



No Supervisado

Introduccion PCA



Reducir la dimensionalidad: El “reductor”



Reductor de Dimensionalidad (by Fluzo Corp.)

SELECCIÓN DE FEATURES:

- Mantiene features originales
- Mecanismos Supervisados y no supervisados
- Pérdida de relaciones con las features “eliminadas”

TRANSFORMACION DE FEATURES:

- No mantiene las features originales
- El mecanismo más común es la “Proyección”. Ej: **PCA**



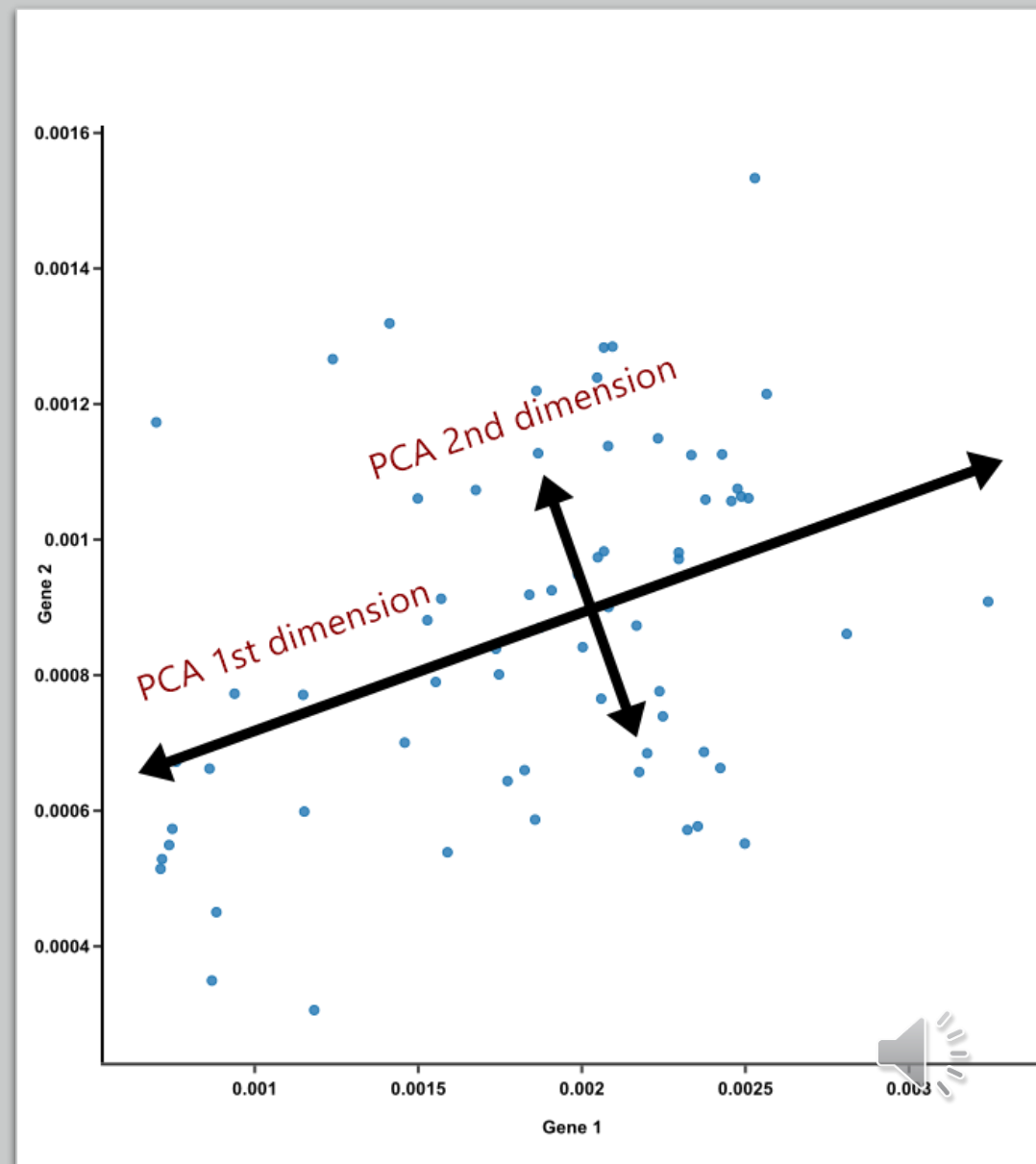
¿Qué es el PCA?

Principal Component Analysis (PCA)

Método estadístico que permite simplificar la complejidad de espacios muestrales con muchas dimensiones a la vez que conserva su información.

Supóngase que existe una muestra con n individuos cada uno con p variables (X_1, X_2, \dots, X_p), es decir, el espacio muestral tiene p dimensiones. PCA permite encontrar un número de factores subyacentes ($z < p$) que explican aproximadamente lo mismo que las p variables originales.

Cada una de estas z nuevas variables recibe el nombre de componente principal.



¿Qué es el PCA?

1. Si partimos de un dataset de m features $(x_1, x_2, x_3, \dots, x_m)$ la PCA va a encontrar m componentes principales (o features transformadas) de esta forma:



¿Qué es el PCA?

1. Si partimos de un dataset de m features $(x_1, x_2, x_3, \dots, x_m)$ la PCA va a encontrar m componentes principales (o features transformadas) de esta forma:

$$CP_1 = c_{10}x_0 + c_{11}x_1 + c_{12}x_2 + c_{13}x_3 + \dots + c_{1m}x_m$$

$$CP_2 = c_{20}x_0 + c_{21}x_1 + c_{22}x_2 + c_{23}x_3 + \dots + c_{2m}x_m$$

$$CP_3 = c_{30}x_0 + c_{31}x_1 + c_{32}x_2 + c_{33}x_3 + \dots + c_{3m}x_m$$

...

$$CP_m = c_{m0}x_0 + c_{m1}x_1 + c_{m2}x_2 + c_{m3}x_3 + \dots + c_{mm}x_m$$



¿Qué es el PCA?

1. Si partimos de un dataset de m features $(x_1, x_2, x_3, \dots, x_m)$ la PCA va a encontrar m componentes principales (o features transformadas) de esta forma:

$$CP_1 = c_{10}x_0 + c_{11}x_1 + c_{12}x_2 + c_{13}x_3 + \dots + c_{1m}x_m$$

$$CP_2 = c_{20}x_0 + c_{21}x_1 + c_{22}x_2 + c_{23}x_3 + \dots + c_{2m}x_m$$

$$CP_3 = c_{30}x_0 + c_{31}x_1 + c_{32}x_2 + c_{33}x_3 + \dots + c_{3m}x_m$$

...

$$CP_m = c_{m0}x_0 + c_{m1}x_1 + c_{m2}x_2 + c_{m3}x_3 + \dots + c_{mm}x_m$$

2. La gracia está en cómo se calculan los coeficientes c_{ij} y en el hecho de que las componentes principales están ordenadas según su poder informativo (es decir CP_1 es más informativa que CP_2 , etc) de forma que si queremos tener un “valor informativo” determinado respecto del original escogeremos las n primeras componentes principales.



¿Qué es el PCA?

1. Si partimos de un dataset de m features $(x_1, x_2, x_3, \dots, x_m)$ la PCA va a encontrar m componentes principales (o features transformadas) de esta forma:

$$CP_1 = c_{10}x_0 + c_{11}x_1 + c_{12}x_2 + c_{13}x_3 + \dots + c_{1m}x_m$$

$$CP_2 = c_{20}x_0 + c_{21}x_1 + c_{22}x_2 + c_{23}x_3 + \dots + c_{2m}x_m$$

$$CP_3 = c_{30}x_0 + c_{31}x_1 + c_{32}x_2 + c_{33}x_3 + \dots + c_{3m}x_m$$

...

$$CP_m = c_{m0}x_0 + c_{m1}x_1 + c_{m2}x_2 + c_{m3}x_3 + \dots + c_{mm}x_m$$

2. La gracia está en cómo se calculan los coeficientes c_{ij} y en el hecho de que las componentes principales están ordenadas según su poder informativo (es decir CP_1 es más informativa que CP_2 , etc) de forma que si queremos tener un “valor informativo” determinado respecto del original escogeremos las n primeras componentes principales.

3. Elegidas las n componentes principales las features de cada instancia pasan a ser los valores de esas n componentes que resultan de sustituir sus valores para cada features original (x_i) en las ecuaciones anteriores



¿Qué es el PCA?

1. Si partimos de un dataset de m features $(x_1, x_2, x_3, \dots, x_m)$ la PCA va a encontrar m componentes principales (o features transformadas) de esta forma:

$$CP_1 = c_{10}x_0 + c_{11}x_1 + c_{12}x_2 + c_{13}x_3 + \dots + c_{1m}x_m$$

$$CP_2 = c_{20}x_0 + c_{21}x_1 + c_{22}x_2 + c_{23}x_3 + \dots + c_{2m}x_m$$

$$CP_3 = c_{30}x_0 + c_{31}x_1 + c_{32}x_2 + c_{33}x_3 + \dots + c_{3m}x_m$$

...

$$CP_m = c_{m0}x_0 + c_{m1}x_1 + c_{m2}x_2 + c_{m3}x_3 + \dots + c_{mm}x_m$$

2. La gracia está en cómo se calculan los coeficientes c_{ij} y en el hecho de que las componentes principales están ordenadas según su poder informativo (es decir CP_1 es más informativa que CP_2 , etc) de forma que si queremos tener un "valor informativo" determinado respecto del original escogeremos las n primeras componentes principales.

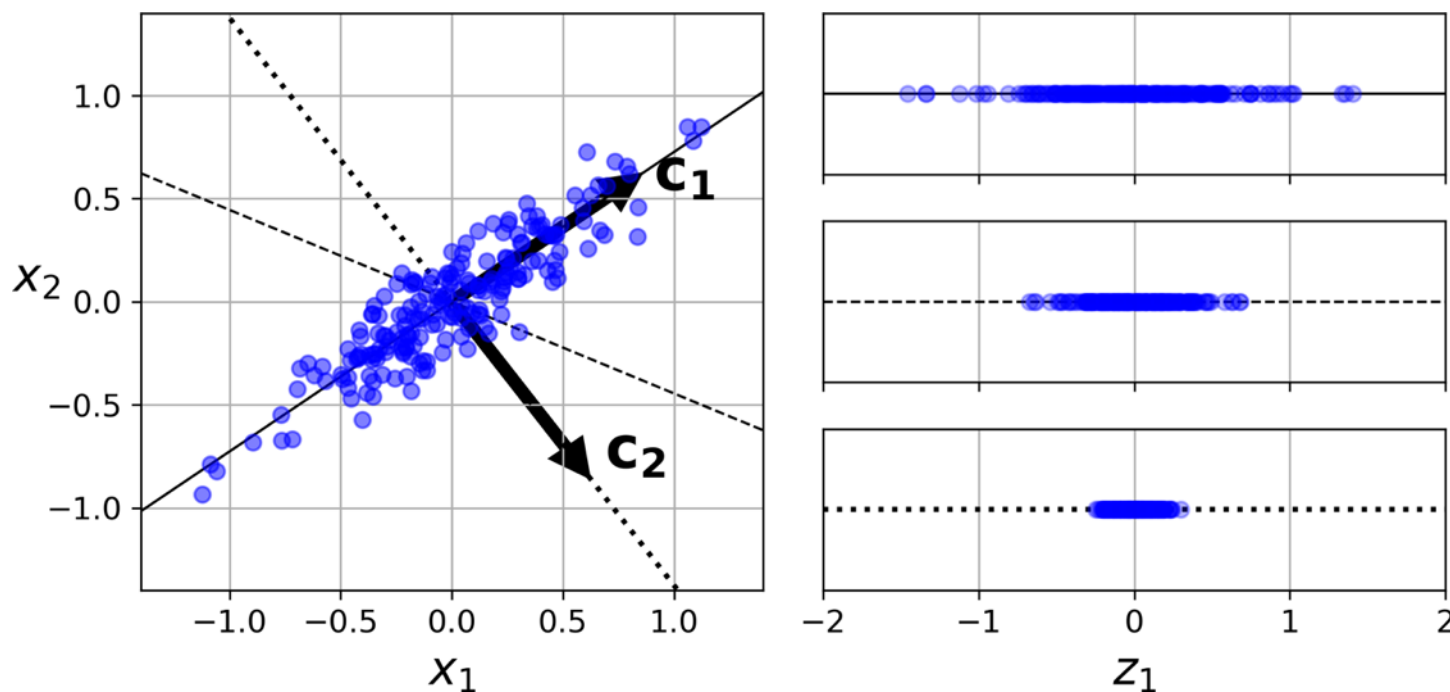
3. Elegidas las n componentes principales las features de cada instancia pasan a ser los valores de esas n componentes que resultan de sustituir sus valores para cada features original (x_i) en las ecuaciones anteriores



¿Qué es el PCA?

Identifica los hiperplanos que maximizan la varianza y proyecta los datos en esos hiperplanos, de tal manera que minimicemos la pérdida de información. O sea que vamos a buscar “la proyección” de los puntos que mantenga la mayor cantidad de varianza (cambio = información) de los datos originales

Las líneas discontinuas de la siguiente imagen serían una proyección con muy poca varianza de la variable.

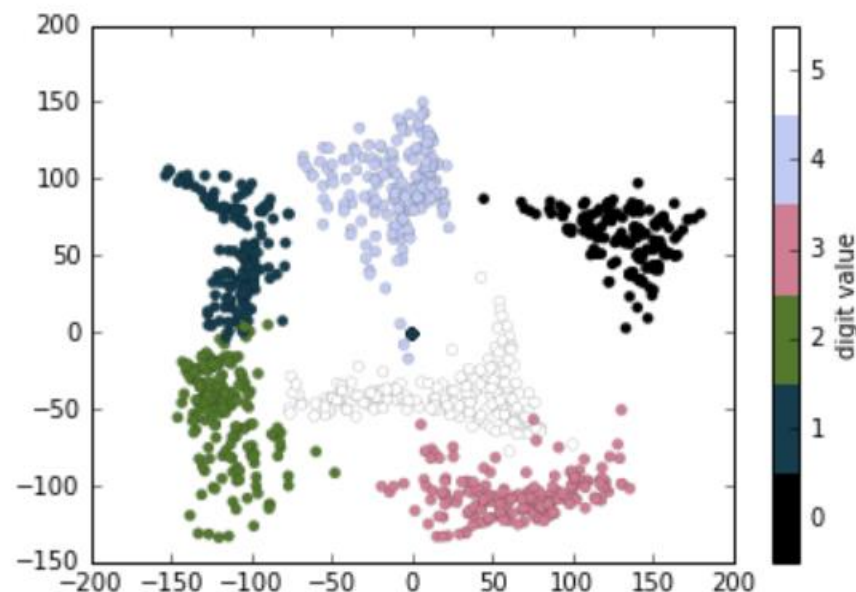


¿Para qué se usa PCA?

Extracción de patrones en los datos: convierte el dataset para ver similitudes y diferencias en los datos.

Feature reduction: Comprimir la información de un dataset en menos variables. Imprescindible con datasets de cientos o miles de features.

Visualización para clasificación: Datasets de más de tres variables son imposibles de representar en una gráfica. Con PCA podemos.



¿Para qué se usa PCA?

Extracción de patrones en los datos: convierte el dataset para ver similitudes y diferencias en los datos.

Feature reduction: Comprimir la información de un dataset en menos variables. Imprescindible con datasets de cientos o miles de features.

Visualización para clasificación: Datasets de más de tres variables son imposibles de representar en una gráfica. Con PCA podemos.

