

ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

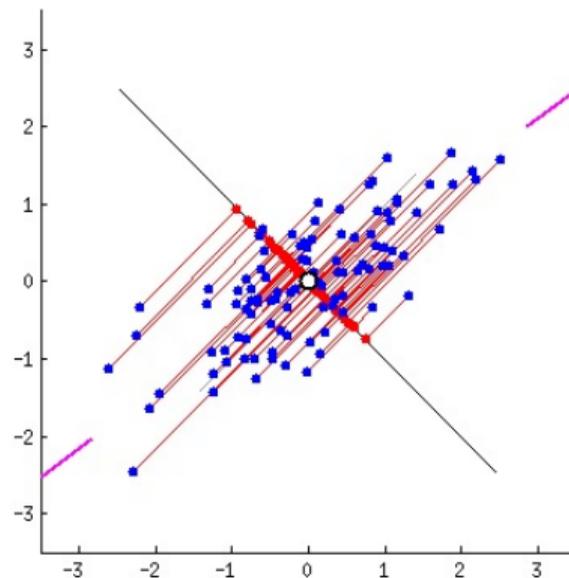
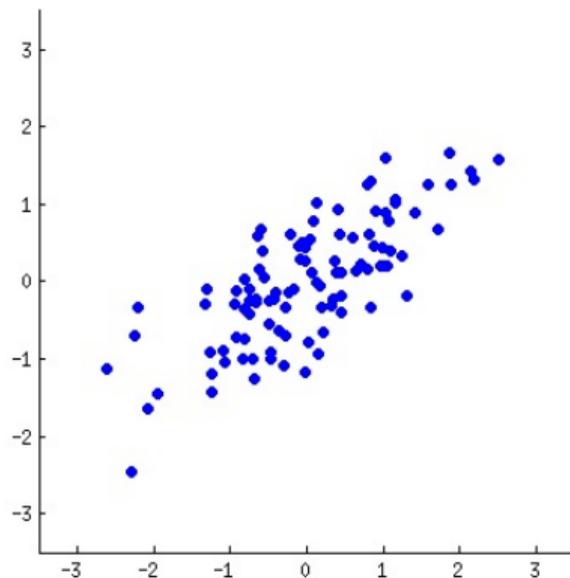
ALAN REYES-FIGUEROA
INTELIGENCIA ARTIFICIAL

(AULA 07) 03.FEBRERO.2025

Componentes principales

Objetivo: encontrar una estructura subyacente en los datos.

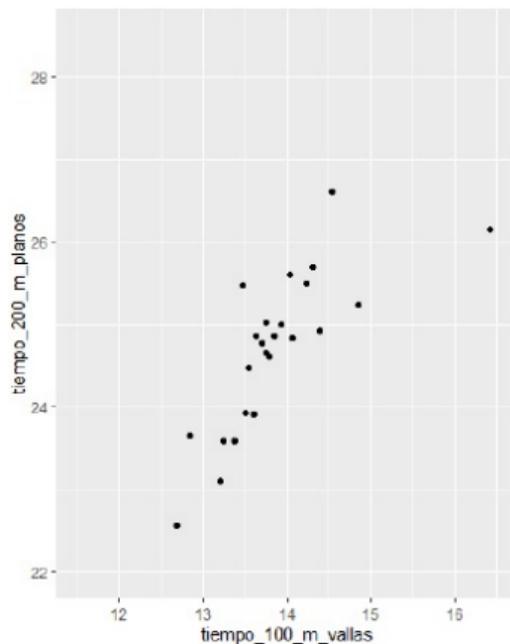
- Proyectar a un subespacio adecuado.



Componentes principales

Ejemplo: Atletismo, pruebas de 100m y 200m.

100m vallas	200m planos
12.69	22.56
12.85	23.65
13.2	23.1
13.61	23.92
13.51	23.93
13.75	24.65
13.38	23.59
13.55	24.48
13.63	24.86
13.25	23.59
13.75	25.03
13.24	23.59
13.85	24.87
13.71	24.78
13.79	24.61
13.93	25
13.47	25.47
14.07	24.83
14.39	24.92
14.04	25.61
14.31	25.69
14.23	25.5
14.85	25.23
14.53	26.61
16.42	26.16

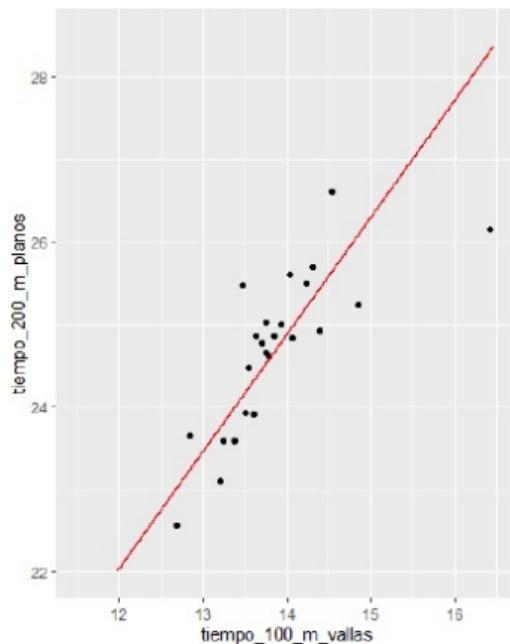


Se observa cierta estructura.

Componentes principales

Ejemplo: Atletismo, pruebas de 100m y 200m.

100m vallas	200m planos
12.69	22.56
12.85	23.65
13.2	23.1
13.61	23.92
13.51	23.93
13.75	24.65
13.38	23.59
13.55	24.48
13.63	24.86
13.25	23.59
13.75	25.03
13.24	23.59
13.85	24.87
13.71	24.78
13.79	24.61
13.93	25
13.47	25.47
14.07	24.83
14.39	24.92
14.04	25.61
14.31	25.69
14.23	25.5
14.85	25.23
14.53	26.61
16.42	26.16



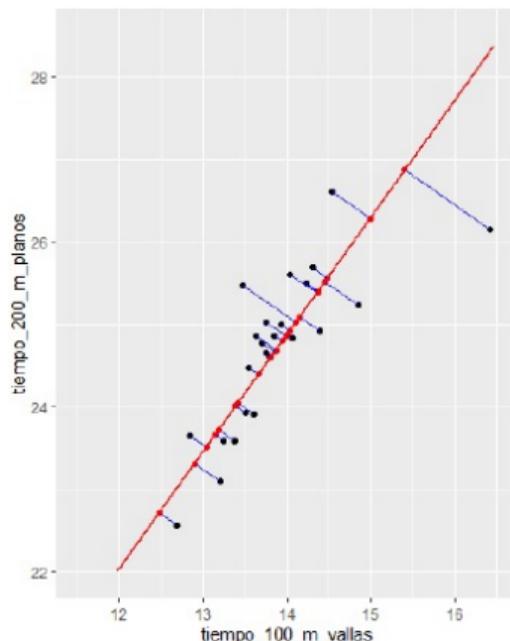
Se observa cierta estructura.

Karl Pearson (1901), describir con una recta.

Componentes principales

Ejemplo: Atletismo, pruebas de 100m y 200m.

100m vallas	200m planos
12.69	22.56
12.85	23.65
13.2	23.1
13.61	23.92
13.51	23.93
13.75	24.65
13.38	23.59
13.55	24.48
13.63	24.86
13.25	23.59
13.75	25.03
13.24	23.59
13.85	24.87
13.71	24.78
13.79	24.61
13.93	25
13.47	25.47
14.07	24.83
14.39	24.92
14.04	25.61
14.31	25.69
14.23	25.5
14.85	25.23
14.53	26.61
16.42	26.16



Se observa cierta estructura.

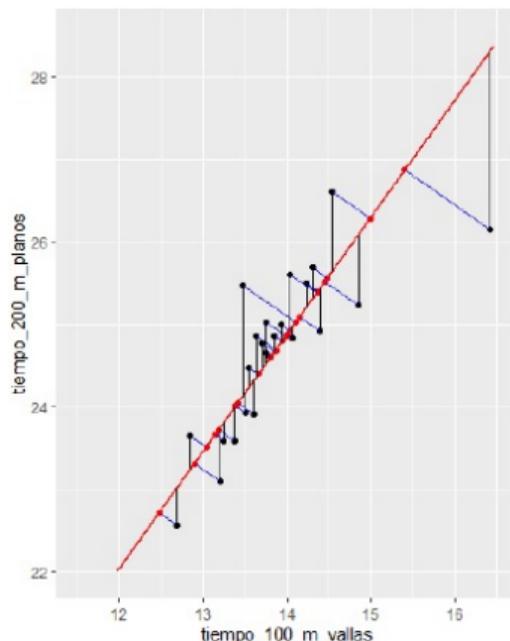
Karl Pearson (1901), describir con una recta.

Hotelling (1933), relación entre variables $g(X_1, X_2)$.

Componentes principales

Ejemplo: Atletismo, pruebas de 100m y 200m.

100m vallas	200m planos
12.69	22.56
12.85	23.65
13.2	23.1
13.61	23.92
13.51	23.93
13.75	24.65
13.38	23.59
13.55	24.48
13.63	24.86
13.25	23.59
13.75	25.03
13.24	23.59
13.85	24.87
13.71	24.78
13.79	24.61
13.93	25
13.47	25.47
14.07	24.83
14.39	24.92
14.04	25.61
14.31	25.69
14.23	25.5
14.85	25.23
14.53	26.61
16.42	26.16



Se observa cierta estructura.

Karl Pearson (1901), describir con una recta.

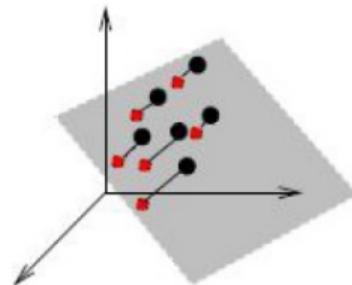
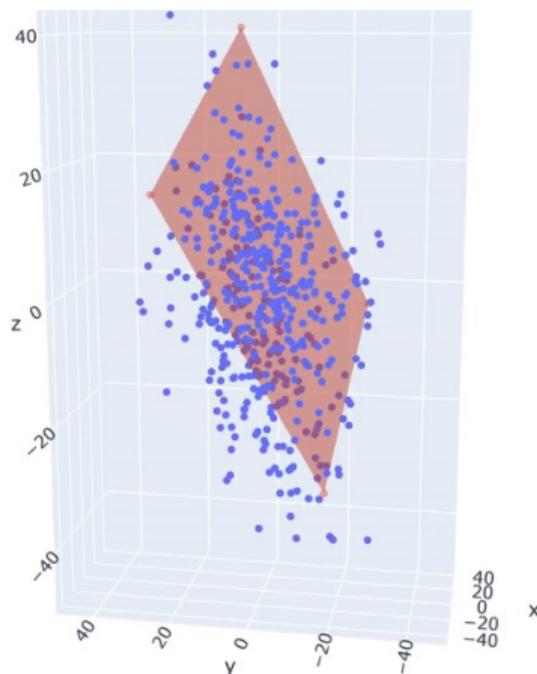
Hotelling (1933), relación entre variables $g(X_1, X_2)$.

No confundir con regresión, Incorporar incertidumbre.

Componentes principales

Ejemplo: Atletismo, pruebas de 100m, 200m y salto de longitud.

100m vallas	200m planos	salto long
12.69	22.56	7.27
12.85	23.65	6.71
13.2	23.1	6.68
13.61	23.92	6.25
13.51	23.93	6.32
13.75	24.65	6.33
13.38	23.59	6.37
13.55	24.48	6.47
13.63	24.86	6.11
13.25	23.59	6.28
13.75	25.03	6.34
13.24	23.59	6.37
13.85	24.87	6.05
13.71	24.78	6.12
13.79	24.61	6.08
13.93	25	6.4
13.47	25.47	6.34
14.07	24.83	6.13
14.39	24.92	6.1
14.04	25.61	5.99
14.31	25.69	5.75
14.23	25.5	5.5
14.85	25.23	5.47
14.53	26.61	5.5
16.42	26.16	4.88

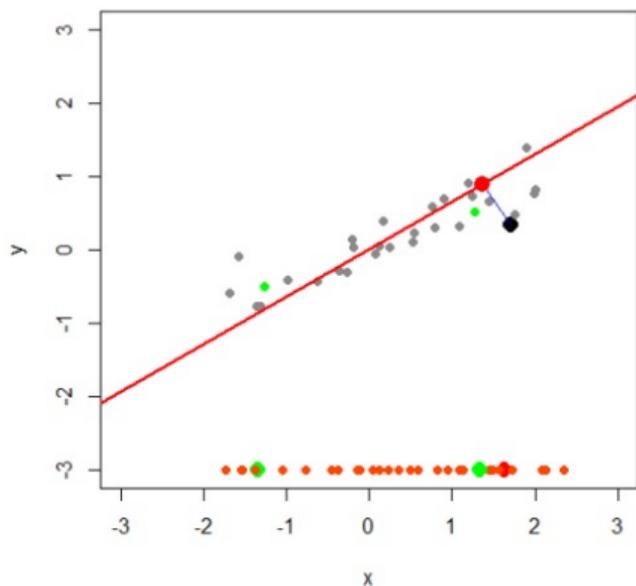


Componentes principales

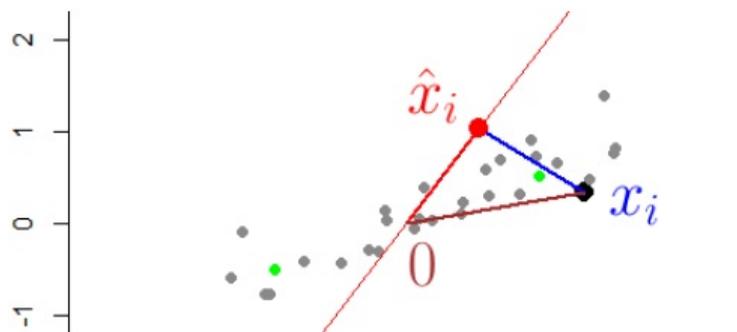
Ejemplo: Compresión de imágenes digitales.



Componentes principales



- Buscamos direcciones informativas (estructura)
informativo = máxima variabilidad
- Buscamos minimizar el error de reconstrucción.



Componentes principales

Obs! Los dos enfoques anteriores son equivalentes.

Prueba:

Denotemos X la v.a. que corresponde a los datos ($X \in \mathbb{R}^2$ en el ejemplo).

Por simplicidad, supongamos que los datos \mathbf{x}_i están centrados (i.e. $\mathbb{E}(X) = \mathbf{0}$).

$$\begin{aligned}MSS &= \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 = \frac{1}{n} \sum_{i=1}^n \|(\mathbf{x}_i - \hat{\mathbf{x}}_i) + \hat{\mathbf{x}}_i\|^2 \\ &= \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2 + \frac{1}{n} \sum_{i=1}^n \|\hat{\mathbf{x}}_i\|^2 = \text{Reconstruction error} + \text{Var}(X).\end{aligned}$$

MSS es fija, luego minimizar el error de reconstrucción equivale a maximizar la varianza de los datos.

Componentes principales

Enfoque probabilístico:

Matriz de datos

$$\mathbb{X} = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1d} \\ X_{21} & X_{22} & \dots & X_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nd} \end{pmatrix}.$$

- Consideramos $X = (X_1, \dots, X_d) \in \mathbb{R}^d$ como variable aleatoria, y los datos $\mathbf{x}_i = (x_{i1}, \dots, x_{id}) \in \mathbb{R}^d$, para $i = 1, 2, \dots, n$ como muestra de X .
- Supondremos que conocemos la ley \mathbb{P}_X .
- Supondremos también que $\mathbb{E}(X) = \mathbf{0}$ (los datos están centrados). En consecuencia, $\text{Var}(X) = \mathbb{X}^T \mathbb{X}$.

Componentes principales

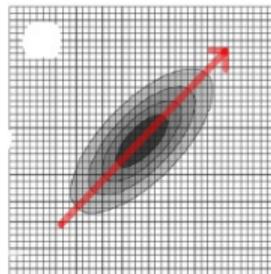
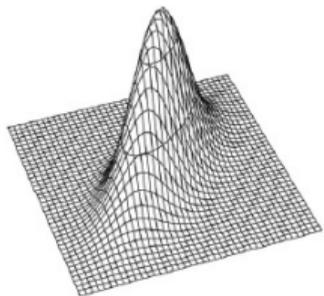
Caso particular 1D: (proyectamos a un subespacio 1-dimensional).

Suponga que proyectamos a un subespacio $\langle l \rangle \Rightarrow \langle l, X \rangle = l^T X$.

Buscamos maximizar

$$\max_{\|l\|=1} \text{Var}(l^T X) = \max_{l \neq 0} \frac{\text{Var}(l^T X)}{l^T l} = \max_{l \neq 0} \frac{l^T \text{Var}(X) l}{l^T l} = \max_{l \neq 0} \frac{l^T (\mathbb{X}^T \mathbb{X}) l}{l^T l}.$$

(cociente de Rayleigh).



Teorema (Teorema espectral / Descomposición espectral)

En otras palabras, A puede escribirse como una suma de matrices de rango 1

$$\begin{aligned} A &= \begin{pmatrix} & & & \\ & & & \\ & & & \\ & & & \end{pmatrix} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \dots & \\ & & & \lambda_d \end{pmatrix} \begin{pmatrix} T_1 \\ T_2 \\ \dots \\ T_d \end{pmatrix} \\ &= \sum_{i=1}^d \lambda_i T_i T_i^T. \end{aligned}$$

Componentes principales

Comentario:

Para $1 \leq k \leq d$, la suma

$$\hat{A}_k = \sum_{i=1}^k \lambda_i i_i^T,$$

es una matriz de rango k siempre que los $\lambda_i \neq 0$ (ya que los i son independientes).

Veremos más adelante, que esta es la mejor aproximación de rango k de la matriz A .