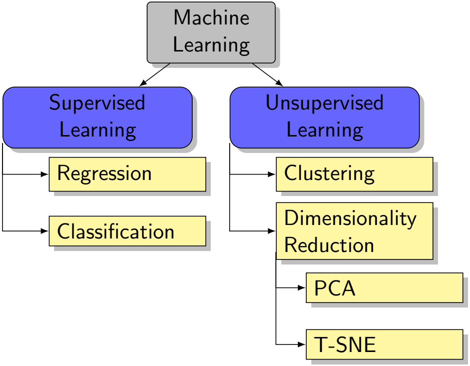
Reducción de dimensionalidad

Dr. Mauricio Toledo-Acosta

Diplomado Ciencia de Datos con Python

Table of Contents

Introducción



Reducción de dimensionalidad

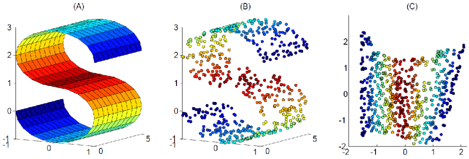
Reducción de dimensionalidad

La reducción de la dimensionalidad es la transformación de los datos de un espacio de alta dimensión a un espacio de baja dimensión, de manera que la representación de baja dimensión conserve algunas propiedades significativas de los datos originales.

Reducción de dimensionalidad

Reducción de dimensionalidad

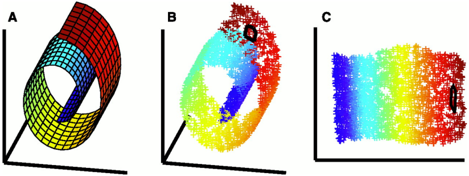
La reducción de la dimensionalidad es la transformación de los datos de un espacio de alta dimensión a un espacio de baja dimensión, de manera que la representación de baja dimensión conserve algunas propiedades significativas de los datos originales.



Reducción de dimensionalidad

Reducción de dimensionalidad

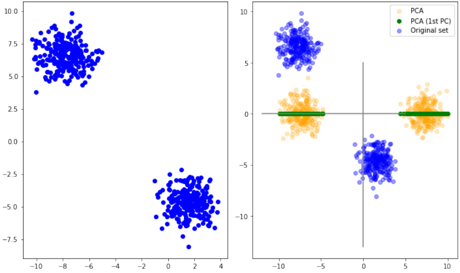
La reducción de la dimensionalidad es la transformación de los datos de un espacio de alta dimensión a un espacio de baja dimensión, de manera que la representación de baja dimensión conserve algunas propiedades significativas de los datos originales.



Reducción de dimensionalidad

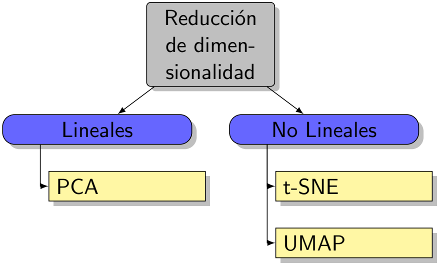
Reducción de dimensionalidad

La reducción de la dimensionalidad es la transformación de los datos de un espacio de alta dimensión a un espacio de baja dimensión, de manera que la representación de baja dimensión conserve algunas propiedades significativas de los datos originales.



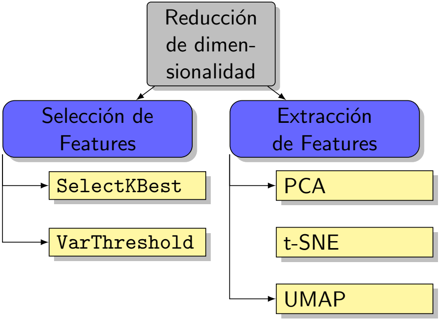
Clasificación de los Métodos

Desde un punto de vista matemático, se trata de transformar los datos de un espacio de alta dimensión a un espacio de baja dimensión.



Clasificación de los Métodos

Desde un punto de vista computacional, se trata de reducir las features que definen a los datos para hacer más tratables las tareas del Machine Learning.



Utilidad

¿Para qué sirve la reducción de dimensionalidad?

- Visualización.

Utilidad

¿Para qué sirve la reducción de dimensionalidad?

- Visualización.
- Extraer la información más importante de los datos.

Utilidad

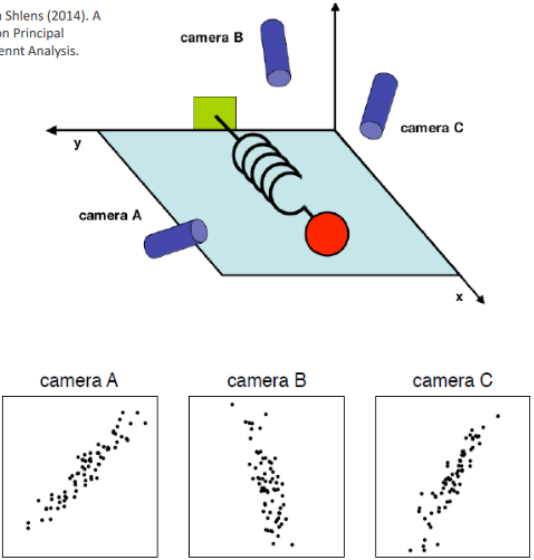
¿Para qué sirve la reducción de dimensionalidad?

- Visualización.
- Extraer la información más importante de los datos.
- Obtener features para fines de clasificación.

Table of Contents

PCA

Jonathon Shlens (2014). A tutorial on Principal Componennt Analysis.



PCA

PCA puede pensarse como el ajuste de un elipsoide D-dimensional al conjunto de datos, donde cada eje del elipsoide representa una componente principal. Si algún eje del elipsoide es pequeño, entonces la varianza a lo largo de ese eje también es pequeña.

PCA

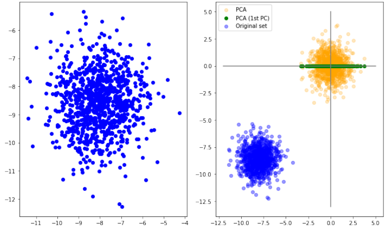
PCA puede pensarse como el ajuste de un elipsoide D-dimensional al conjunto de datos, donde cada eje del elipsoide representa una componente principal. Si algún eje del elipsoide es pequeño, entonces la varianza a lo largo de ese eje también es pequeña.

PCA transforma los datos a un nuevo sistema de coordenadas de tal manera que la mayor varianza se sitúa en la primera coordenada, la segunda mayor varianza en la segunda coordenada, y así sucesivamente.

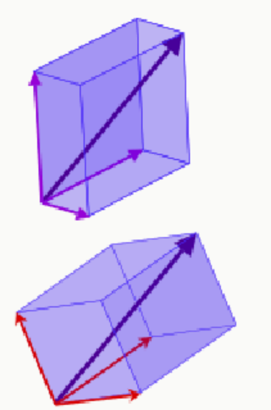
PCA

PCA puede pensarse como el ajuste de un elipsoide D-dimensional al conjunto de datos, donde cada eje del elipsoide representa una componente principal. Si algún eje del elipsoide es pequeño, entonces la varianza a lo largo de ese eje también es pequeña.

PCA transforma los datos a un nuevo sistema de coordenadas de tal manera que la mayor varianza se sitúa en la primera coordenada, la segunda mayor varianza en la segunda coordenada, y así sucesivamente.



Cambios de Base



- Las coordenadas de un punto en son los coeficientes de los vectores canónicos unitarios

$$\begin{aligned} e_1 &= (1, 0, ..., 0), \\ &\dots \\ e_D &= (0, 0, ..., 1). \end{aligned}$$

- Todo punto puede ser expresado en una infinidad de bases.
- Para cambiar de base hay que multiplicar el vector por la matriz

$$\begin{pmatrix} v_1^{(1)} & \dots & v_1^{(D)} \\ & \cdots & \\ v_D^{(1)} & \dots & v_D^{(D)} \end{pmatrix}$$

PCA

¿Cómo obtenemos la nueva base de coordenadas para PCA? Esta base de vectores deben ser las direcciones de máxima varianza, es decir, las componentes principales.

PCA

¿Cómo obtenemos la nueva base de coordenadas para PCA? Esta base de vectores deben ser las direcciones de máxima varianza, es decir, las componentes principales.

Consideremos la matriz de covarianza

$$C_X = \frac{1}{N} X \cdot X^T.$$

PCA

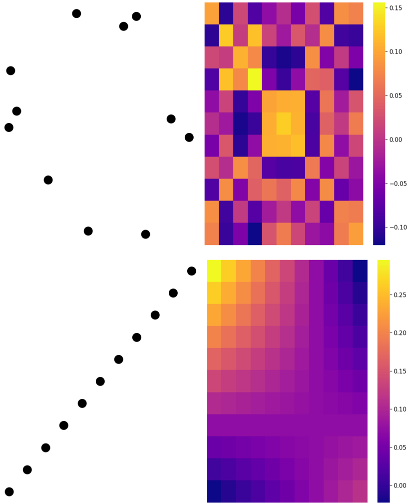
¿Cómo obtenemos la nueva base de coordenadas para PCA? Esta base de vectores deben ser las direcciones de máxima varianza, es decir, las componentes principales.

Consideremos la matriz de covarianza

$$C_X = \frac{1}{N} X \cdot X^T.$$

- Los términos altos en la diagonal corresponden a una varianza alta.
- Los términos altos fuera de la diagonal corresponden a una redundancia alta.

Matriz de covarianza



Matriz de Covarianza

De acuerdo con lo anterior, para nuestra matriz de covarianza C_X deseamos:

- Minimizar la redundancia, medida por la magnitud de la covarianza.
- Maximizar la varianza.

Matriz de Covarianza

De acuerdo con lo anterior, para nuestra matriz de covarianza C_X deseamos:

- Minimizar la redundancia, medida por la magnitud de la covarianza.
- Maximizar la varianza.

Esto, en términos de matrices, quiere decir Diagonalizar. Es decir, encontrar matrices P y D tales que

$$D = P \cdot C_X \cdot P^T$$

Matriz de Covarianza

De acuerdo con lo anterior, para nuestra matriz de covarianza C_X deseamos:

- Minimizar la redundancia, medida por la magnitud de la covarianza.
- Maximizar la varianza.

Esto, en términos de matrices, quiere decir Diagonalizar. Es decir, encontrar matrices P y D tales que

$$D = P \cdot C_X \cdot P^T$$

Esto se hace encontrando los eigenvector y eigenvalores de C_X.

Eigenvalores y Eigenvectores

Cada matriz M de tamaño $n \times n$ se puede ver como una transformación del espacio $\mathbb{R}^n$ en el espacio $\mathbb{R}^n$. Es decir, toma una punto $p \in \mathbb{R}^n$ y devuelve otro vector $M \cdot p \in \mathbb{R}^n$. ¿Cómo es este punto respecto al inicial?

Eigenvalores y Eigenvectores

Cada matriz M de tamaño $n \times n$ se puede ver como una transformación del espacio $\mathbb{R}^n$ en el espacio $\mathbb{R}^n$. Es decir, toma una punto $p \in \mathbb{R}^n$ y devuelve otro vector $M \cdot p \in \mathbb{R}^n$. ¿Cómo es este punto respecto al inicial?

Hay vectores especiales, dependiendo de M, que son transformados en un múltiplo de ellos mismos.

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$A \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$A \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Eigenvalores y Eigenvectores

Cada matriz M de tamaño $n \times n$ se puede ver como una transformación del espacio $\mathbb{R}^n$ en el espacio $\mathbb{R}^n$. Es decir, toma una punto $p \in \mathbb{R}^n$ y devuelve otro vector $M \cdot p \in \mathbb{R}^n$. ¿Cómo es este punto respecto al inicial?

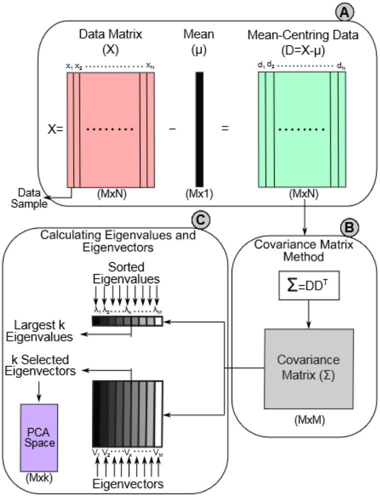
Hay vectores especiales, dependiendo de M, que son transformados en un múltiplo de ellos mismos.

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$A \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$A \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

El proceso


A
Data Matrix
(X)
Mean
(μ)
Mean-Centring Data
(D=X-μ)
X=
(MxN)
(Mx1)
(MxN)
Data
Sample
B
Covariance Matrix
Method
Σ=DD^T
Covariance
Matrix (Σ)
(MxM)
C
Calculating Eigenvalues and
Eigenvectors
Sorted
Eigenvalues
Largest k
Eigenvalues
k Selected
Eigenvectors
PCA
Space
(Mxk)
Eigenvectors

Conexiones entre la geométricas y la estadística

- Promedio/Centroide = Media.
- Norma = Varianza.
- Ángulo entre vectores = Covarianza.

Ventajas y Desventajas

Ventajas

- Permite eliminar variables correlacionadas.
- Permite la visualización de datos.
- Puede ayudar a reducir el overfitting.

Ventajas y Desventajas

Ventajas

- Permite eliminar variables correlacionadas.
- Permite la visualización de datos.
- Puede ayudar a reducir el overfitting.

Desventajas

- Suele requiere de un escalamiento de datos antes.
- Se pierde información.
- Perdemos la interpretabilidad de las variables de entrada.