x_k:

$$\sigma_{x_k} = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_k^{(i)} - \bar{x}_k\right)^2$$

Varianza





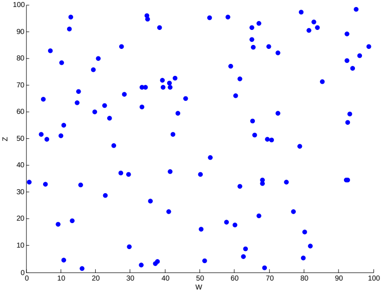
$\sigma = 9.49$

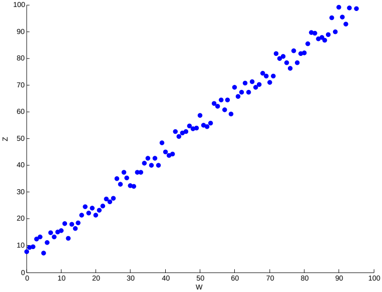
$\sigma = 1.52$

Varianza de n observaciones de una variable

$$\sigma_{X_1} = \frac{1}{n-1} \sum_{i=1}^{n} (x_1^{(i)} - \bar{x}_1)^2$$

Covarianza





$\sigma_{W,Z} = 93.07$ $\qquad$ $\sigma_{W,Z} = 847.71$

$$\sigma_{W,Z} = \frac{1}{n-1} \sum_{i=1}^{n} (w^{(i)} - \bar{w})(z^{(i)} - \bar{z})$$

Variables con mayor covarianza inducen presencia de cierta de dependencia o relación

Covarianza entre variables

Dadas de n observaciones de dos variables x_k y x_j:

$$\sigma_{x_k,x_j} = \frac{1}{n-1}\sum_{i=1}^{n}(x_k^{(i)} - \bar{x}_k)(x_j^{(i)} - \bar{x}_j) = \frac{1}{n-1}(x_k - \bar{x}_k)^t(x_j - \bar{x}_j)$$

Matriz de Covarianza

Sea $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ una secuencia de n datos, con $x^{(i)} \in \mathbb{R}^3$

$$X = \begin{array}{c} \begin{pmatrix} 10.898 - \bar{x}_1 & 17.477 - \bar{x}_2 & 7.4443 - \bar{x}_3 \\ 11.018 - \bar{x}_1 & 18.777 - \bar{x}_2 & 9.1042 - \bar{x}_3 \\ 10.157 - \bar{x}_1 & 15.845 - \bar{x}_2 & 9.6744 - \bar{x}_3 \\ 10.065 - \bar{x}_1 & 16.378 - \bar{x}_2 & 7.8452 - \bar{x}_3 \\ 11.072 - \bar{x}_1 & 18.227 - \bar{x}_2 & 9.0744 - \bar{x}_3 \\ \vdots & \vdots & \vdots \\ 28.889 - \bar{x}_1 & 38.937 - \bar{x}_2 & 22.877 - \bar{x}_3 \end{pmatrix} \\ \begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ x_1 & x_2 & x_3 \end{array} \end{array}$$

$$M_X = \begin{pmatrix} \sigma_{x_1,x_1} & \sigma_{x_1,x_2} & \sigma_{x_1,x_3} \\ \sigma_{x_2,x_1} & \sigma_{x_2,x_2} & \sigma_{x_2,x_3} \\ \sigma_{x_3,x_1} & \sigma_{x_3,x_2} & \sigma_{x_3,x_3} \end{pmatrix}$$

$$M_X = \frac{1}{n-1} X^t X$$

¿Cómo expresar *mejor* nuestros datos?

Cambio de base: $\widehat{X}^t = PX^t$

- Disminuir la redundancia en los datos ($\downarrow$ covarianza)

Cómo?

- Diagonalizar la matriz de covarianza

¿Cómo expresar *mejor* nuestros datos?

Cambio de base: $\widehat{X}^t = PX^t$

- Disminuir la redundancia en los datos ($\downarrow$ covarianza)

Cómo?

- Diagonalizar la matriz de covarianza

¿Cómo expresar *mejor* nuestros datos?

Cambio de base: $\widehat{X}^t = PX^t$

Sea P ortogonal y $M_{\widehat{X}}$ la matriz de covarianza de $\widehat{X}$

$$
\begin{aligned}
M_{\widehat{X}} &= \frac{1}{n-1}\widehat{X}^t\widehat{X} \\
&= \frac{1}{n-1}(PX^t)(XP^t) \\
&= P\,\frac{X^tX}{n-1}\,P^t \\
M_{\widehat{X}} &= P\,M_X\,P^t
\end{aligned}
$$

Toda matriz simétrica es diagonalizable

Teorema

Si $B \in \mathbb{R}^{n\times n}$ es simétrica entonces existe una base ortonormal de autovectores $\{v_1, \dots, v_n\}$ asociados a autovalores $\lambda_1, \dots, \lambda_n$

Entonces: $B \left(\begin{array}{c|c|c} v_1 & \dots & v_n \end{array} \right) = \left(\begin{array}{c|c|c} Bv_1 & \dots & Bv_n \end{array} \right) =$

$$\left(\begin{array}{c|c|c} \lambda_1 v_1 & \dots & \lambda_n v_n \end{array} \right) = \left(\begin{array}{c|c|c} v_1 & \dots & v_n \end{array} \right) \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

$$\Longrightarrow B = \left(\begin{array}{c|c|c} v_1 & \dots & v_n \end{array} \right) \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \left(\begin{array}{c|c|c} v_1 & \dots & v_n \end{array} \right)^t$$

¿Cómo expresar *mejor* nuestros datos?

Cambio de base: $\widehat{X}^t = P X^t \Longrightarrow M_{\widehat{X}} = P\, M_X\, P^t$

M_X es simétrica $\Rightarrow$ tiene una base ortonormal de autovectores $\Rightarrow$ es diagonalizable $\Rightarrow M_X = V D V^t$

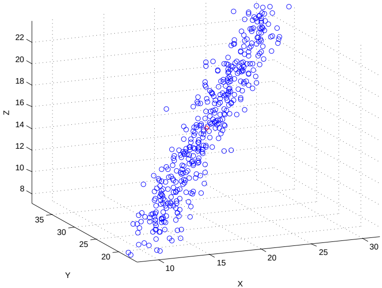
$$\begin{aligned} M_{\widehat{X}} &= P\, M_X\, P^t \\ &= P(VDV^t)P^t \qquad \text{elijo } P := V^t \\ &= V^t(VDV^t)V \\ &= (V^tV)D(V^tV) \\ &= D \end{aligned}$$

Con esta elección de P se diagonaliza la matriz de covarianzas de los datos transformados.

Matriz de Covarianza

Sea $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ una secuencia de n datos, con $x^{(i)} \in \mathbb{R}^3$

$$X = \begin{array}{c} \begin{pmatrix} 10.898 - \bar{x}_1 & 17.477 - \bar{x}_2 & 7.4443 - \bar{x}_3 \\ 11.018 - \bar{x}_1 & 18.777 - \bar{x}_2 & 9.1042 - \bar{x}_3 \\ 10.157 - \bar{x}_1 & 15.845 - \bar{x}_2 & 9.6744 - \bar{x}_3 \\ 10.065 - \bar{x}_1 & 16.378 - \bar{x}_2 & 7.8452 - \bar{x}_3 \\ 11.072 - \bar{x}_1 & 18.227 - \bar{x}_2 & 9.0744 - \bar{x}_3 \\ \vdots & \vdots & \vdots \\ 28.889 - \bar{x}_1 & 38.937 - \bar{x}_2 & 22.877 - \bar{x}_3 \end{pmatrix} \\ \begin{array}{ccc} \uparrow \qquad\qquad & \uparrow & \qquad\qquad \uparrow \\ x_1 \qquad\qquad & x_2 & \qquad\qquad x_3 \end{array} \end{array}$$



$$M_X = \begin{pmatrix} 35 & 34.27 & 23.67 \\ 34.274 & 35.26 & 23.80 \\ 23.675 & 23.80 & 17.43 \end{pmatrix}$$

$$M_X = \frac{1}{n-1} X^t X$$

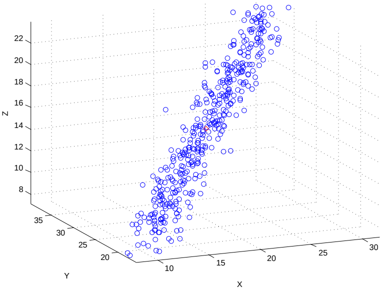
Diagonalización Matriz de Covarianza

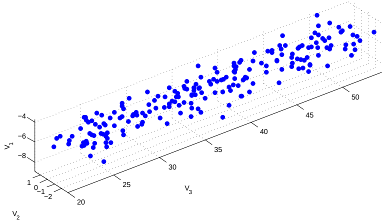
$$M_X = \begin{pmatrix} 35 & 34.27 & 23.67 \\ 34.27 & 35.26 & 23.80 \\ 23.67 & 23.80 & 17.43 \end{pmatrix} =$$

$$= \underbrace{\begin{pmatrix} 0.62 & 0.46 & 0.63 \\ -0.75 & 0.14 & 0.63 \\ 0.20 & -0.87 & 0.44 \end{pmatrix}}_{V} \underbrace{\begin{pmatrix} 0.85 & 0 & 0 \\ 0 & 0.97 & 0 \\ 0 & 0 & 85.8 \end{pmatrix}}_{D} \underbrace{\begin{pmatrix} 0.62 & -0.75 & 0.20 \\ 0.46 & 0.14 & -0.87 \\ 0.63 & 0.63 & 0.44 \end{pmatrix}}_{V^t}$$

Transformación de los datos

Cambio de base ($P = V^t$): $\widehat{X}^t = V^t X^t$

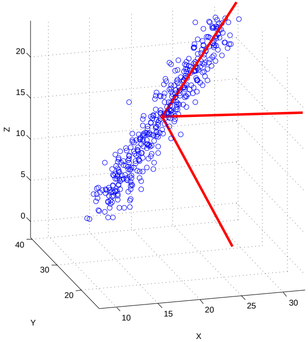




$$M_X = \begin{pmatrix} 35 & 34.27 & 23.67 \\ 34.27 & 35.26 & 23.80 \\ 23.67 & 23.80 & 17.43 \end{pmatrix} \qquad M_{\widehat{X}} = \begin{pmatrix} 85.8 & 0 & 0 \\ 0 & 0.97 & 0 \\ 0 & 0 & 0.85 \end{pmatrix}$$

Propiedades del cambio de base

- disminuye redundancias
- minimiza el error
- captura la dinámica de interés de los datos (si se ordena autovalores según módulo)



Usando la descomposición en valores singulares

Consideremos la matriz de datos $A \in \mathbb{R}^{n \times m}$

- n muestras y m variables
- Supongamos que las muestras en A tienen media 0
- Definimos a $X = A/\sqrt{n-1}$
- Sea $X = U\Sigma V^t$ la descomposición SVD de X
- V es ortogonal con autovectores de $X^tX = (A^tA)/(n-1)$

Los columnas de V contienen los autovectores
de la matriz de covarianza

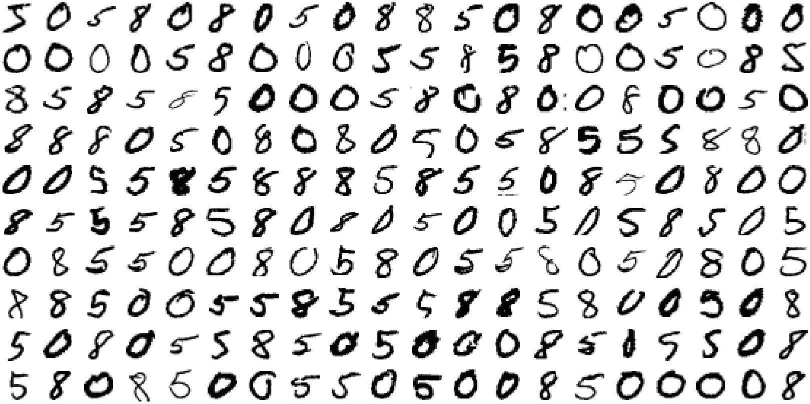
Reconociendo dígitos

Base de dates MNIST

- 6000 dígitos
- Cada dígito es una imagen de 28×28 píxeles

Reconociendo dígitos

Ejemplo con base de sólo imágenes de los dígitos 0, 5 y 8.



Reconociendo dígitos

Ejemplo con dígitos 0, 5 y 8.

- La base MNIST tiene 17195 imágenes de estos 3 dígitos
- $28 \times 28 = 784$ píxeles por imagen (cada pixel, una variable)
- Sea $x^{(1)}, x^{(2)}, \ldots, x^{(17195)}$ cada una de las imágenes dispuestas en un vector en $\mathbb{R}^{784}$
- Sea $\mu = \left(\sum_{i=1}^{17195} x^{(i)}\right) / 17195$ la imagen promedio.
- Definimos X como una matriz en $\mathbb{R}^{17195 \times 784}$
- El vector $\frac{(x^{(i)} - \mu)^t}{\sqrt{783 - 1}}$ en la fila i-ésima de X

Reconociendo dígitos

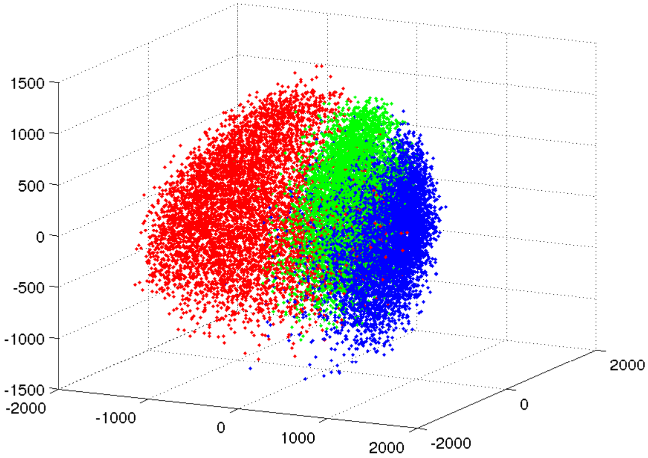
Ejemplo con dígitos 0, 5 y 8.

- Calculamos $X = U\Sigma V^t$ la descomposición SVD de X
- Las columnas de V tienen los autovectores de la matriz de covarianza $X^t X$
- Transformamos cada imagen: $\widehat{x}^{(i)} = V^t x^{(i)}$

Reconociendo dígitos

Ejemplo con dígitos 0, 5 y 8.

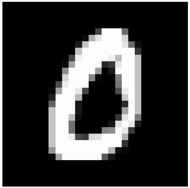
- Graficamos en $\mathbb{R}^3$ las primeras 3 componentes de cada $\widehat{x}^{(i)}$



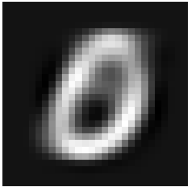
Reconociendo dígitos

Reconstrucción con menos componentes

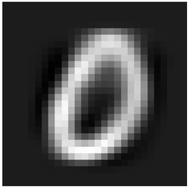
digito 0 $\in$ base



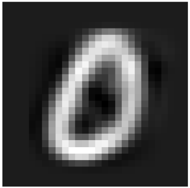
$K = 5$



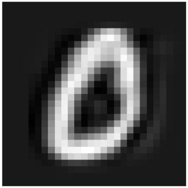
$K = 10$



$K = 30$



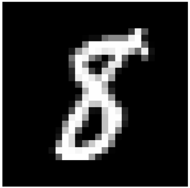
$K = 60$



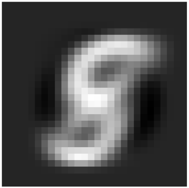
Reconociendo dígitos

Reconstrucción con menos componentes

digito 8 $\in$ base

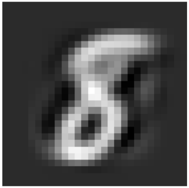


$K = 5$

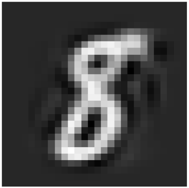


$K = 10$

$K = 30$



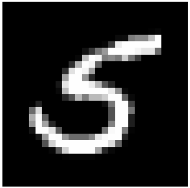
$K = 60$



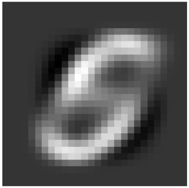
Reconociendo dígitos

Reconstrucción con menos componentes

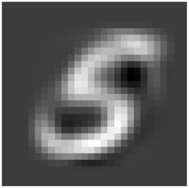
digito 5 $\in$ base



$K = 5$

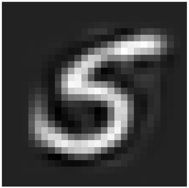


$K = 10$



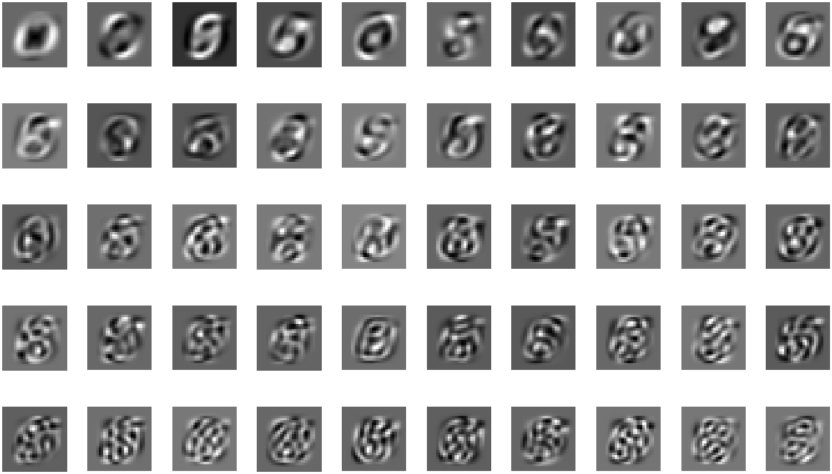
$K = 30$

$K = 60$



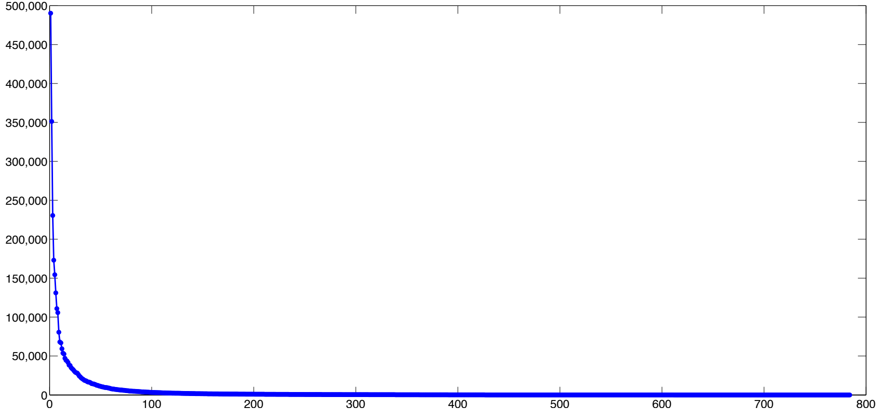
Reconociendo dígitos

‘AUTODIGITOS’ (autovectores de la matriz de covarianza)



Reconociendo dígitos

Varianzas (plot de la diagonal de la matriz Σ)



Cálculo de autovectores/autovalores

Métodos a utilizar

- Algoritmo QR
- Método de la potencia + deflación

Cálculo de autovectores/autovalores

Algoritmo QR

repeat

Factorizar $A_k = Q_k R_k$;

$A_{k+1} \leftarrow R_k Q_k$;

until A_{k+1} *sea triangular inferior* ;

- Notar que A_{k+1} y A_k son matrices semejantes:

$$\begin{aligned} A_{k+1} &= R_k Q_k \\ &= Q_k^t Q_k R_k Q_k \\ &= Q_k^{-1} A_k Q_k \end{aligned}$$

Cálculo de autovectores/autovalores

Algoritmo QR

- Luego de varias iteraciones
  $A_0 \curvearrowright A_1 \curvearrowright A_2 \curvearrowright \cdots \curvearrowright A_k \longrightarrow$matriz triangular inferior
- Si la matriz inicial es simétrica, el algoritmo QR tiende a una matriz diagonal
- Comenzando con $A_0 = A$

$$
\begin{aligned}
A_{k+1} &= R_k Q_k \\
&= Q_k^{-1} A_k Q_k \\
&= Q_k^{-1} Q_{k-1}^{-1} A_{k-1} Q_{k-1} Q_k \\
&= \underbrace{Q_k^{-1} Q_{k-1}^{-1} \ldots Q_0^{-1}}_{V^{-1}} A_0 \underbrace{Q_0 \ldots Q_{k-1} Q_k}_{V}
\end{aligned}
$$

- Obtenemos autovectores en V: $A = V A_{k+1} V^{-1}$

Cálculo de autovectores/autovalores

Algoritmo QR - Condición de parada

- Sea $A^{(k)} = (a_{ij}^{(k)})$ la matriz de $m \times m$ obtenida en la k-ésima iteración
- $A^{(k)}$ simétrica tiende a una matriz diagonal
- Definimos como condición de corte:

$$\sum_{i=2}^{m} \sum_{j=1}^{i-1} |a_{ij}^{(k)}| < \delta$$

  donde $\delta \in \mathbb{R}_+$ es un parámetro de la implementación
- Es decir, el método se interrumpe si la suma de los módulos de los elementos de la parte triangular inferior de $A^{(k)}$ sea inferior a δ

Cálculo de autovectores/autovalores

Método de la potencia + deflación

for $i = 1$ **to** k **do**

$v_i \leftarrow$MetodoPotencia(A) ;

$\lambda_i \leftarrow (v_i^t A v_i)/(v_i^t v_i)$;

$A \leftarrow A - \lambda_i v_i v_i^t$; `// deflación`

end

Propiedad: Sea $A \in \mathbb{R}^{n\times n}$ una matriz con autovalores distintos $\lambda_1 > \lambda_2 > \ldots > \lambda_n$ y una base ortonormal de autovectores.

- La matriz $A - \lambda_1 v_1 v_1^t$ tiene autovalores $0, \lambda_2, \ldots, \lambda_n$, con autovectores asociados $v_1, \ldots, v_n$

$$\boxed{\text{Este método solo funciona si tengo autovalores todos distintos}}$$