Ttitle Principal component analysis

Afdasdadf

# Introducción y motivación

En épocas modernas la data abunda y los análisis sobre ella han llevado a desarrollar técnicas y procedimientos para su entendimiento que permean a todas las ramas de la ciencia.

Dentro de dichos estudios, el análisis multivariado ha probado ser de mucha utilidad para el procesamiento de datos y el descubrimiento de patrones. De igual manera la cantidad abrumadora de datos que se presentan día a día requieren que los métodos para procesar datos sean eficientes y conserven propiedades fundamentales en el análisis, es para responder esta necesidad que surge el análisis de componentes principales.

El análisis de componentes principales surge como un método muy eficiente y poderoso para reducir la dimensionalidad de la data sin sacrificar mucha calidad informativa a la hora de ser aplicado.

Particularmente, el análisis de componentes principales o bien conocido por sus siglas en inglés (PCA), consiste en reducir la dimensión de la data en unas variables llamadas componentes principales (PC) mientras se retiene tanta variabilidad proveniente de la data como sea posible.

Este trabajo está apuntado a explicar el método de componentes principales, su estimación, su relación con el análisis matricial y finalmente una aplicación directa del mismo.

# Definición de los componentes principales

Suponga que *x* es un vector de *m* variables aleatorias y que se desea analizar la estructura de varianzas y covarianzas de las *m* variables. Si las *m* variables aleatorias son pocas, no es inconveniente alguno analizar las varianzas y covarianzas una a una. Lamentablemente cuando el número de variables escala la complejidad del análisis también lo hace ya que sería necesario analizar covarianzas.

El PCA se concentra en las varianzas de las variables para derivar los valores de sus PC.

El primer paso es entonces buscar una función lineal de los elementos de *x* que tengan la máxima varianza, donde es un vector de *m* constantes . Tal que:

Luego de encontrar esta función lineal se repite el mismo procedimiento para buscar una función que cumpla todas las propiedades antes mencionadas y que además esté no correlacionado con el anterior vector. Este procedimiento se sigue hasta llegar a la función que es una función que maximice la varianza dado que este vector no está correlacionado con ninguno de las anteriores *k-1* funciones. Dado que se necesitan vectores no correlacionados estos mismos tienen un número limitado que asciende a *m*, el número de variables.

Afortunadamente, en la mayoría de veces los primeros *w* PC van a explicar la mayoría de la varianza y se presenta que este *w << m* por lo que resulta en un método efectivo de reducción de dimensiones.

# Derivación de los componentes principales

Luego de haber definido los componentes principales se necesita un método de cálculo, que sea eficiente y que cumpla con todos los requerimientos antes mencionados para los PC.

Considere la matriz de covarianza conocida , ahora, aplicando la derivación de los PC anteriormente expuesta se obtiene que la función *k*-ésima es donde es un vector propio de la matriz , asociado a *k*-ésimo valor propio .

## Derivación para 1 componente principal

Para derivar entonces los PCs consideremos la función el vector maximiza Dado el actual planteamiento de la ecuación es evidente que el máximo no ve a ser alcanzado para un número finito de , por lo que una restricción de normalización debe de ser impuesta. La restricción planteada para esta derivación es . Esta restricción es equivalente a plantear que la suma de los cuadrados de los elementos del vector sea igual a 1. En resumen, el problema de optimización se ve de la siguiente manera:

La manera estándar de resolver estos problemas es usar multiplicadores de LaGrange, por lo que se seguirá este camino para resolver algebraicamente el problema.

El problema de LaGrange modificado queda de la siguiente manera:

Ahora:

De manera equivalente:

Donde es la matriz identidad de dimensión *m*. Es evidente, que es un valor propio de y es su vector propio correspondiente. Para decidir cuál de los *m* vectores propios hace que tenga varianza máxima es útil notar que la cantidad a maximizar es:

Por lo que debe de ser tan grande como sea posible. Por lo tanto es el vector propio que corresponde a el mayor valor propio de la matriz y .

Por lo general el *k*-ésimo PC corresponde al *k*-ésimo vector propio asociado al *k*-ésimo valor propio

El mismo procedimiento se debe de seguir para encontrar los otros PC, en este trabajo se explorará el caso donde *m=2*, esto se debe a que el procedimiento de des correlacionar los PC no es trivial.

## Derivación para *2* componentes

El procedimiento para derivar el primer componente es estándar por lo que se asumirá que este PC ya fue encontrado. Para encontrar el segundo componente se establece entonces que dicho componente debe de sujeto a que no exista correlacion alguna con el primer componente ya establecido, por lo que esto se traduce en donde *cov* es la función de covarianza entre dos variables aleatorias.

Nótese que:

Por lo que solucionando cualquier de las ecuaciones resultantes sería equivalente a especificar una correlación de 0 entre los dos vectores.

Ahora si se toma la última ecuación y se establece una restricción de normalización para alcanzar el máximo en valores finitos de se tiene que el problema de optimización a resolver es:

De la misma manera que antes, se resuelve por multiplicadores de LaGrange, la ecuación con los multiplicadores queda:

Derivando con respecto a se tiene:

Multiplicando a derecha por se obtiene:

Los dos primeros terminos son 0 y dada la restricción de normalidad de primer componente la ecuación se reduce a:

Por lo que la ecuación de queda:

Equivalente a:

Donde es evidente que es de nuevo un valor propio de

Finalmente sabiendo que , necesita ser lo más grande posible, asumiendo que la matriz no tiene valores propios repetidos por lo que este lambda corresponde al segundo valor propio más grande para la matriz.

De manera análoga se puede demostrar que para el tercero, cuarto,…, *m*-ésimo PC, los vectores de coeficientes corresponden a los vectores propios y los lambdas a sus valores propios correspondientes.

# Propiedades matemáticas y estadísticas de los componentes principales poblacionales

El primer paso en todo análisis matemático es establecer un modelo que sea analíticamente manejable y del cual se puedan estudiar propiedades sin complicaciones y refinamientos innecesarios en un primer acercamiento.

Es por ello que en este informe se procede primero a analizar las propiedades de los componentes provenientes de matrices poblacionales (covarianzas y correlaciones) para luego redefinir las mismas en casos muestrales donde más detalle es necesario.

## Propiedades matemáticas de los componentes principales poblacionales

Considere el vector *z* cuyo *k*-ésimo elemento corresponde al PC número *k*, entonces se tiene que:

Donde la matriz *A* es la matriz ortogonal cuya *k*-ésima columna, es el *k*-ésimo vector propio de . Es entonces natural pensar que los PC estén definidos como una transformación ortonormal lineal de *x*. Adicionalmente de la derivación algebraica de los PC se tiene que:

Donde es la matriz diagonal cuyo *k*-ésimo elemento es , el *k*-ésimo valor propio de . Ahora es necesario considerar dos representaciones adicionales de la ecuación anterior que van a resultar útiles en derivaciones futuras:

Esto sucede porque A es una matriz ortogonal.

### Propiedad de los primeros elementos

*Para todo entero* ***q, con 1 < q < p****, considere la transformación ortonormal:*

*Donde* ***y*** *es un vector de q elementos y Bt es una matriz de dimensión q, y sea la matriz de varianza-covarianza para* ***y****. Entonces la traza de denotada por , es maximizada tomando , donde consiste en las primeras* ***q*** *columnas de A*

***PRUEBA***

Sea la *k*-ésima columna de B; como las columnas de **A** forman una base para el espacio *m*-dimensional, entonces se tiene:

Donde son constantes definidas. Teniendo en cuenta que , con la matriz *(mxq)* con su *(j,k)-*ésimo elemento y :

Donde es la *j*-ésima columna de *C*. Entonces:

Ahora

Esto se debe a que ***A*** es una matriz ortogonal y las columnas de ***B*** son ortonormales. Por lo tanto

Y las columnas de la matrix ***C*** son ortonormales. La matriz ***C*** puede ser pensada como las primeras *q* columnas de la matriz ortogonal *(mxm)* ***D***. Ahora como las filas de la matriz ***D*** son ortonormales satisfacen , ahora como las filas de ***C*** consisten en las primeras filas de la matriz ***D***, se tiene que , por lo que:

Además, como es el coeficiente de en (13), la suma de los coeficientes es q (14) y ninguno de los coeficientes puede exceder 1 (15). Entonces

Será máximo si se puede encontrar un conjunto de para el cual:

Ahora, si consideramos , entonces

Dado que estos cumplen las condiciones anteriores podemos concluir entonces que alcanza su máximo cuando .

### Propiedad de los últimos elementos

*Para todo entero* ***q, con 1 < q < p****, considere la transformación ortonormal:*

*Donde* ***y*** *es un vector de q elementos y Bt es una matriz de dimensión q, y sea la matriz de varianza-covarianza para* ***y****. Entonces la traza de denotada por , es minimizada tomando , donde consiste en las últimas* ***q*** *columnas de A*

La prueba de esta propiedad es similar a la anterior pero en vez de encontrar los elementos que maximizan se hace con los elementos que minimizan, esto haría que la prueba se haga de menor a mayor.

La implicación estadística de esta segunda propiedad tiene propiedades estadísticas que permiten su uso en detección de outsiders o en exploración de relaciones estadísticas quasi-constantes lineales entre variables.

### Descomposición espectral

*La matriz admite una descomposición espectral y esta es:*

***PRUEBA***

De (10) se sabe que:

Si se expande el producto matricial del lado derecho se tiene que:

Como es necesario.

Un resultado interesante que deriva de lo anterior sería:

Este resultado presenta una utilidad estadística en la manera que nos muestra que no solo podemos descomponer la varianza conjunta de todos los elementos del vector de variables aleatorias ***x*** en contribuciones decrecientes por los PC, sino que podemos descomponer toda la matriz de covarianzas en contribuciones de de cada PC. Aunque no son estrictamente decrecientes los elementos de tienen la tendencia de decrecer mientras *j* incremente, esto se debe a que los valores propios decrecen a medida de que *j* incrementa mientras que tienden a tener el mismo tamaño debido a la condición de normalización impuesta para derivar los resultados:

Es interesante notar que *la propiedad de los primeros componentes* enfatiza en el hecho de que los PC explican de una manera efectiva tanto de la como sea posible , pero también se puede ver que de manera intuitiva son buenos explicando los componentes que no provengan de la diagonal. Particularmente este hecho es de utilidad si se aplica a matrices de correlacion donde los coeficientes están estandarizados entre 0 y 1.

### Propiedad de maximo determinante

*Como en las propiedades de los primeros y últimos componentes, considere la transformación:*

*Si denota el determinante de la matriz de covarianza* **y***, entonces es maximizado cuando*

***PRUEBA***

Considere cualquier entero, *k*, entre 1 y *q*, ahora sea . Data esta definición se sabe que donde dim denota la dimensión del espacio. El *k-*ésimo valor propio, de satisface:

Ahora suponga que , son los valores propios de la matriz y que son los vectores propios correspondientes. Ahora, sea con dim () = k.

Entonces, para todo vector no nulo en se sabe que:

Ahora considere el subespacio de la forma para en

Este resultado es un resultado general de la dimensión de dos espacios vectoriales:

Pero:

Entonces,

Por lo que se puede concluir que existe un vector no nulo en de la forma para un en , y también sucede que:

Viendo entonces que el *k*-ésimo valor propio de al *k-*ésimo valor propio de para k=1,…, q. Esto quiere decir que:

Finalmente si se toma ***B*** como entonces los valores propios de son:

Ahora, esta propiedad tiene una gran importancia práctica, en estadística multivariada el determinante de una matriz de covarianza se conoce como la varianza generalizada, esta se usa como un símil a la dispersión que puedan tener los datos. La raíz cuadrada de eso es proporcional al volumen ocupado por el espacio *m-*dimensional que ocupe la distribución de las variables aleatorias.

## Propiedades geométricas de los componentes principales

La idea es motivar la interpretación y el razonamiento de las propiedades de los componentes principales por medio de una aproximación geométrica.

### Propiedad de los ejes principales

*Considere la familia de los elipsoides m-dimensionales*

*Los componentes principales definen le eje principal de estos elipsoides.*

***PRUEBA***

Los componentes principales son definidos por la transformación antes presentada y dado que *A* es ortogonal, la transformación existe y es . Remplazando en la expresión original:

Es bien conocido que los vectores propios de son los mismos que los de , y que los valores propios de son los recíprocos de los de , asumiendo que son estrictamente positivos (como en el caso de la matriz de covarianza). Es entonces de una correspondencia directa con un resultado antes expuesto que:

Esta última ecuación puede ser re-escrita como:

Esta ecuación es el prototipo estándar de ecuación para un elipsoide definida por sus ejes principales.

También permite extraer la información de que sus ejes principales son proporcionales a .