Inteligencia Artificial LFIS419

Clase 6: Reducción de dimensionalidad

Profesor: Jorge Arevalo (jorge.arevalo@uv.cl jab@meteo.uv.cl)

Martes 2 de mayo de 2023

Objetivos de la sesión

- Introducir el concepto de reducción de dimensionalidad
- Profundizar en los conceptos básicos de ML
- Introducir el uso de PCA
- Propiciar la interacción entre estudiantes y profesor

Problema

- Análisis de datos multivariados es importante
- Espacios con más de 3 dimensiones son difíciles de visualizar

- Se busca representar los datos de una manera que facilite el análisis

Beneficios

- Se reduce la carga computacional en las siguientes etapas
- Podría reducirse el ruido
- Proyectar los datos en un espacio de dimension pequeña es útil para visualizar los datos

Objetivos PCA

- Reducir dimensionalidad
- Identificar variables subyacentes con significado
- Perder la menor cantidad de información (manteniendo la mayor variabilidad)

Conceptos básicos (varianza)

- Desviación estándar
  - Promedio de la distancia entre cada dato y la media
- Varianza
  - Cuadrado de la SD

Conceptos básicos (covarianza)

- Cómo los datos 2D varían respecto de su medio

$$\operatorname{cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

- Cov(X,Y)>0: ambas ”crecene juntas”
- Cov(X,Y)<0: una “crece” la otra “deccrece”
- Cov(X,Y)=0: las dimensiones son independientes

Conceptos básicos (matriz de covarianza)

- Matriz de covarianza contiene la covarianza entre cada par de dimensiones
- Cov(X,X) = varianza de la columna X

$$C = \begin{pmatrix} \text{cov}(x,x) & \text{cov}(x,y) & \text{cov}(x,z) \\ \text{cov}(y,x) & \text{cov}(y,y) & \text{cov}(y,z) \\ \text{cov}(z,x) & \text{cov}(z,y) & \text{cov}(z,z) \end{pmatrix}$$

Conceptos básicos (valores y vectores propios)

- Para una martiz **A**, los vectores **x** que tienen la misma dirección que **Ax** se llaman vectores propios de **A**

- **lambda** son los valores propios de **A**
- Para obtener los vectores propios, calcular las raíces de det(A-lambda I)=0 y resolver para cada lambda

Conceptos básicos (valores y vectores propios)

- Valores propios miden la cantidad de varianza explicada por cada PC

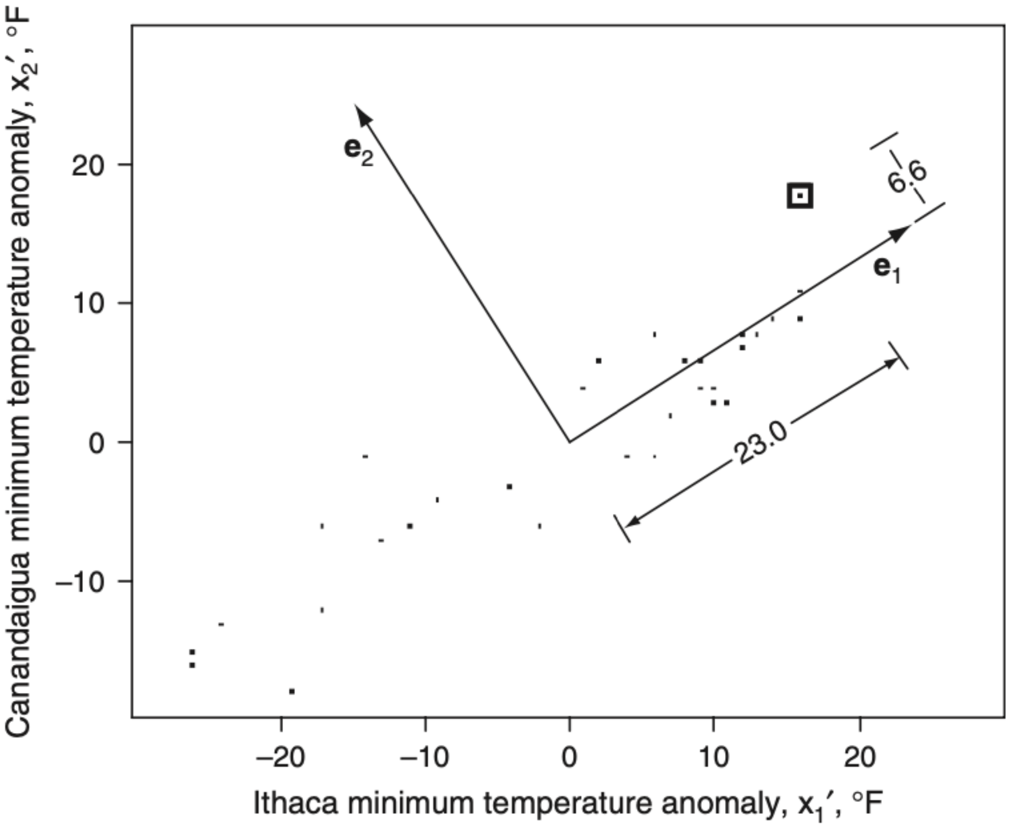
- Vectores propios corresponden a las PC no-correlacionadas. Direcciones de cada componente.

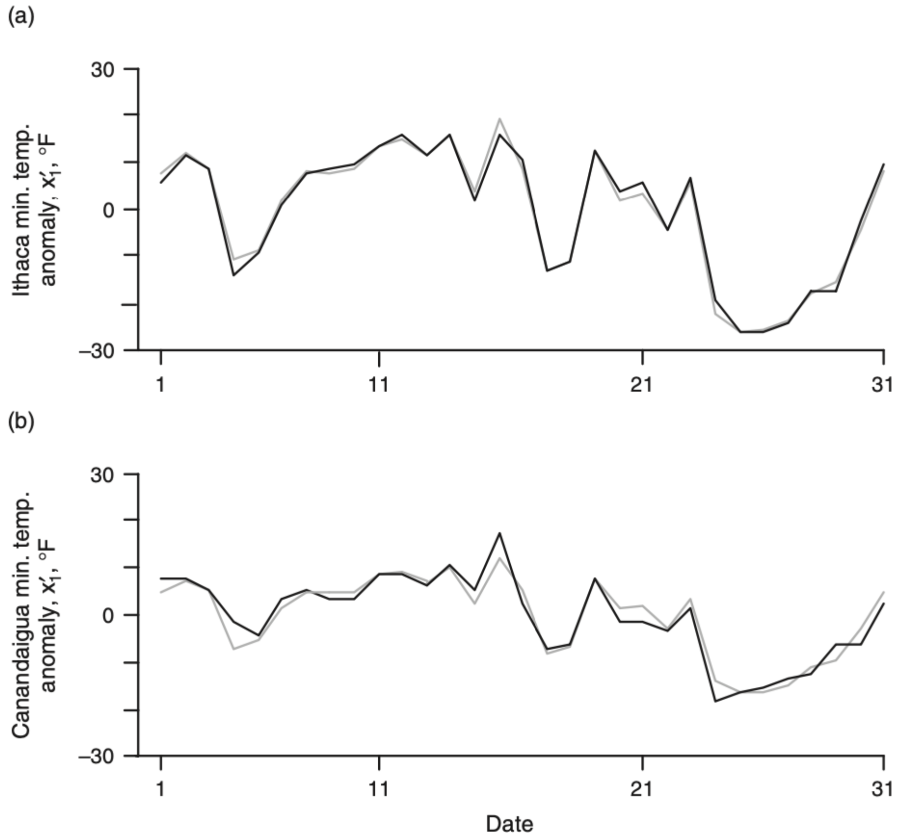
Conceptos básicos (componente principal)

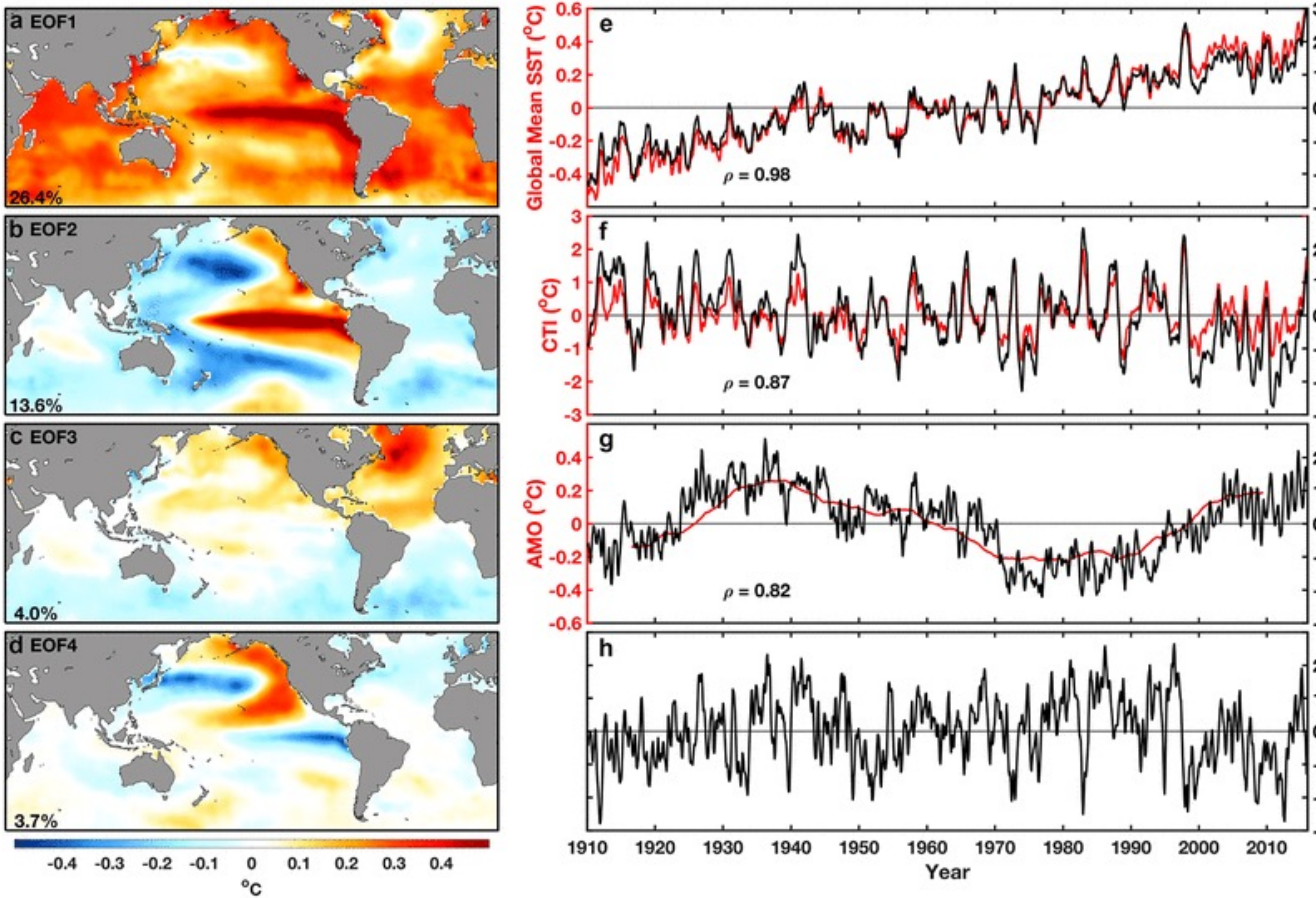
- Estimados desde los vectores propios de la matriz de covarianza
- Corresponden a las proyecciones de los datos originales en los vectores propios
- Son combinaciopnes lineales de las variables originales.
- Las primeras PC contienen la mayor parte de la variabilidad de los datos originales.

Statistical Methods in the Atmospheric Sciences, 2nd Ed., Wilks

$$\mathbf{x}'(t) = \begin{bmatrix} x'_1(t) \\ x'_2(t) \end{bmatrix} \approx \mathbf{e}_1 u_1(t) = \begin{bmatrix} .848 \\ .530 \end{bmatrix} u_1(t).$$






a EOF1
26.4%
b EOF2
13.6%
c EOF3
4.0%
d EOF4
3.7%
-0.4 -0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4
°C
e
Global Mean SST (°C)
PC1
ρ = 0.98
f
CTI (°C)
PC2
ρ = 0.87
g
AMO (°C)
PC3
ρ = 0.82
h
PC4
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010
Year