

SOBRE EL ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

Isadore Nabi

I.	GENERALIDADES CONCEPTUALES Y MATEMÁTICAS DEL PCA	1
II.	INTUICIÓN GEOMÉTRICA Y MARCO CONCEPTUAL DEL PCA	5
III.	INTEPRETACIÓN DE LOS RESULTADOS DEL PCA	32
III.I.	Generalidades	32
III.II.	Interpretación de los Valores Característicos	32
III.II.I.	Criterios para la selección de los valores característicos	33
III.III.	Interpretación de los Componentes Principales	34
III.IV.	Distancia de Mahalanobis	36
III.V.	Gráfica de Sedimentación (Scree Plot)	37
III.VI.	Gráfica de Puntuaciones	38
III.VII.	Gráfica de Influencias	39
III.VIII.	Gráfica de Doble Proyección	40
III.IX.	Gráfica de Valores Atípicos	41
III.X.	Algunas Consideraciones Finales	42
IV.	INTUICIÓN GEOMÉTRICA EN \mathbb{R}^3	43
V.	REFERENCIAS	48

I. GENERALIDADES CONCEPTUALES Y MATEMÁTICAS DEL PCA

Según (Adler, 2012, pág. 96), “Los objetos en \mathbb{R} pueden tener muchas propiedades asociadas a ellos, llamadas atributos. *Estas propiedades explican qué representa un objeto y cómo debe ser interpretado por \mathbb{R} .* Con mucha frecuencia, la única diferencia entre dos objetos similares es que tienen atributos diferentes.”

Así, los atributos o variables de un conjunto de datos son las propiedades que lo caracterizan cualitativamente, mientras que los datos que tales variables contienen lo caracterizan cuantitativamente. Es en este sentido que se habla del concepto de *dimensionalidad*, el cual es utilizado para expresar la cardinalidad o medida del conjunto de atributos que caracterizan cualitativamente tal o cual conjunto de datos. El concepto de *dimensionalidad intrínseca* cristaliza la noción del número mínimo de variables necesarias para caracterizar cualitativamente y cuantitativamente un determinado conjunto de datos.

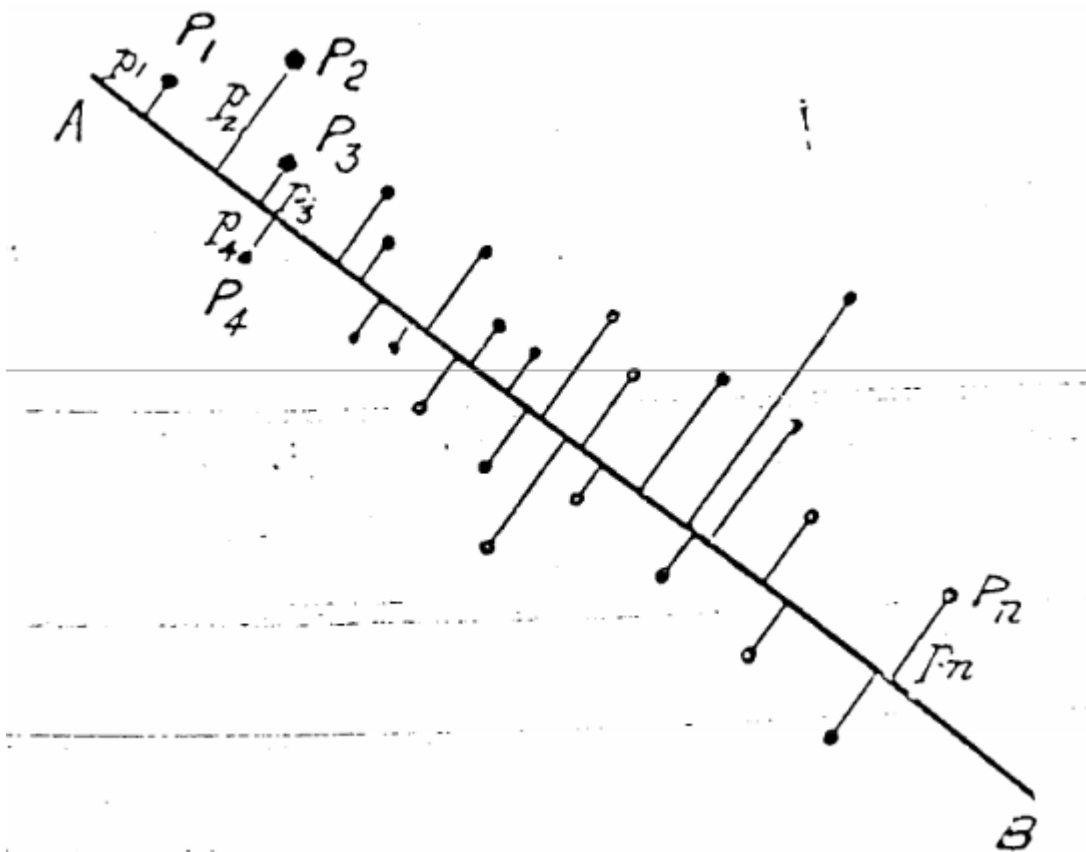
La *reducción de dimensionalidad* es un proceso que tiene como objetivo encontrar, dado el conjunto original de atributos o variables de un conjunto de datos, el mínimo subconjunto de tal conjunto de variables necesario para caracterizar de manera óptima cualitativa y cuantitativamente el conjunto de datos estudiado, lo cual pertenece a los dominios de la Optimización Combinatoria. En la búsqueda de reducir la dimensionalidad de un conjunto de datos surge el problema de la *maldición de la dimensionalidad*.

Este término acuñado por Richard Bellman en 1957 consiste, según sus propias palabras en que: “Sin embargo, hay algunos detalles a considerar. En primer lugar, la solución analítica eficaz de un gran número de ecuaciones incluso simples como, por ejemplo, ecuaciones lineales, es un asunto difícil. Bajar la mirada incluso con una solución computacional suele tener una serie de dificultades de naturaleza tanto burda como sutil. En consecuencia, la determinación de este máximo definitivamente no es rutinaria cuando el número de es grande. Todo esto puede subsumirse bajo el título “la maldición de la dimensionalidad.” (Bellman, 1972, pág. ix).

Intuitivamente, como se señala en (Wikipedia, 2020), esta categoría conceptualiza el hecho de que el común denominador de estos problemas radica en que cuando aumenta la dimensionalidad del conjunto de datos, el volumen del espacio (entendido como la estructura matemática conocida como *espacio muestral*) aumenta tan rápido que los datos disponibles se vuelven escasos. Esta escasez es problemática para cualquier método que requiera significación estadística. *Para obtener un resultado estadísticamente sólido y confiable, la cantidad de datos necesarios para respaldar el resultado a menudo aumenta exponencialmente con la dimensionalidad.* Además, la organización y la búsqueda de datos a menudo se basan en la detección de áreas donde los objetos forman grupos con propiedades similares; sin embargo, en datos de alta dimensión, todos los objetos parecen ser escasos y

diferentes en muchos aspectos, lo que impide que las estrategias comunes de organización de datos sean eficientes.

Es en el escenario antes descrito que el célebre estadístico marxista Karl Pearson planteó, en (Pearson, 1901), una metodología estadística que empleaba la técnica matemática de encontrar “la recta de mejor ajuste”, es decir, aquella ecuación de la recta que minimice la suma de las distancias entre cada observación (o punto x_n) y su respectiva proyección a esa dicha recta, en donde las distancias se asumen como ortogonales a la recta en cuestión, tal como se presenta a continuación.



Fuente: (Pearson, 1901, pág. 560).

Esta metodología estadística es actualmente conocida en el contexto del Aprendizaje Automático como *Análisis de Componentes Principales* (PCA, por su nombre en inglés). Como se señala en (Jolliffe, 2002, pág. 1), la idea central del análisis de componentes principales, desde la perspectiva de la Ciencia de Datos,

es reducir la dimensionalidad de un conjunto de datos que posee un gran número de variables interrelacionadas logrando retener tanto como sea posible la variabilidad actualmente presente en el conjunto de datos estudiado, puesto que esto implica que la distribución de tal conjunto de datos dentro del espacio muestral tras realizar la transformación ortogonal será isométricamente equivalente a la distribución del conjunto de datos original.

Así, supóngase que x es un vector compuesto por p variables aleatorias y que tanto la varianza de las p variables aleatorias como también la estructura de las covarianzas o correlaciones entre las p variables son de interés en una investigación. Salvo que p sea pequeño (pocas variables) o que la estructura de las correlaciones sea muy simple, no será de mucha utilidad simplemente considerar las p varianzas y todas las $\frac{1}{2}p(p - 1)$ correlaciones o covarianzas.

Un enfoque alternativo para superar este problema es el propuesto por Pearson, que desde lo planteado anteriormente consiste en buscar un subconjunto de variables derivadas del conjunto de variables original (cuya cardinalidad será estrictamente menor a la del original) que preserve la mayor parte de información dadas tales varianzas y correlaciones o covarianzas. Nótese que en (Jolliffe, 2002, pág. 1) al hablar de “la mayor parte de información” no se está hablando únicamente de una preservación cuantitativa, sino y dado que lo anterior se realiza mediante optimización combinatoria de forma analítica e iterativamente a nivel computacional-informático, se está hablando también de una preservación cualitativa, lo que implica que la metodología planteada por Pearson, sea que se ejecute analítica o numéricamente, cristaliza la noción de preservar “la información esencial”.

Aunque el PCA no ignora las covarianzas y correlaciones, se enfoca en las varianzas. El primer paso de esta metodología estadística es encontrar una función lineal $\alpha_1' x$ de elementos x que tengan varianza máxima, en donde α_1 es un vector

de p constantes $\alpha_{11}, \alpha_{12}, \dots, \alpha_{1p}$ y $'$ denota una transposición (en el sentido del Álgebra Lineal), tal que:

$$\alpha'_1 x = \alpha_{11} x_1 + \alpha_{12} x_2 + \dots + \alpha_{1p} x_p = \sum_{j=1}^p \alpha_{1j} x_j$$

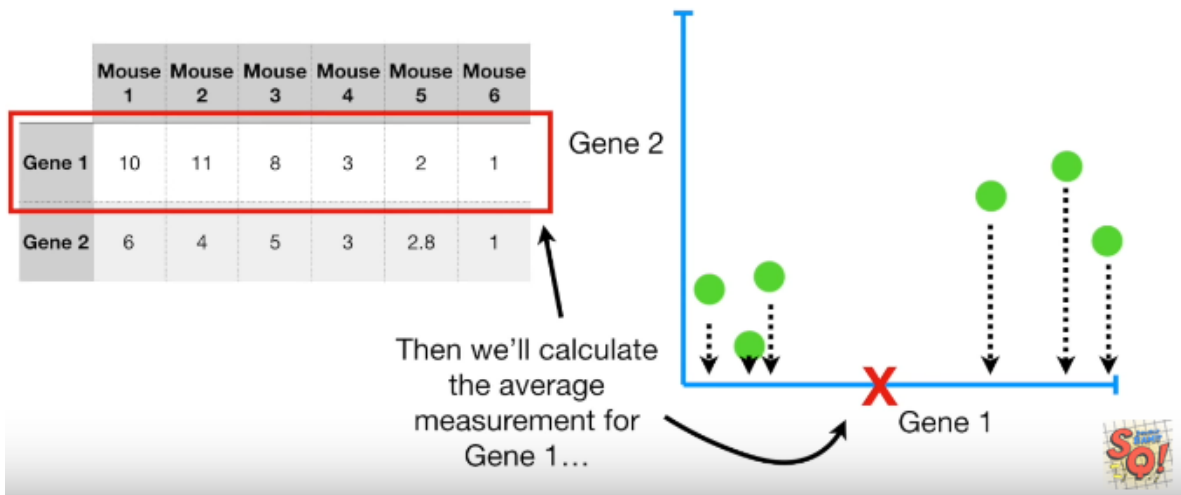
El segundo paso es buscar una función lineal $\alpha'_2 x$ no correlacionada con $\alpha'_1 x$ siendo de varianza máxima y así sucesivamente en los $k - \text{ésimos}$ pasos siguientes hasta encontrar una función lineal $\alpha'_k x$ con varianza máxima sujeta a la restricción de estar no-correlacionada con $\alpha'_1 x, \alpha'_2 x, \dots, \alpha'_{k-1} x$. La $k - \text{ésima}$ variable derivada del conjunto de datos original, *i.e.*, $\alpha'_k x$, es el $k - \text{ésimo}$ **componente principal**. Así, como señala (Adler, 2012, pág. 357), “El análisis de componentes principales divide un conjunto de variables (posibles correlacionadas) en un conjunto de variables no correlacionadas.”

Lo anterior puede explicarse con un ejemplo en un espacio de dos dimensiones tomado de (Starmer, 2018).

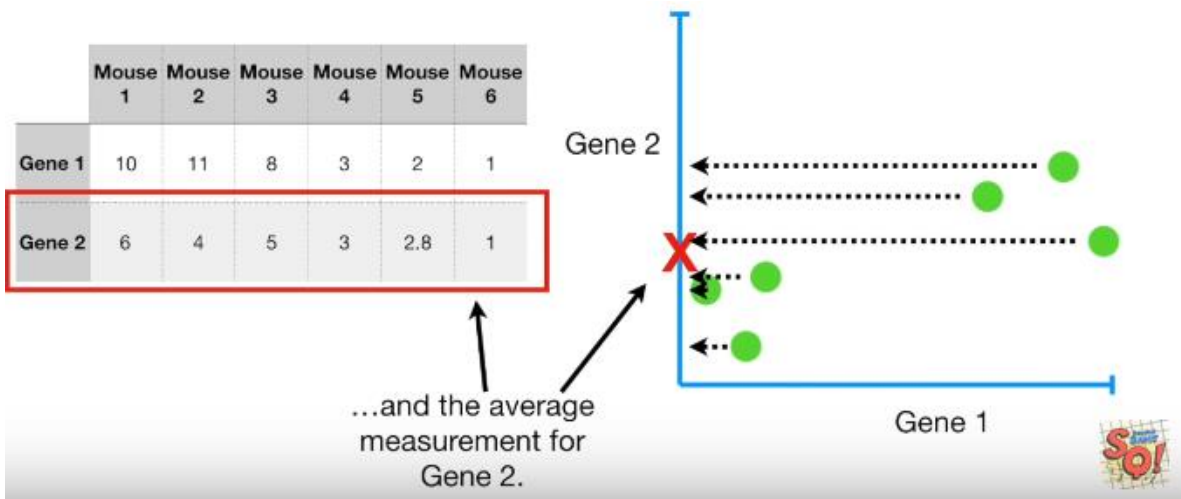
II. INTUICIÓN GEOMÉTRICA Y MARCO CONCEPTUAL DEL PCA

Supóngase que se está estudiando la transcripción genética¹ de dos genes en un conjunto de seis ratones y que se dispone de tal información en el cuadro presentado a continuación.

¹ “La transcripción del ADN es el primer proceso de la expresión genética, mediante el cual se transfiere la información contenida en la secuencia del ADN hacia la secuencia de proteína utilizando diversos ARN como intermediarios. Durante la transcripción genética, las secuencias de ADN son copiadas a ARN mediante una enzima llamada ARN polimerasa (ARNp) la cual sintetiza un ARN mensajero que mantiene la información de la secuencia del ADN. De esta manera, la transcripción del ADN también podría llamarse síntesis del ARN mensajero.” (Wikipedia, 2021).



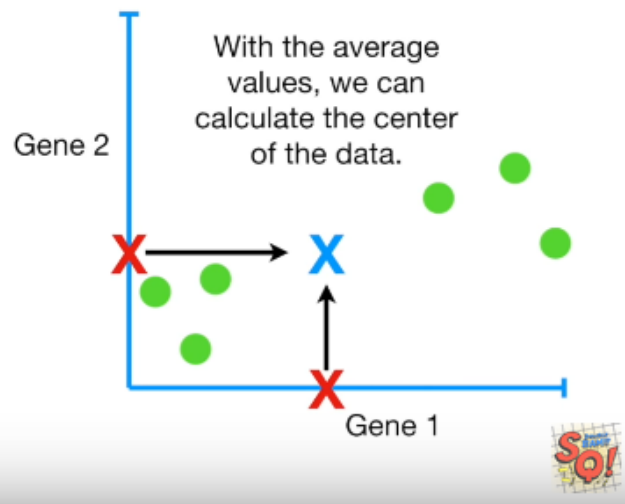
Fuente: (Starmer, 2018).



Fuente: (Starmer, 2018).

Con los valores promedio de la transcripción genética es posible calcular el centro de la distribución del conjunto de datos, tal como se muestra a continuación.

	Mouse 1	Mouse 2	Mouse 3	Mouse 4	Mouse 5	Mouse 6
Gene 1	10	11	8	3	2	1
Gene 2	6	4	5	3	2.8	1



Fuente: (Starmer, 2018).

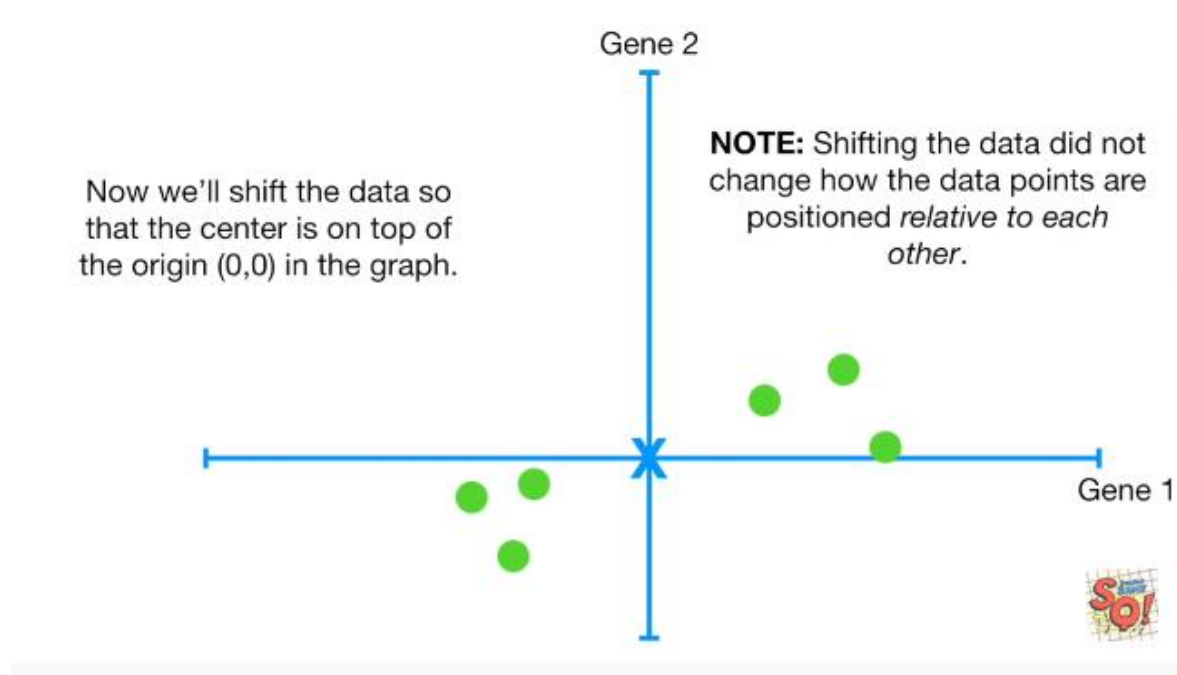
Así, es posible realizar transformaciones² sobre el conjunto de datos para situar en el centro de su distribución el origen, que para el caso bidimensional es el punto (0,0), sin afectar la posición relativa de las observaciones dentro de la estructura espacial conocida como *distribución* del conjunto de datos.

² Como se verifica en (Weisstein, Transformation, 2021), una transformación T (i.e., un mapa, una función) sobre un dominio D toma elementos $X \in D$ y los transforma (mediante operaciones matemáticas) en elementos $Y \in D$, en donde el rango (i.e., la imagen) de T es definida como $Rango(T) = T(D) = \{T(X) : X \in D\}$. Los tipos de transformaciones geométricas existentes son resumidas en el siguiente cuadro:

Transformation	Characterization
dilation	center of dilation, scale decrease factor
expansion	center of expansion, scale increase factor
reflection	mirror line or plane
rotation	center of rotation, rotation angle
shear	invariant line and shear factor
stretch (1-way)	invariant line and scale factor
stretch (2-way)	invariant lines and scale factors
translation	displacement vector

Fuente: (Weisstein, Transformation, 2021).

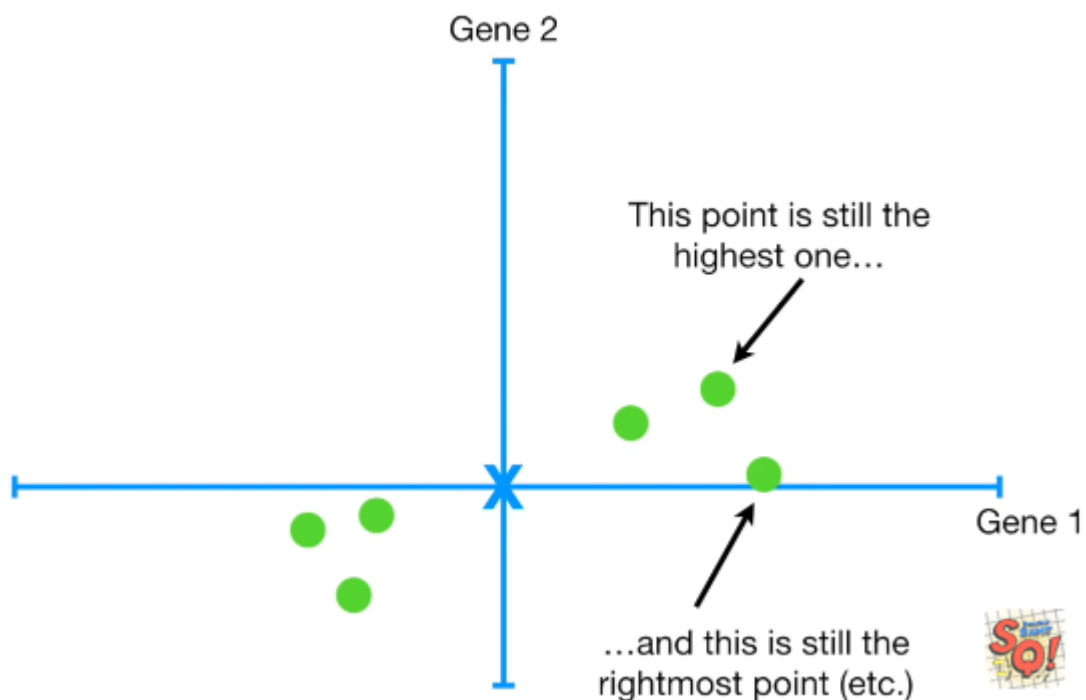
Que no se altere la posición relativa de los objetos que conforman una determinada estructura implica que su topología se conserva, *i.e.*, existe una función isomórfica que relaciona ambas estructuras o, dicho de otra forma, existe un isomorfismo que relaciona ambas estructuras. La existencia de una función isomórfica que relaciona estas estructuras implica que entre la estructura algebraica original y la nueva estructura algebraica existe una función biyectiva (que establece una relación uno-a-uno entre los elementos de ambas estructuras) que transforma una estructura en la otra preservando las operaciones definidas con antelación dentro de la estructura original (sean estas cuales sean) y, además, que es posible encontrar para estas estructuras (la original y la nueva resultante de la transformación) sus respectivas estructuras inversas (que es una generalización del simple concepto algebraico de inverso multiplicativo).



Fuente: (Starmer, 2018).

Al conservarse la topología de las estructuras se conserva también, puesto que se están estudiando espacios normados (espacios vectoriales que tienen una norma - una función que mide la longitud entre objetos-), la norma dentro de dicha

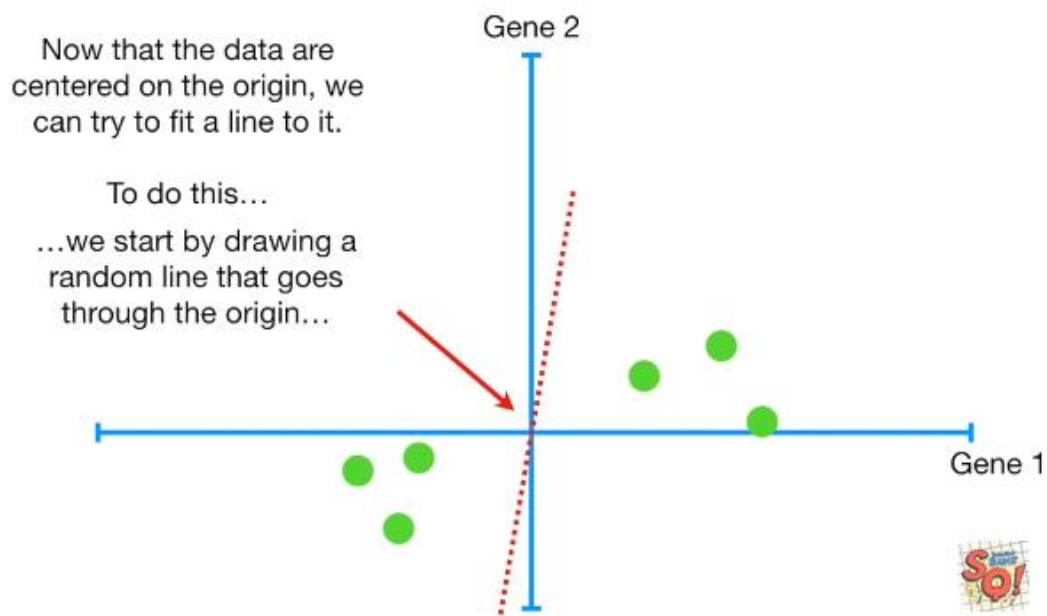
estructura. Finalmente, al conservarse la norma, puesto que esto ocurre en el contexto de los espacios euclidianos (y en los espacios de Hilbert, que son su generalización), espacios para los cuales la métrica es inducida por una norma, también se conserva la métrica. Así, en el contexto del Álgebra Lineal, una isometría (preservación de la métrica -función que sirve para medir distancias dentro del espacio-) implica un isomorfismo, puesto que una isometría es un isomorfismo en el contexto de espacios métricos, como se verifica en (Stack Exchange, 2015) y (Wikipedia, 2021)³.



³ Como se señala en (Wikipedia, 2020), en el siglo XX se ha precisado en matemáticas la noción intuitiva de estructura, siguiendo la concepción de Aristóteles de la materia y la forma, según la cual cada estructura es un conjunto X dotado de ciertas operaciones (como la suma o el producto) o de ciertas relaciones (como una ordenación) o ciertos subconjuntos (como en el caso de la topología), etc. En este caso, el conjunto X es la materia y las operaciones, relaciones, etc., en él definidas, son la forma. La concepción de Platón, filosóficamente equivocada, de que la forma es lo que importa, se recoge en Matemáticas con el concepto de isomorfismo. Debido a lo anterior, los isomorfismos son estructuras matemáticas que no están sujetas a una clasificación en concreto (como resultado de su equivalencia), lo que se conoce en esta ciencia como que las estructuras matemáticas deben clasificarse *salvo* isomorfismos, que no denota excepción o exclusión, sino inclusión o equivalencia. Lo anterior implica, en otras palabras, que como pueden pertenecer a cualquier clasificación, no pertenecen a ninguna en concreto.

Fuente: (Starmer, 2018).

Así, ahora que el conjunto de datos está centrado en el origen, prosigue ajustar una recta (la que proporcione el mejor ajuste, la cual puede ser cualquiera -en concreto, su forma matemática dependerá del conjunto de datos del que se trate-) que corte el espacio (partiendo del origen) en dos partes, tal como se muestra a continuación.

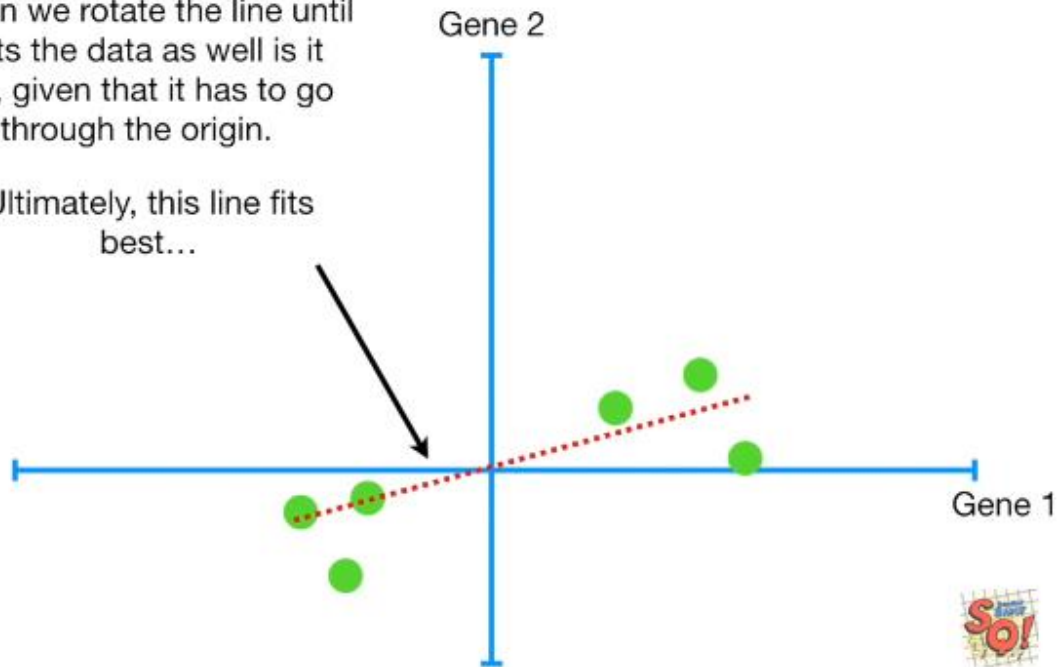


Fuente: (Starmer, 2018).

Para cuantificar cuán bueno es el ajuste de un conjunto de datos a una recta, la metodología PCA realiza proyecciones ortogonales. Como se verifica en (MIT, 2021, pág. 1), una proyección es una transformación lineal (*loc. cit.*) en la que no se emplean coordenadas matriciales a nivel puramente matemático, sin embargo, algo distinto ocurre cuando se requiere su estimación empírica-computacional.

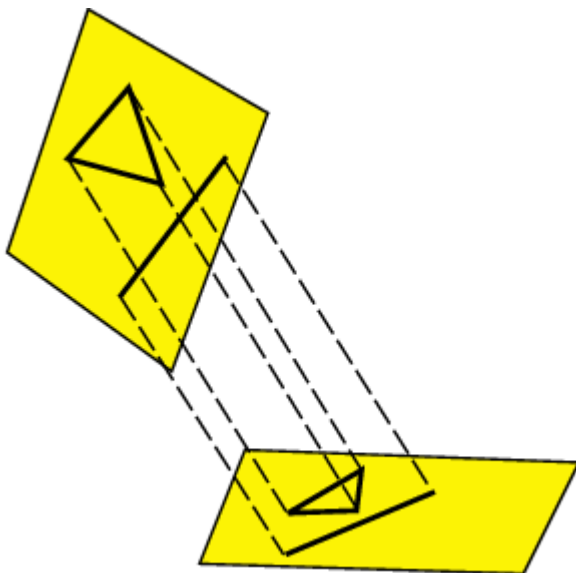
...then we rotate the line until it fits the data as well as it can, given that it has to go through the origin.

Ultimately, this line fits best...



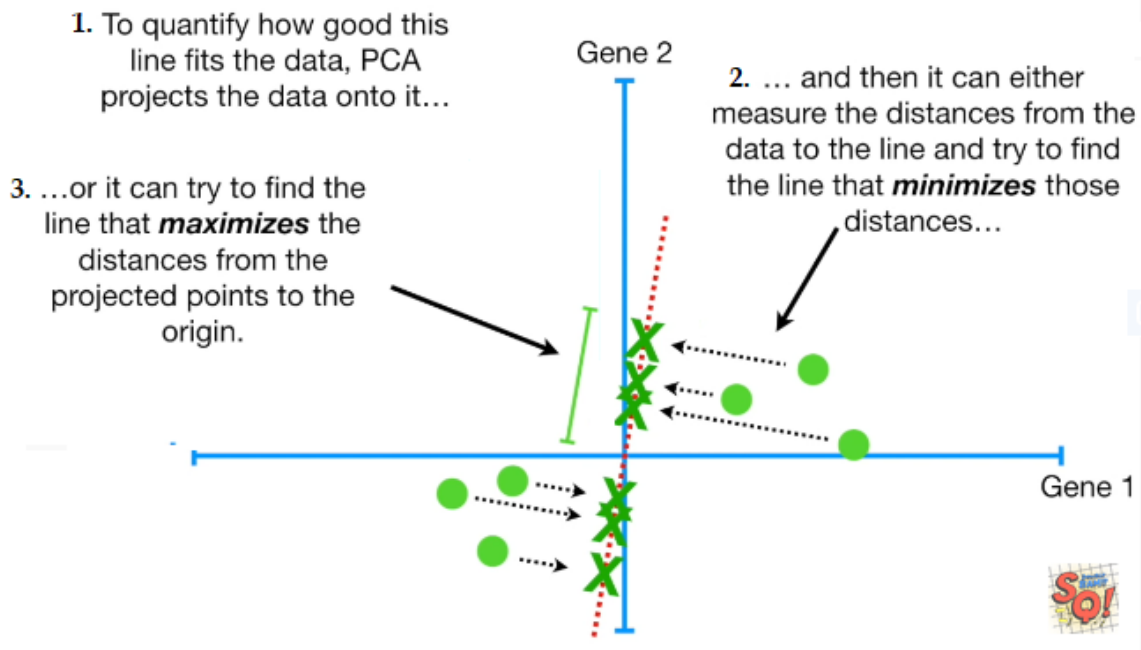
Fuente: (Starmer, 2018).

Como se señala en (Weisstein, Projection, 2021), una proyección toma el siguiente aspecto geométrico a nivel de dos dimensiones.



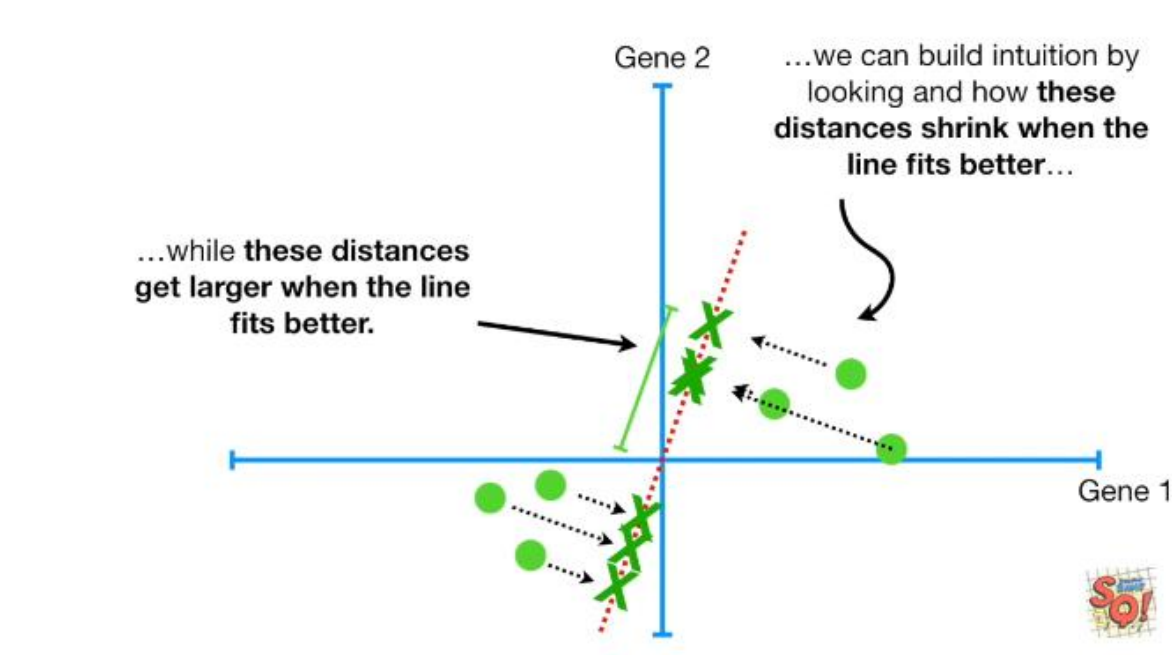
Fuente: (Weisstein, Projection, 2021).

Como señala (Weisstein, Projection, 2021), una proyección es la transformación de puntos y líneas en un plano en otro plano conectando los puntos correspondientes en los dos planos con líneas paralelas. La rama de la geometría que se ocupa de las propiedades e invariantes de las figuras geométricas proyectadas se denomina *geometría proyectiva*, descubierta por el gran matemático ruso Nikolái Lobachevski.



Fuente: (Starmer, 2018).

Que una proyección sea ortogonal significa que las líneas paralelas que conectan los puntos con sus proyecciones son perpendiculares (forman un ángulo de 90 grados) a la recta a la cual se están ajustando los puntos. Respecto a la equivalencia planteada en la imagen anterior respecto a los puntos de partida 2 (minimizar) y 3 (maximizar), en el escenario 2 el proceso de optimización se está realizando de forma analítica (mediante el proceso de optimización matemática antes descrito), mientras que en el escenario 3 se está realizando de forma iterativa a través de métodos numéricos.



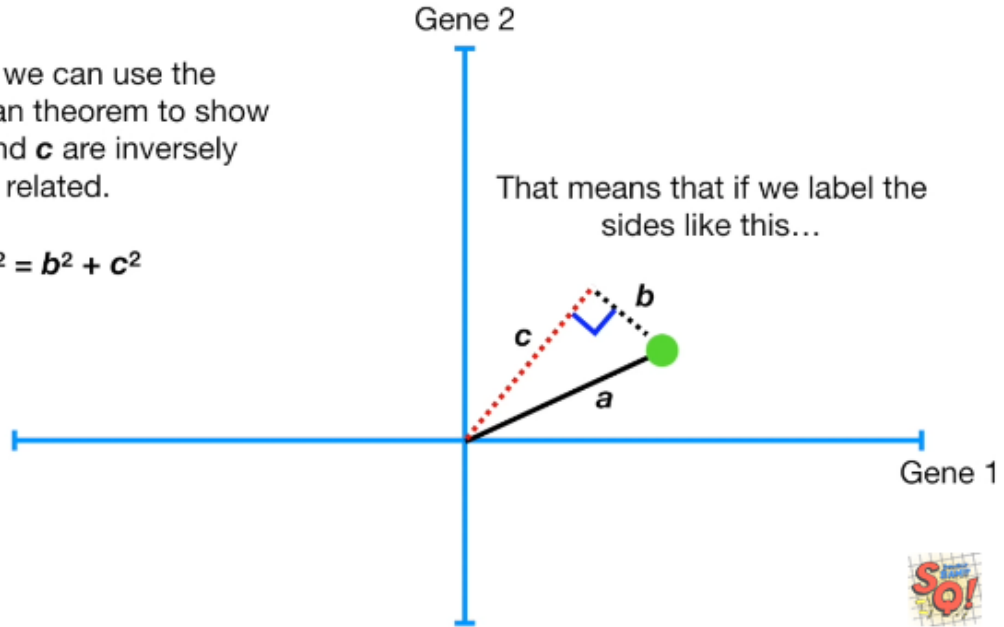
Fuente: (Starmer, 2018).

En relación a la estructura que en conjunto forman cada una de las observaciones (o puntos de datos), cada observación o punto está fijo y, por consiguiente, su distancia al origen será también fija. Esto significa, como ya se vio, que la distancia de un punto al origen no cambia cuando la línea punteada roja de las imágenes anteriores (que representa la recta a la cual se desean ajustar las observaciones mediante transformaciones ortogonales) sufre alguna transformación, en este caso una rotación. Así, cuando se proyecta el punto hacia la recta de mejor ajuste (asumida como la línea punteada roja) se genera un ángulo de 90 grados entre la recta de mejor ajuste y la recta paralela mediante la cual se hace la proyección del punto a la recta de mejor ajuste en cuestión. Lo anterior se muestra a continuación tomando de referencia únicamente un punto, con la finalidad de facilitar su comprensión.

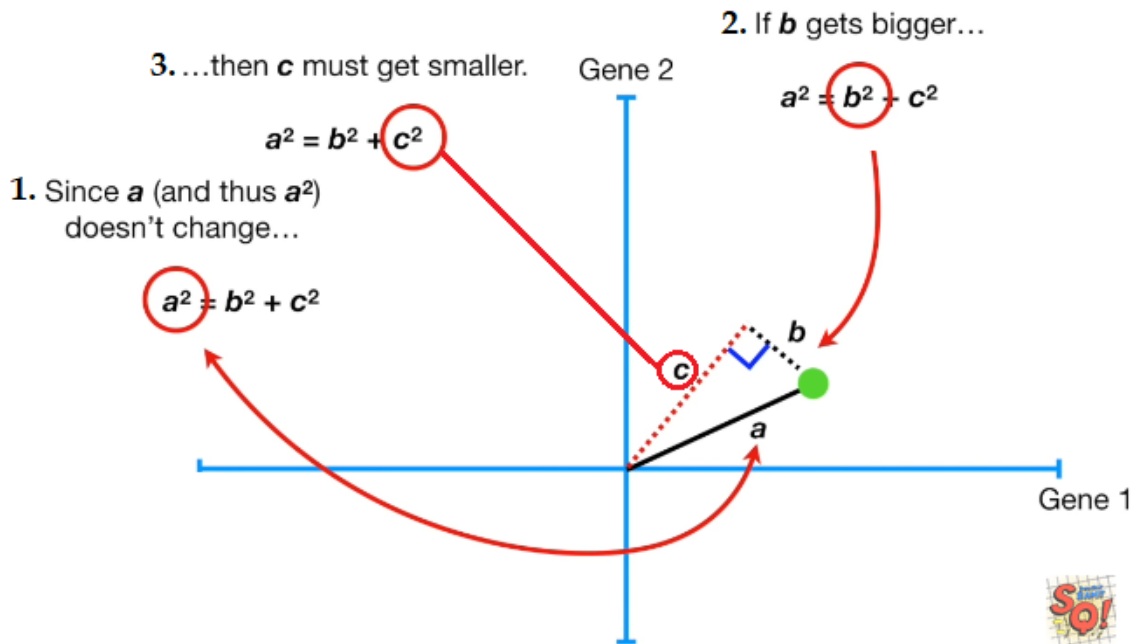
...then we can use the Pythagorean theorem to show how **b** and **c** are inversely related.

$$a^2 = b^2 + c^2$$

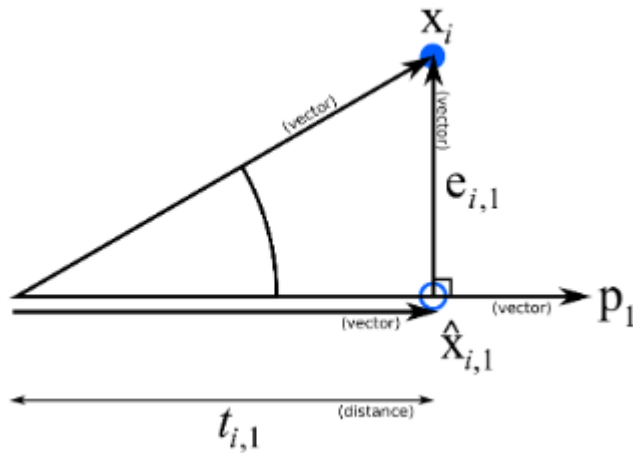
That means that if we label the sides like this...



Fuente: (Starmer, 2018).

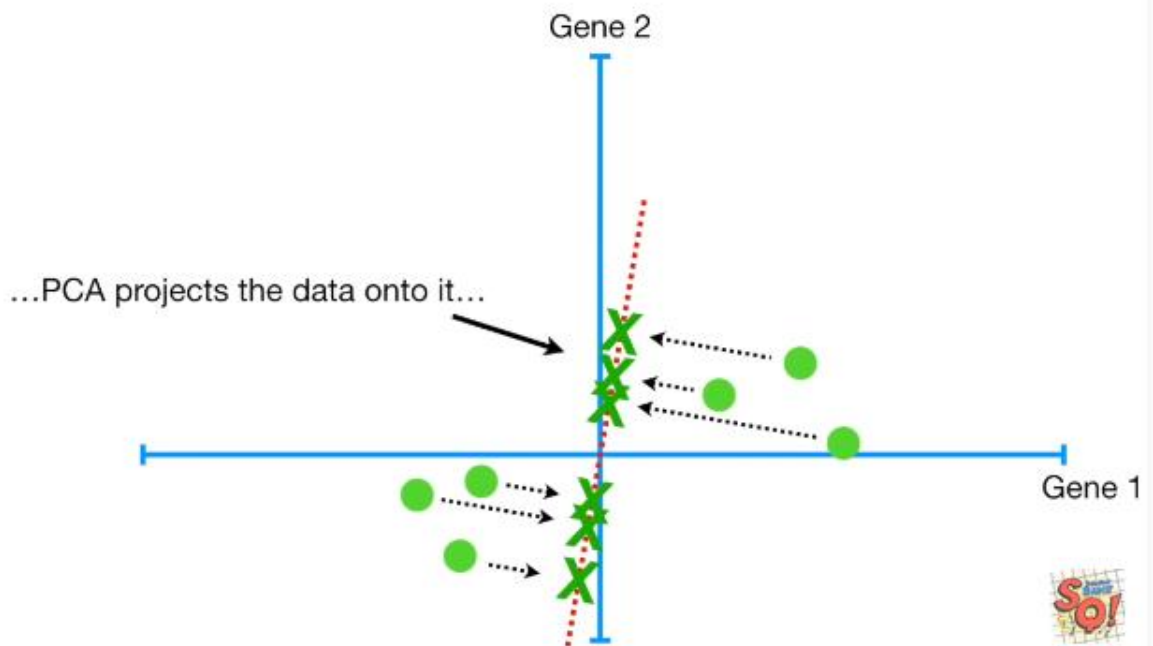
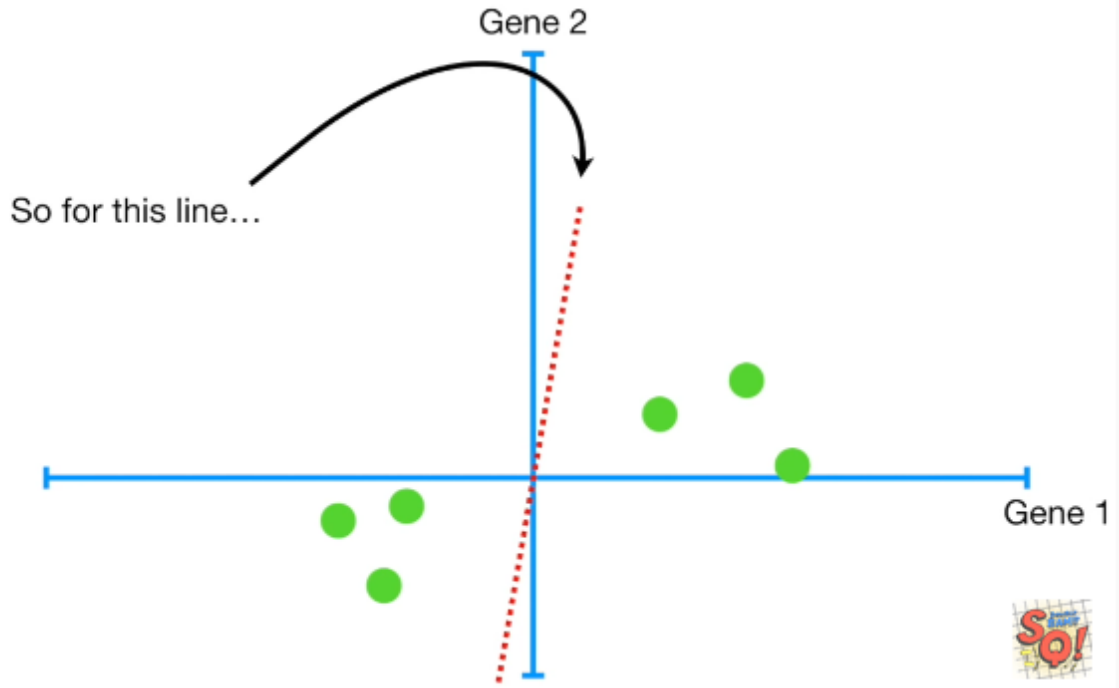


Fuente: (Starmer, 2018).

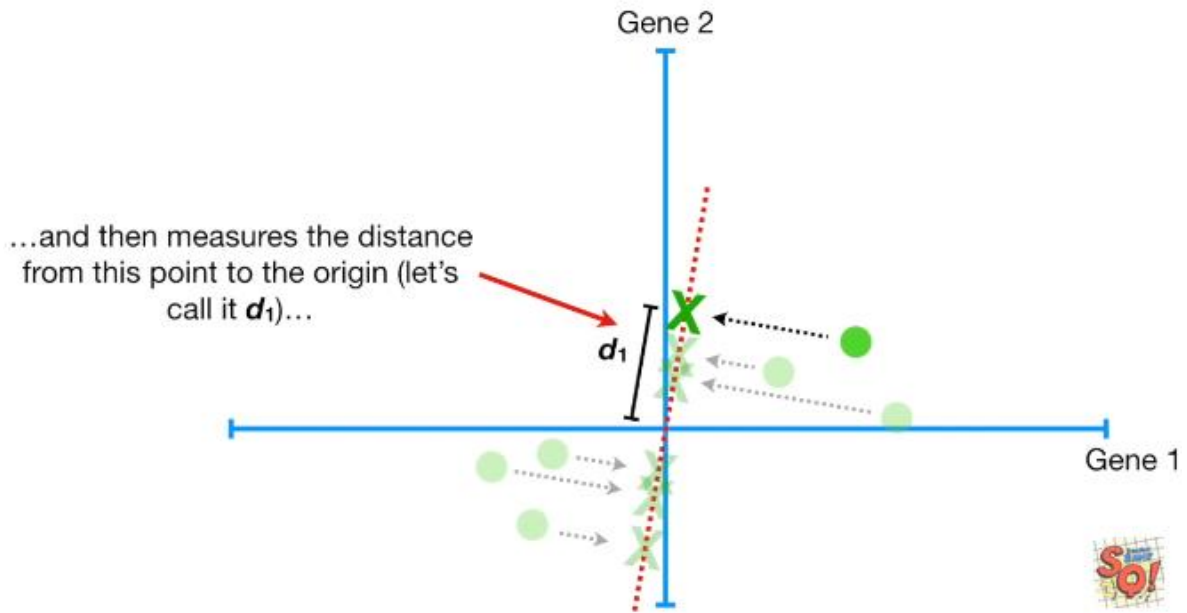


Fuente: (Dunn, 2021, pág. 342).

Así, el PCA puede realizarse a través de dos metodologías. La primera es la metodología analítica, la cual consiste en minimizar la suma de las distancias entre cada punto original y su respectiva proyección ortogonal a esa misma recta. La segunda es la metodología numérica, la cual consiste en construir una recta que maximice la distancia de los puntos proyectados al origen, como señala el genetista de la Universidad de Carolina del Norte en Chapel Hill. La razón por la que numéricamente se hace a través de la maximización y no de la minimización (como ocurre analíticamente) es porque el costo computacional (sea de computadores informáticos o de simplemente computar) es inferior si se maximiza la suma de las distancias al cuadrado desde las observaciones (puntos de datos) proyectadas hasta el origen en lugar de minimizar las distancias del punto a la recta que le sirve de proyección ortogonal.

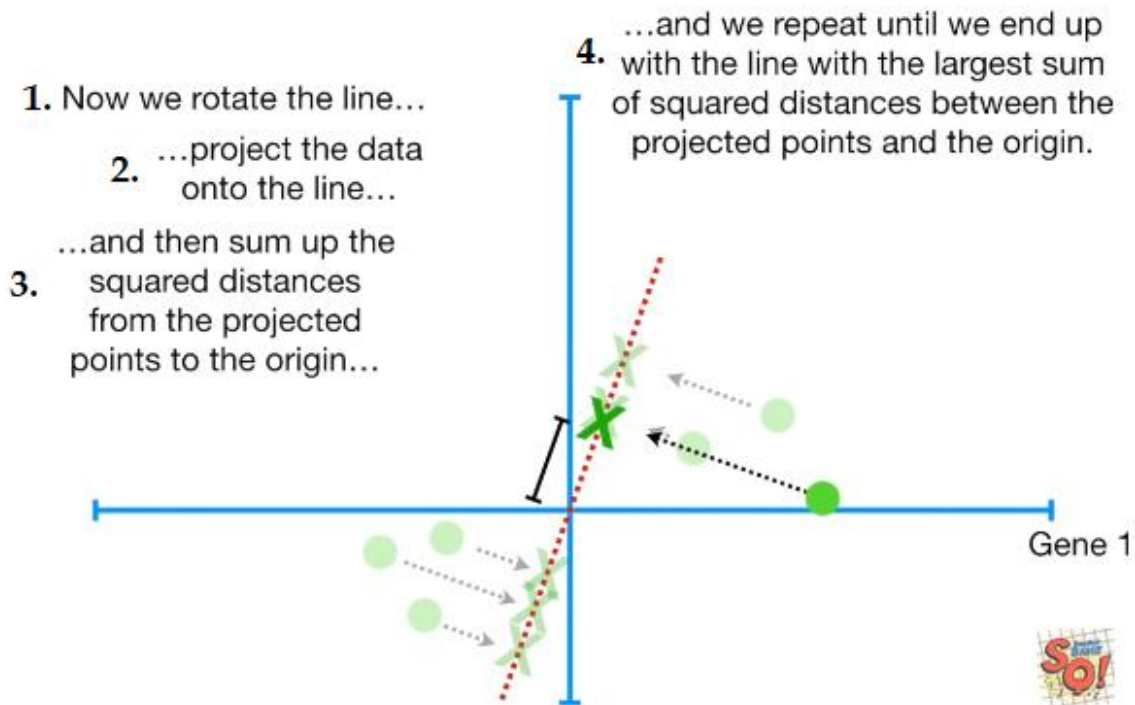


Fuente: (Starmer, 2018).



Fuente: (Starmer, 2018).

Y así se hará con todas las demás distancias (correspondientes a los demás puntos de datos u observaciones), obteniendo el resultado presentado a continuación.

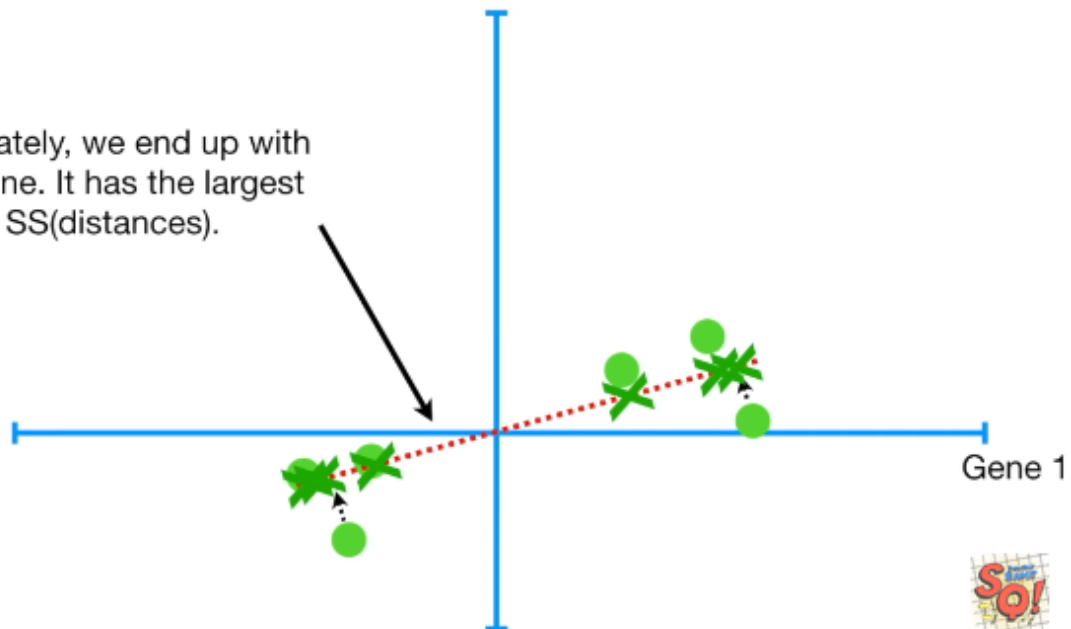


$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2 = \text{sum of squared distances} = \text{SS}(\text{distances})$$

Fuente: (Starmer, 2018).

$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2 = \text{sum of squared distances} = \text{SS}(\text{distances})$$

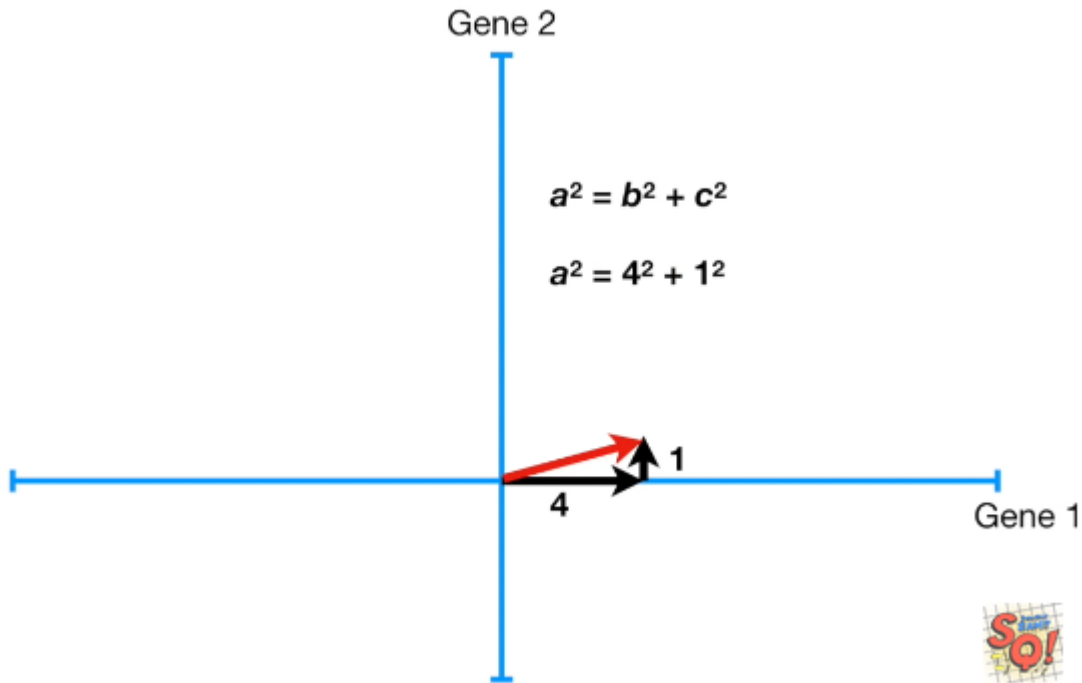
Ultimately, we end up with this line. It has the largest SS(distances).



Fuente: (Starmer, 2018).

El procedimiento, descrito en términos de su intuición geométrica en las cinco imágenes anteriores, es el correspondiente a la metodología numérica. La línea roja que contiene la distancia más larga es conocida como *componente principal 1* (PC1), en donde a manera de ejemplo, existe una pendiente con un valor de 0.25, lo cual no es otra cosa que el valor de la derivada evaluada en ese punto, con la interpretación teórica usual de la derivada (sea como pendiente o como tasa de cambio instantánea). Esta derivada α_{1j} es de importancia fundamental al describir la forma en que los puntos de datos se distribuyen en el plano cartesiano, puesto que en conjunto con los valores concretos que toman las variables x_j , específicamente mediante una combinación lineal, conforman los PC_i , que es precisamente la metodología para realizar inferencias sobre la distribución de las observaciones dentro del espacio muestral.

Computacionalmente hablando, los PC se encuentran calculando las distancias mediante la desigualdad de Minkowski (en espacios \mathcal{L}_p , que son la generalización natural de los espacios euclidianos), mediante la desigualdad Cauchy-Bunyakovsky-Schwarz en general (aplicable hasta espacios \mathcal{L}_2) o mediante el teorema de Pitágoras en estructuras matemáticas más simples (puramente euclidianas), tal como se presenta a continuación.



Fuente: (Starmer, 2018).

De lo anterior se desprende que el valor de a es de 4.12 unidades de medición (cualesquiera que estas sean). Tras realizar las mediciones anteriores, debe procederse a calcular el primer componente principal (PC1) a través de la técnica de *descomposición en valores singulares* empleada para calcular desde la metodología numérica los componentes principales, en donde tales valores se expresan matemáticamente como las raíces cuadradas de los valores característicos. Lo anterior expresa, en su forma más general, realizar un reajuste métrico al conjunto de relaciones fundamentales (cristalizadas en las funciones que orquestan a los valores característicos, que expresan a su vez la esencia del sistema) que describen el fenómeno natural o social de estudio, que tiene por objetivo cuantificar tales relaciones esenciales con la máxima precisión posible.

Después de la explicación anterior conviene regresar a las intuiciones geométricas expuestas por Josh Starmer, específicamente antes mostrar su exposición sobre los *valores característicos* y los *valores singulares* en términos de tales intuiciones, es necesario remitir al lector a la investigación (Nabi, Una Interpretación

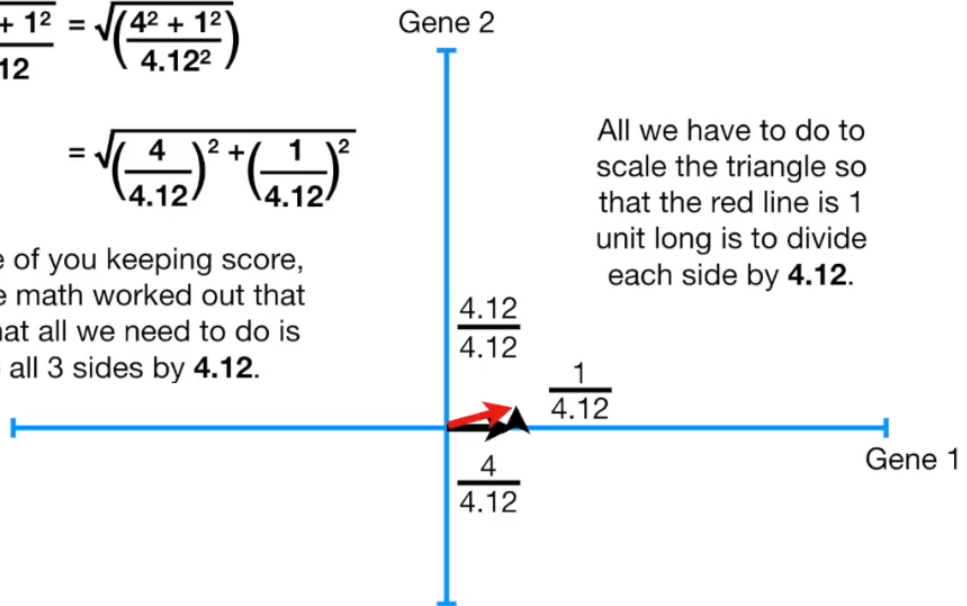
Multidisciplinaria de los Espacios Característicos, Vectores Característicos y Valores Característicos, 2021) para comprender el significado más general de los valores característicos en el contexto de los espacios característicos.

Así, lo primero que se debe hacer con el punto de dato u observación estudiada (y así para todas las demás observaciones (geoméricamente Josh Starmer está trabajando únicamente con una observación, pero las acciones ejecutadas sobre esa observación deben ser igualmente ejecutadas sobre las $n - 1$ observaciones restantes) es re-escalarla, específicamente estandarizarla. Como se verifica en (Nabi, ¿Por qué se realiza un ajuste por re-escalamiento, normalización o estandarización sobre los datos en el contexto del aprendizaje automático?, 2021), estandarizar implica transformar una variable de tal forma que su comportamiento estocástico pueda ser modelado por una distribución normal estándar y, en el análisis puramente geométrico realizado por Starmer, esto se ejecuta meramente mediante la división de todos los lados del triángulo rectángulo generado (implícito en el uso del teorema de Pitágoras) por la hipotenusa de dicho triángulo (que en el ejemplo es de 4.12) y así, la desviación estándar (que está expresada en la hipotenusa del triángulo en cuestión) será de longitud unitaria y, al realizar esto con las $n - 1$ observaciones restantes, el promedio de los valores estandarizados será de cero, distribuyéndose los puntos de datos u observaciones como una distribución normal con media nula y varianza unitaria.

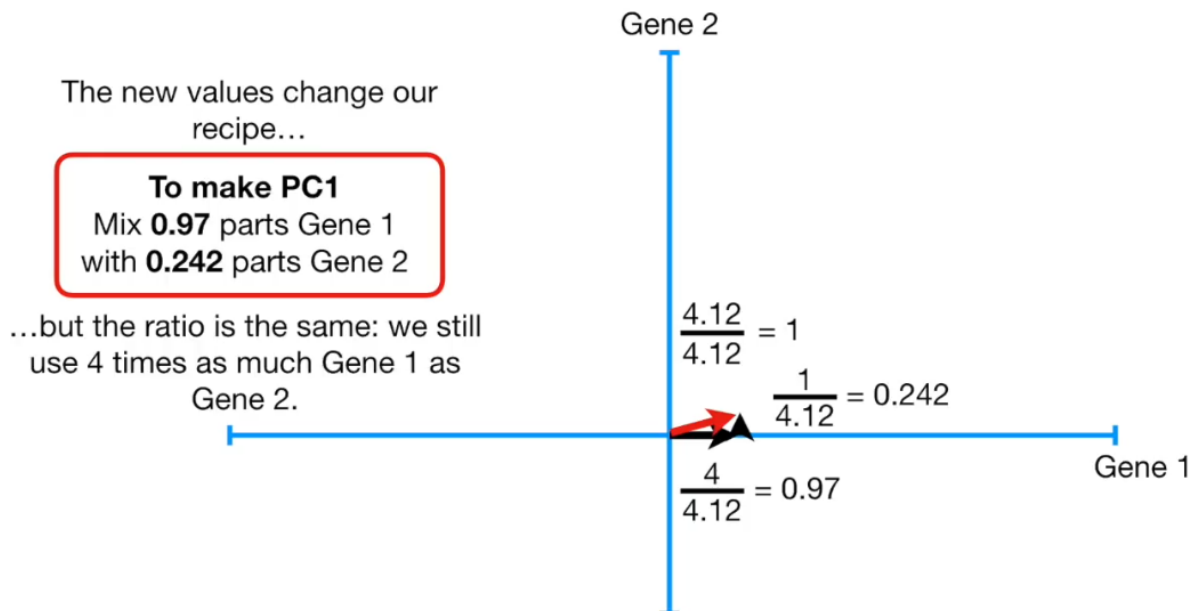
$$\frac{4.12}{4.12} = \frac{\sqrt{4^2 + 1^2}}{4.12} = \sqrt{\left(\frac{4^2 + 1^2}{4.12^2}\right)}$$

$$= \sqrt{\left(\frac{4}{4.12}\right)^2 + \left(\frac{1}{4.12}\right)^2}$$

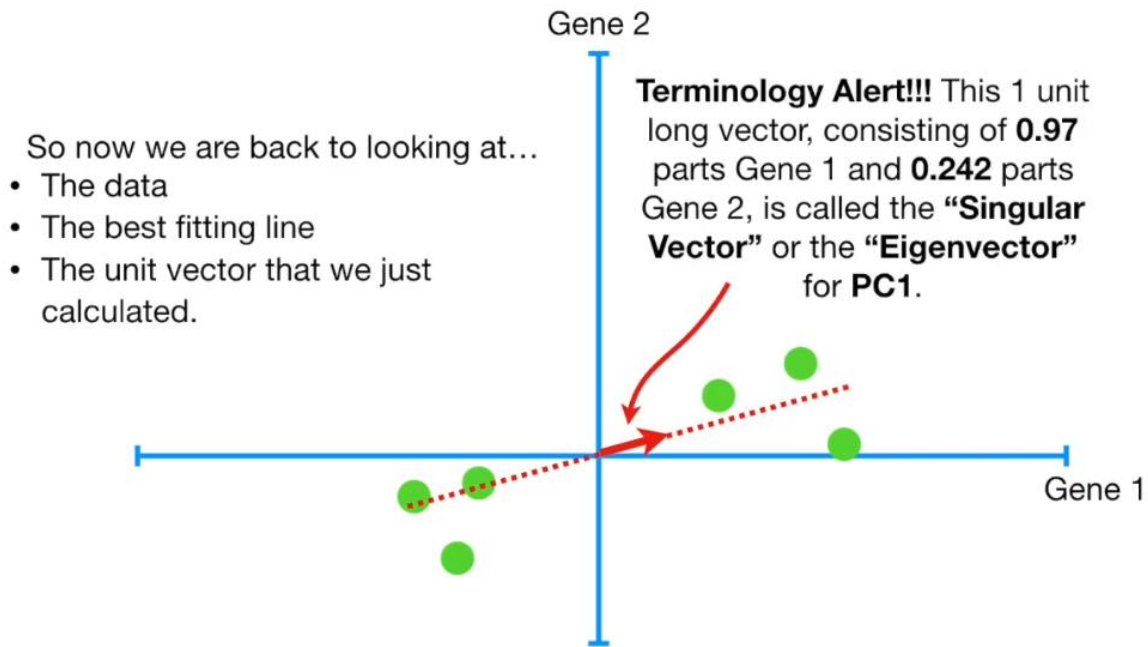
For those of you keeping score, here's the math worked out that shows that all we need to do is divide all 3 sides by **4.12**.



Fuente: (Starmer, 2018).



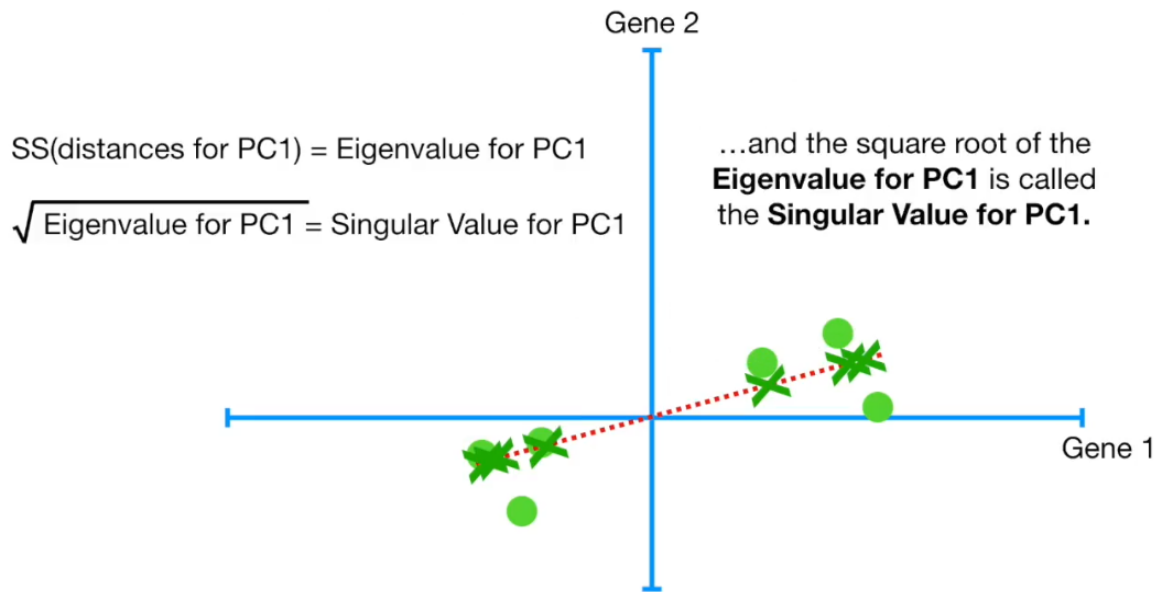
Fuente: (Starmer, 2018).



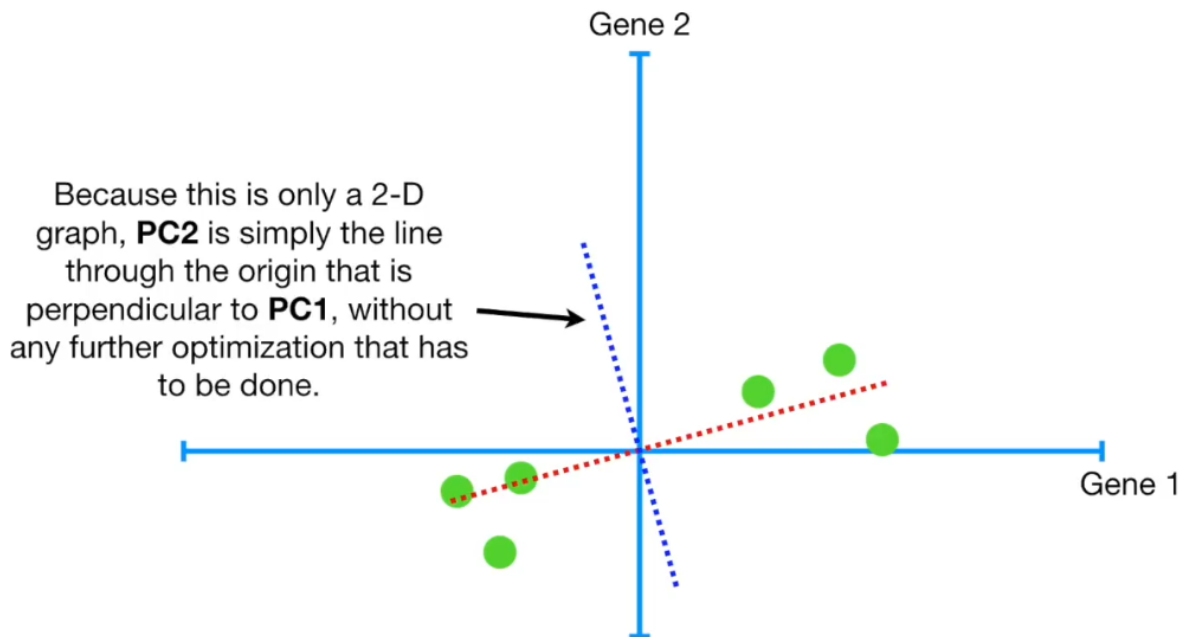
Fuente: (Starmer, 2018).

Cada distancia estandarizada (cada hipotenusa de longitud unitaria), compuesta parcialmente por una de las variables que conforman el fenómeno de estudio y parcialmente por la otra variable de estudio y (así será para el caso de $n - \text{ésimas}$ variables de estudio) en las proporciones expuestas por Starmer [que obedecen a que los valores de la pendiente, la cual expresa las proporciones en que todas las variables de estudio conforman la estructura del fenómeno de estudio (que son equivalentes a la derivada en ese punto), han sido divididos entre la hipotenusa (cuyo valor es de 4.12)] expresa los vectores característicos del espacio muestral analizado. Así, el numerador y el denominador de la pendiente o, lo que es lo mismo, las proporciones en que se interrelacionan cada una de las variables que componen al fenómeno analizado para generar una observación de dicho fenómeno (en este caso las variables son el gen 1 y el gen 2), son usualmente conocidas como *puntajes de carga*, que pueden entenderse a nivel puramente estadístico como las ponderaciones de cada variable original al calcular el componente principal.

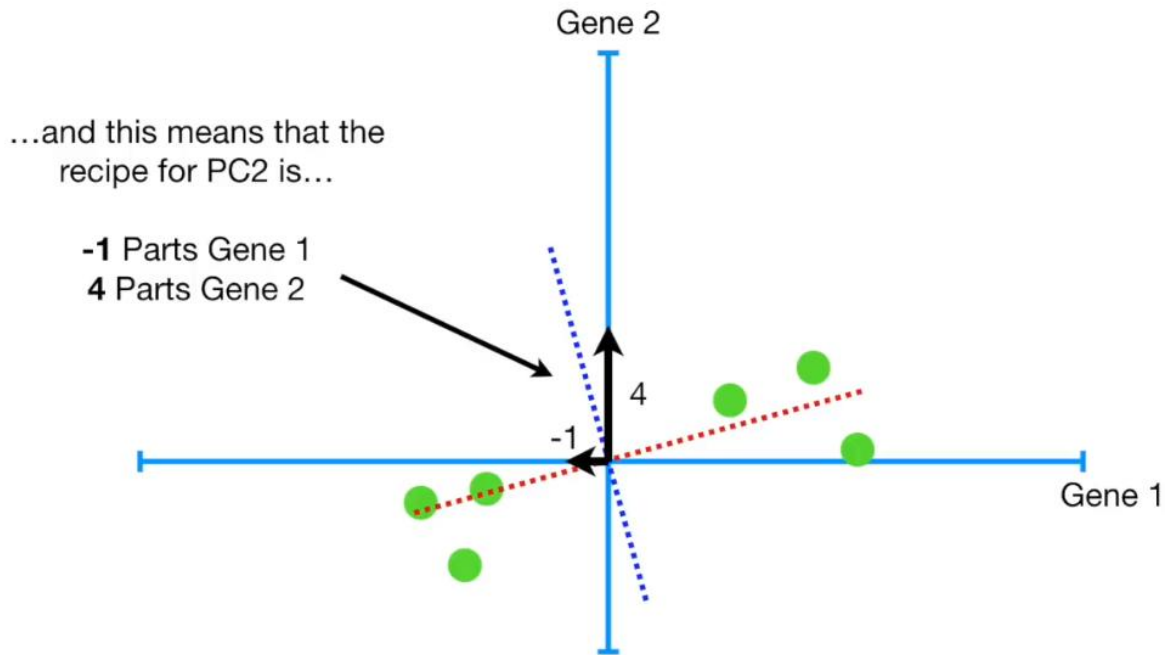
$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2 = \text{sum of squared distances} = \text{SS}(\text{distances})$$



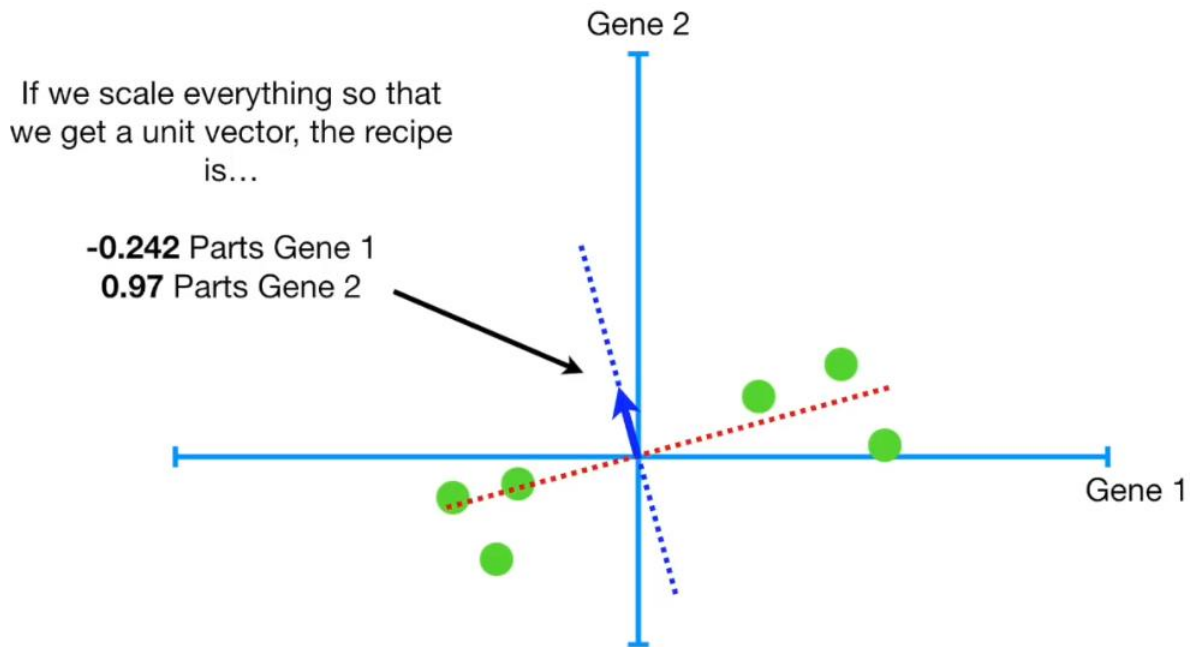
Fuente: (Starmer, 2018).



Fuente: (Starmer, 2018).



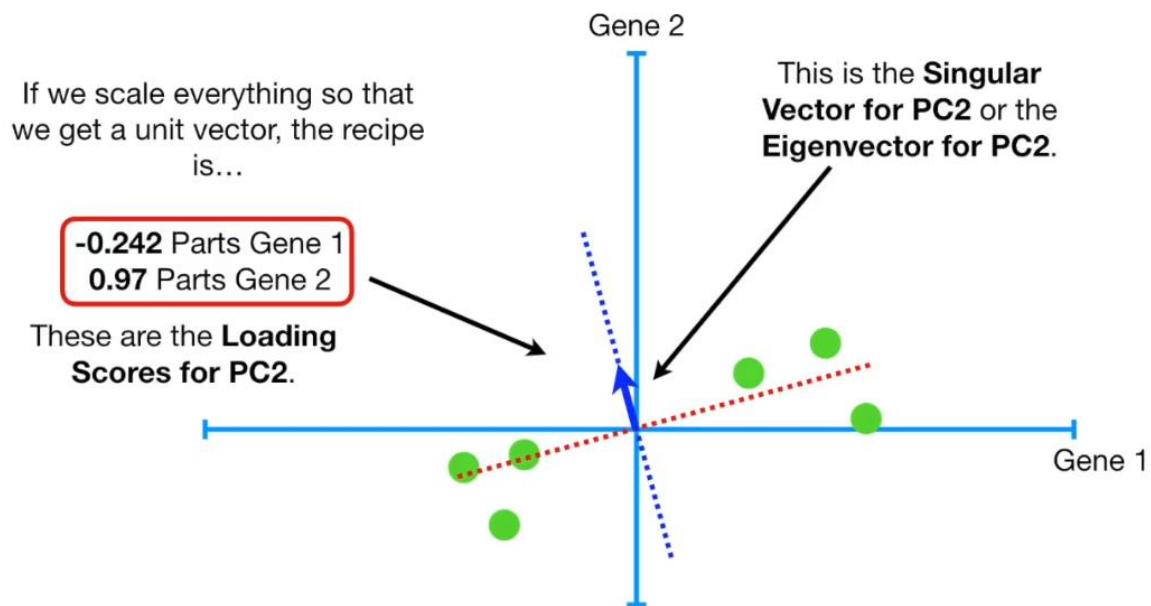
Fuente: (Starmer, 2018).



Fuente: (Starmer, 2018).

Como el lector seguramente sabrá, en álgebra lineal se expresa la independencia entre dos vectores a través de la nulidad de su producto escalar, lo cual se representa geoméricamente a nivel de dos dimensiones como la condición de

perpendicularidad entre dos vectores y a nivel de tres dimensiones como la condición de ortogonalidad entre dos vectores. Así, como la característica de no-correlación (mencionada al inicio de este documento) es precisamente la característica de independencia lineal entre variables (lo que busca cristalizar la intuición de que las variables no tienen relación siquiera al nivel más elemental), entonces la perpendicularidad entre los vectores que menciona Starmer expresa únicamente el segundo paso del proceso iterativo de la metodología numérica del PCA para ir encontrar una segunda recta de mejor ajuste, esta vez sujeta a la restricción de no estar correlacionada con la primera recta que se encontró; y así sucesivamente para todos los siguientes pasos.



Fuente: (Starmer, 2018).

$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2 = \text{sum of squared distances} = \text{SS}(\text{distances})$$

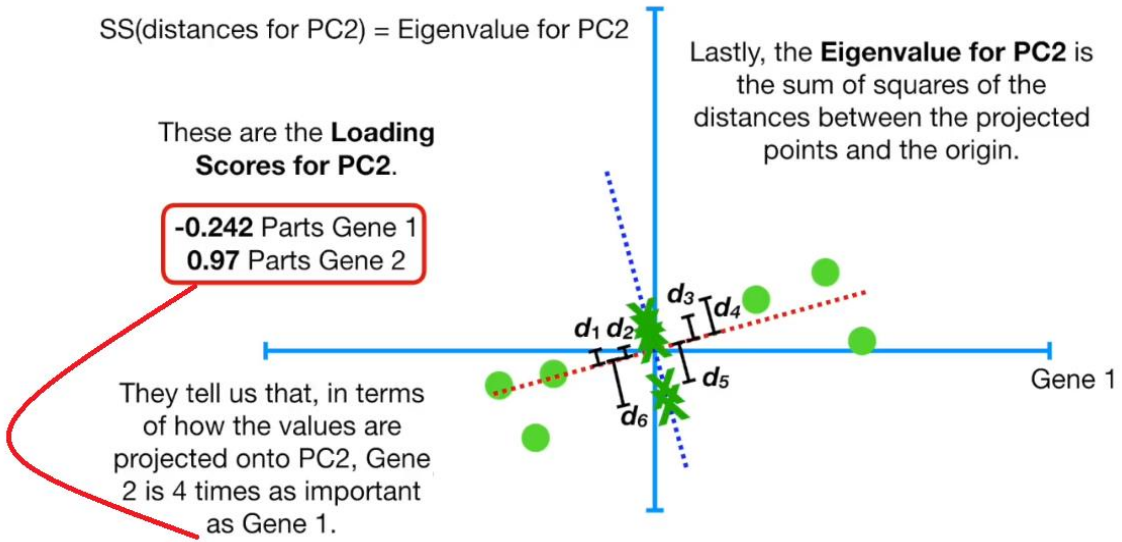
SS(distances for PC2) = Eigenvalue for PC2

Lastly, the **Eigenvalue for PC2** is the sum of squares of the distances between the projected points and the origin.

These are the **Loading Scores for PC2**.

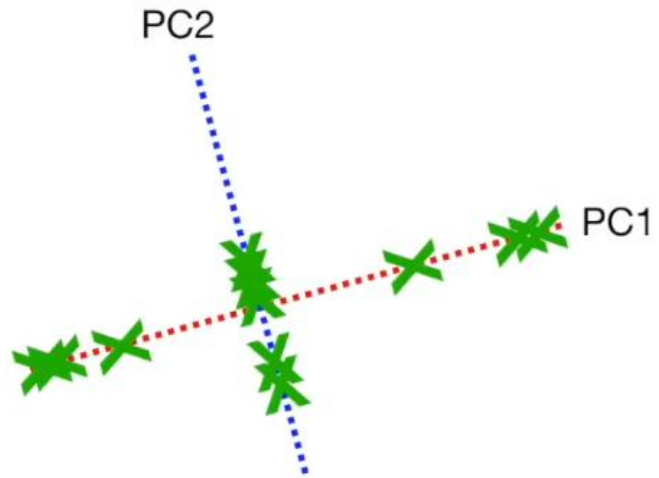
-0.242 Parts Gene 1
0.97 Parts Gene 2

They tell us that, in terms of how the values are projected onto PC2, Gene 2 is 4 times as important as Gene 1.

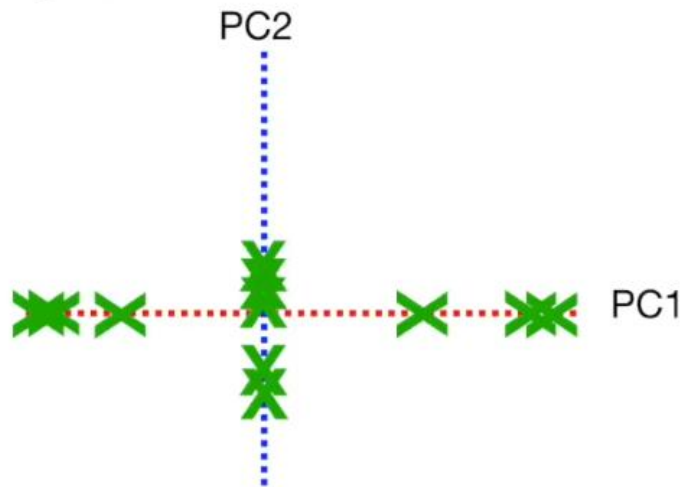


Fuente: (Starmer, 2018).

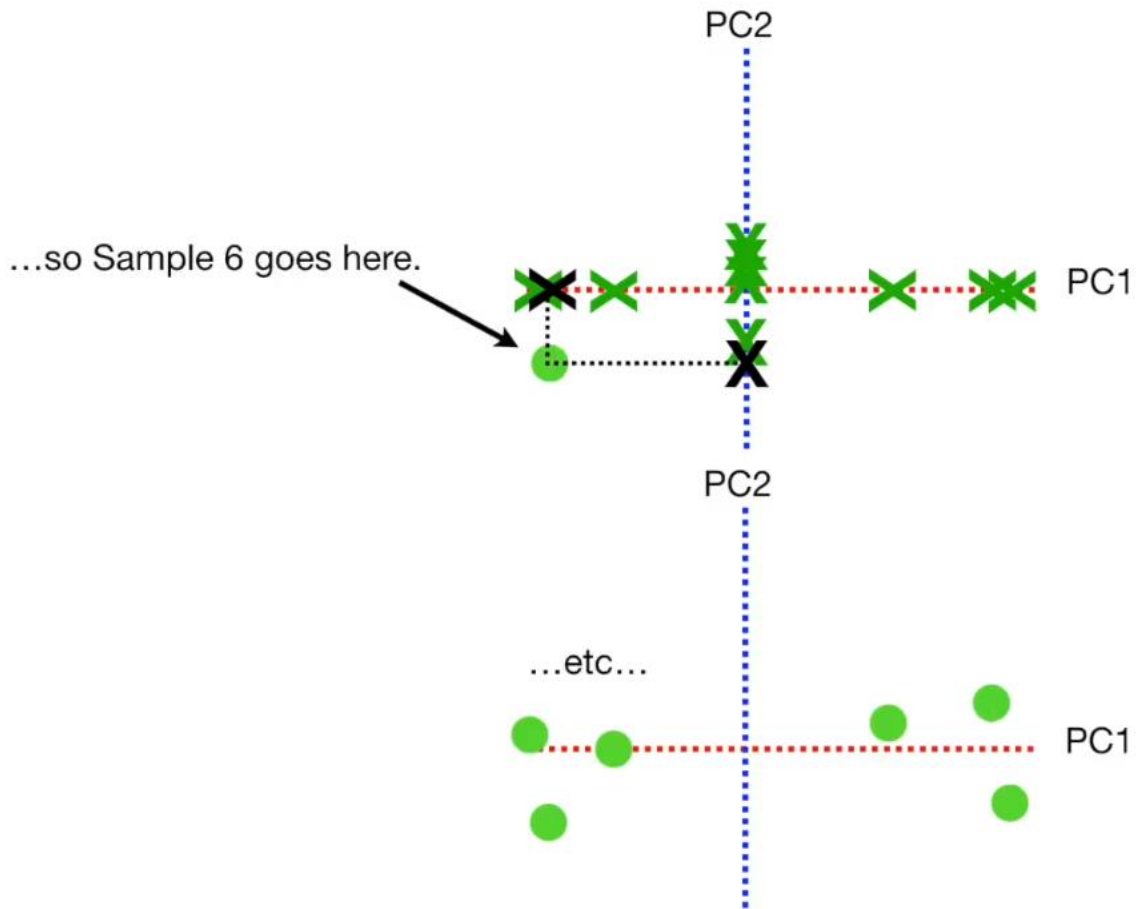
To draw the final PCA plot...



...then we use the projected points to find where the samples go in the PCA plot.



Fuente: (Starmer, 2018).



Fuente: (Starmer, 2018).

2. We can convert them into variation around the origin (0, 0) by dividing by the sample size minus 1 (i.e. $n - 1$).

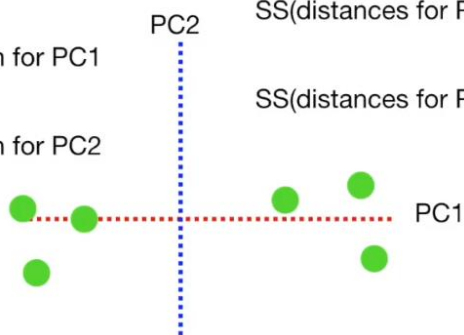
1. Remember the eigenvalues?

$$\frac{SS(\text{distances for PC1})}{n - 1} = \text{Variation for PC1}$$

$$\frac{SS(\text{distances for PC2})}{n - 1} = \text{Variation for PC2}$$

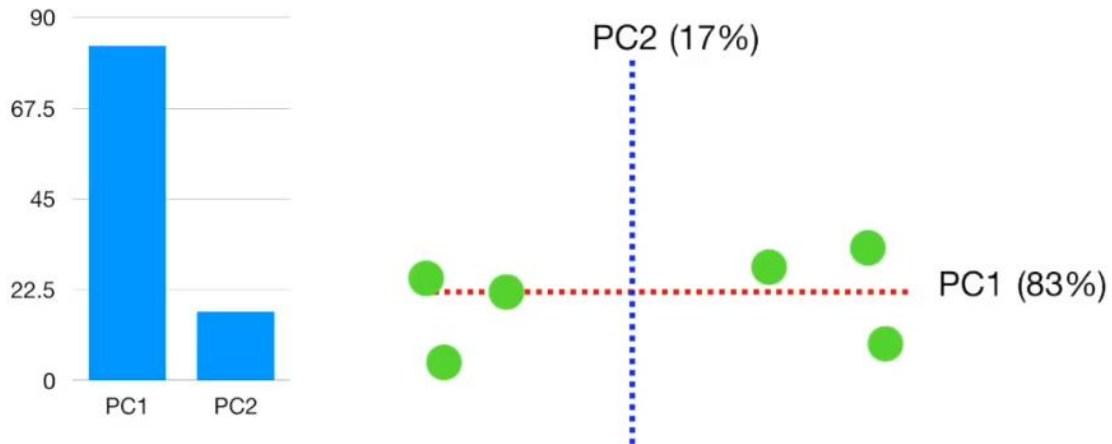
SS(distances for PC1) = Eigenvalue for PC1

SS(distances for PC2) = Eigenvalue for PC2



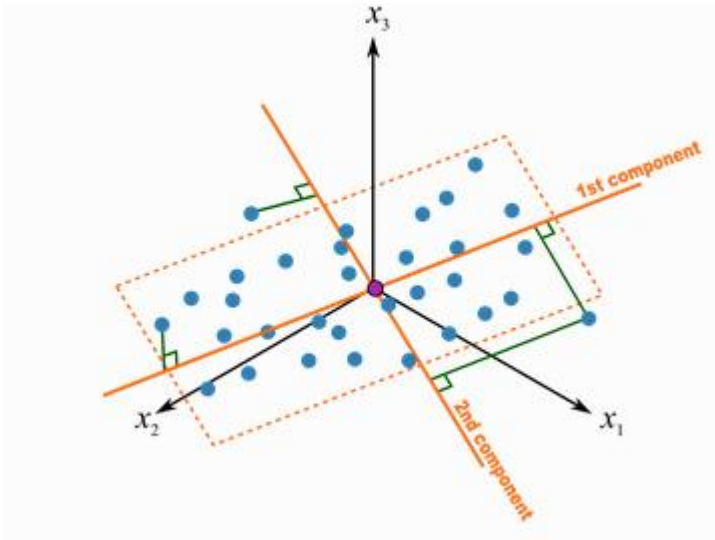
Fuente: (Starmer, 2018).

A **Scree Plot** is a graphical representation of the percentages of variation that each PC accounts for.



Fuente: (Starmer, 2018).

Puesto que los valores característicos fueron obtenidos proyectando las observaciones a cada una de las rectas de mejor ajuste encontradas (con la primera ronda de optimización sin restricciones y con la segunda sujeta a la condición de no-correlación con la primera recta de mejor ajuste), cuyos vectores directores (los que dan la dirección a la recta) son ni más ni menos que los vectores característicos, mientras que los valores característicos resumen para cada recta la varianza o “información perdida” tras el ajuste (mediante la proyección ortogonal) de cada una de las observaciones dentro del espacio muestral analizado. Las intuiciones geométricas anteriores se pueden sintetizar en la gráfica presentada a continuación.



Fuente: (Dunn, 2021, pág. 339).

El procedimiento anteriormente, conocido como *descomposición en valores singulares*, no debe confundirse con el procedimiento de *descomposición en eigenvalores* que, aunque guardan similitudes importantes (fundamentalmente en su planteamiento geométrico), también poseen diferencias fundamentales como lo es su estructura computacional misma, como se puede verificar en (Dunn, 2021, págs. 353-355)⁴.

⁴ Conceptualmente hablando, que es más importante, sus diferencias son sintetizadas en (Alger, 2013) como sigue:

- 1) Los vectores en la matriz de auto-descomposición P no son necesariamente ortogonales, por lo que el cambio de base no es una simple rotación. Por otro lado, los vectores en las matrices U y V en la SVD son ortonormales, por lo que representan rotaciones (y posiblemente volteos).
- 2) En la SVD , las matrices no diagonales U y V no son necesariamente inversas entre sí. Por lo general, no están relacionados entre sí en absoluto. En la descomposición propia, las matrices no diagonales P y P^{-1} son inversas entre sí.
- 3) En la SVD , las entradas en la matriz diagonal Σ son todas reales y no negativas. En la descomposición propia, las entradas de D puede ser cualquier número complejo: negativo, positivo, imaginario, lo que sea.
- 4) La SVD siempre existe para cualquier tipo de matriz rectangular o cuadrada, mientras que la descomposición propia solo puede existir para matrices cuadradas, e incluso entre matrices cuadradas a veces no existe.

III. INTEPRETACIÓN DE LOS RESULTADOS DEL PCA

III.I. Generalidades

El análisis de componentes principales puede realizarse con diversos programas estadísticos, los cuales comprenden R, Python, Minitab, SPSS, entre otros. A continuación, se presenta la salida del análisis de componentes principales arrojado por el software Minitab en su versión 2018, junto con su gráfica de sedimentación, con la finalidad de mostrar la manera en que deben interpretarse los resultados del PCA.

Análisis de componente principal: Ingresos, Educación, Edad, Residencia, ...

Análisis de los valores y vectores propios de la matriz de correlación

Valor propio	3.5476	2.1320	1.0447	0.5315	0.4112	0.1665	0.1254	0.0411
Proporción	0.443	0.266	0.131	0.066	0.051	0.021	0.016	0.005
Acumulada	0.443	0.710	0.841	0.907	0.958	0.979	0.995	1.000

Vectores propios

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Ingresos	0.314	0.145	-0.676	-0.347	-0.241	0.494	0.018	-0.030
Educación	0.237	0.444	-0.401	0.240	0.622	-0.357	0.103	0.057
Edad	0.484	-0.135	-0.004	-0.212	-0.175	-0.487	-0.657	-0.052
Residencia	0.466	-0.277	0.091	0.116	-0.035	-0.085	0.487	-0.662
Empleo	0.459	-0.304	0.122	-0.017	-0.014	-0.023	0.368	0.739
Ahorros	0.404	0.219	0.366	0.436	0.143	0.568	-0.348	-0.017
Deuda	-0.067	-0.585	-0.078	-0.281	0.681	0.245	-0.196	-0.075
Tarj Crédito	-0.123	-0.452	-0.468	0.703	-0.195	-0.022	-0.158	0.058

Fuente: (Minitab, 2019).

III.II. Interpretación de los Valores Característicos

Como se señala en la fuente referida, es posible utilizar la magnitud del valor característico para determinar el número de componentes principales. De esto se desprende el criterio de conservar los vectores característicos con los valores característicos más grandes (que son los factores de escala de estos vectores

característicos). Por ejemplo, según el criterio de Kaiser, se usan solo los componentes principales con valores propios que son mayores que 1.

II.III.I. Criterios para la selección de los valores característicos

Sobre tales criterios es necesario mencionar que, según (Universitat de Girona, 2002):

“Se han dado diversos criterios para determinar el número de factores a conservar. Uno de los más conocidos y utilizados es el criterio o regla de Kaiser (1960) que indicaría lo siguiente: "conservar solamente aquellos factores cuyos valores propios (eigenvalues) son mayores a la unidad". Este criterio es el que suelen utilizar los programas estadísticos por defecto. Sin embargo, este criterio es generalmente inadecuado tendiendo a sobreestimar el número de factores.

Otros criterios propuestos han sido, por ejemplo, el Scree-test de Cattell (1966) consistente en representar en un sistema de ejes los valores que toman los eigenvalues (ordenadas) y el número de factor (abscisas). Sobre la gráfica resultante se traza una línea recta base a la altura de los últimos autovalores (los más pequeños) y aquellos que queden por encima indicarán el número de factores a retener.

Velicer (1976) propone el método MAP (Minimum Average Partial), que implica calcular el promedio de las correlaciones parciales al cuadrado después de que cada uno de los componentes ha sido parcializado de las variables originales. Cuando el promedio de las correlaciones parciales al cuadrado alcanza un mínimo no se extraen más componentes. Este mínimo se alcanza cuando la matriz residual se acerca más a una matriz identidad. Un requisito para utilizar esta regla es que cada uno de los componentes retenidos deben tener al menos dos variables con pesos altos en ellos.

Bartlett (1950, 1951) propone una prueba estadística para contrastar la hipótesis nula de que los restantes $p-m$ autovalores son iguales (siendo p el número original

de variables y m el número de factores o componentes retenidos). Cada autovalor es excluido de manera secuencial hasta que no puede ser rechazada la hipótesis nula a través de una prueba de Ji-cuadrado.

El Análisis Paralelo fue sugerido por Horn (1965) quien señala que a nivel poblacional los autovalores de una matriz de correlaciones para variables no correlacionadas tomarían valor 1. Cuando se generan matrices muestrales basadas en esa matriz poblacional por fluctuaciones debidas al azar los autovalores excederán levemente de 1 y los últimos estarán ligeramente por debajo de 1. Horn propone contrastar los autovalores encontrados empíricamente en los datos reales con los obtenidos a partir de una matriz de variables no correlacionadas basada en el mismo número de variables que los datos empíricos y en el mismo tamaño de muestra. Los componentes empíricos con autovalores superiores a los de la matriz son retenidos.

El método de Razón de Verosimilitud, introducido por Lawley (1940), se trata de un criterio de bondad de ajuste pensado para la utilización del método de extracción de máxima verosimilitud, que se distribuye según Ji-cuadrado. La lógica de este procedimiento es comprobar si el número de factores extraído es suficiente para explicar los coeficientes de correlación observados.

De todos estos criterios los que parecen haber demostrado un mejor funcionamiento son el MAP y el Análisis Paralelo, sin embargo, tienen la desventaja de que no son muy accesibles en la práctica.”

III.III. Interpretación de los Componentes Principales

Los componentes principales son las combinaciones lineales de las variables originales que explican la varianza en los datos. El número máximo de componentes extraídos siempre es igual al número de variables (en caso de que la reducción de dimensionalidad no mejore la calidad del conjunto de datos). Los vectores característicos, compuestos por los coeficientes que corresponden a cada variable, se utilizan para calcular las puntuaciones de los componentes principales.

Los coeficientes indican la ponderación relativa de cada variable en el componente. *De lo anterior se desprende, puesto que se planteó antes que La k – ésima variable derivada del conjunto de datos original (i.e., $\alpha'_k x$) es el k – ésimo componente principal, la conclusión de que el componente principal es el producto escalar entre el espacio característico (que es la matriz conformada por los vectores característicos) multiplicada por el vector columna que contiene las variables originales (este vector columna es el que se transpuso y se denotó con el símbolo ' al inicio de este documento).*

Entonces todos los componentes y (en total p) se pueden expresar como el producto de una matriz formada por los autovectores, multiplicada por el vector x que contiene las variables originales x_1, \dots, x_p

$$y = Ax$$

donde

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$$

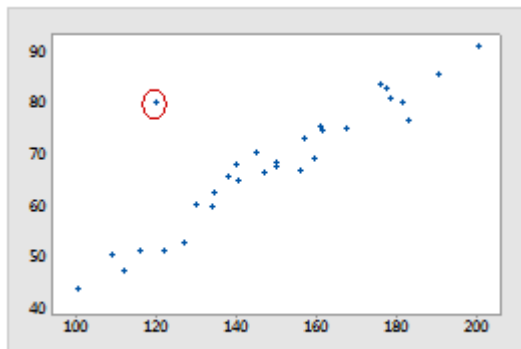
Fuente: (Universidad Carlos III de Madrid, 2006, pág. 10).

Por tanto, la interpretación que debe realizarse, específicamente de la tabla numérica generada por Minitab para el caso hipotético aquí presentado, es que el primer componente principal tiene asociaciones positivas grandes con “Edad”, “Residencia”, “Empleo” y “Ahorros”. Así, es posible interpretar este componente principalmente como una medición de la estabilidad financiera a largo plazo de un solicitante. El segundo componente tiene asociaciones negativas grandes con “Deudas” y “Tarj Crédito”, así que este componente mide principalmente el historial crediticio de un solicitante. El tercer componente tiene asociaciones negativas grandes con “Ingresos”, “Educación” y “Tarj crédito”, así que este componente mide principalmente las calificaciones académicas y de ingresos de un solicitante.

Además, en el análisis de las puntuaciones (los elementos al interior de los vectores característicos) debe recordarse que estas son combinaciones lineales de los datos que se determinan por los coeficientes de cada componente principal (estos coeficientes son los vectores columna x_i en el que se expresa cada atributo o variable analizada, como se explicó dos párrafos atrás). Para obtener la puntuación de una observación, se deben sustituir sus valores en la ecuación lineal del componente principal. Si se utiliza la matriz de correlación, se deben estandarizar las variables para obtener la puntuación correcta de los componentes cuando se usa la ecuación lineal.

Por supuesto, también existen diversos indicadores gráficos que deben considerarse en el estudio estadístico de datos desde el PCA.

III.IV. Distancia de Mahalanobis

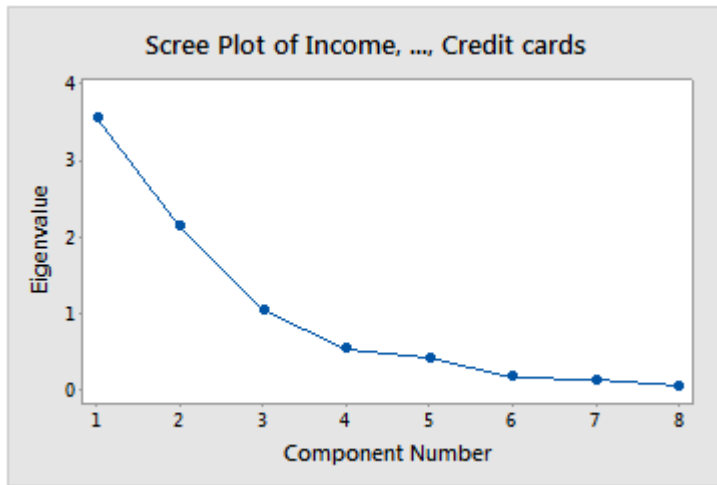


Fuente: (Minitab, 2019).

La distancia de Mahalanobis es la distancia entre un punto de los datos y el centroide de un espacio multivariado (la media general). La distancia de Mahalanobis es utilizada para identificar valores atípicos. Examinar la distancia de Mahalanobis es un método multivariado más potente para detectar valores atípicos que examinar una variable a la vez, porque la distancia toma en cuenta las diferentes escalas entre las variables y las correlaciones entre estas. Por ejemplo, en la gráfica anterior, al considerarse individualmente, ni el valor x ni el valor y del punto de datos encerrado en un círculo es poco usual. Sin embargo, el punto de

datos no se ajusta a la estructura de correlación de las dos variables. Por lo tanto, la distancia de Mahalanobis para este punto es inusualmente grande. Para evaluar si un valor de distancia es suficientemente grande para que la observación se considere un valor atípico, debe utilizarse la gráfica de valores atípicos.

III.V. Gráfica de Sedimentación (Scree Plot)



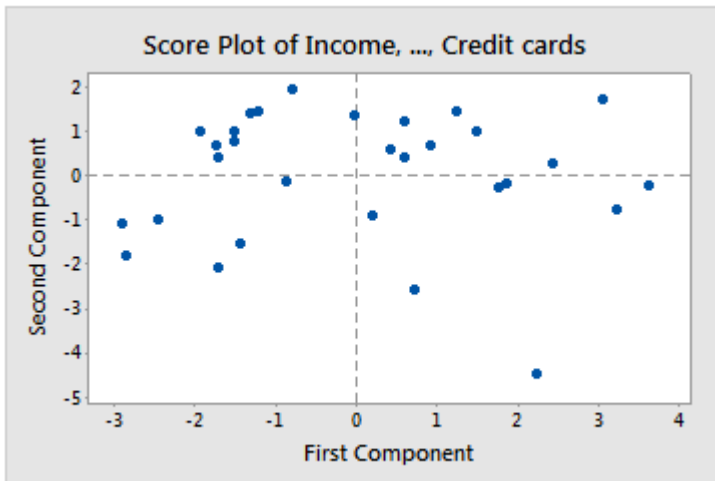
Fuente: (Minitab, 2019).

Respecto al gráfico de sedimentación presentado, señala (Minitab, 2019) que “(...) un gráfico de sedimentación muestra el número del componente principal versus su valor propio correspondiente. La gráfica de sedimentación ordena los valores propios desde el más grande hasta el más pequeño. Los valores propios de la matriz de correlación son iguales a las varianzas de los componentes principales.” Esta gráfica debe para seleccionar el número de componentes que se usarán con base en la magnitud (el “tamaño”) de los valores característicos. *El patrón ideal es una curva pronunciada, seguida de una inflexión y luego de una línea recta. Se deben utilizar los componentes en la curva pronunciada antes del primer punto que inicia la tendencia de línea.*

Así, en la gráfica de sedimentación anteriormente expuesta, se debe realizar la siguiente interpretación: “(...) los valores propios comienzan a formar una línea recta después del tercer componente principal. Por lo tanto, los componentes

principales restantes explican una proporción muy pequeña de la variabilidad (cercana a cero) y probablemente carezcan de importancia.” Además de la gráfica de sedimentación, el PCA brinda cuatro gráficas más.

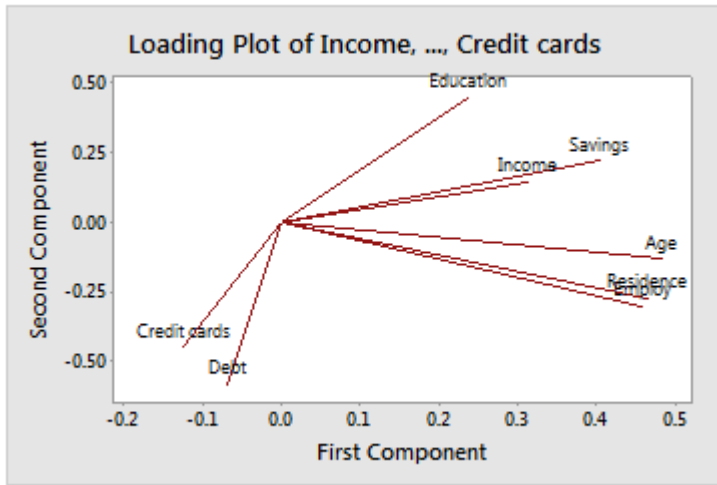
III.VI. Gráfica de Puntuaciones



Fuente: (Minitab, 2019).

La gráfica de puntuaciones representa visualmente las puntuaciones del segundo componente principal versus las puntuaciones del primer componente principal. Si los dos primeros componentes explican la mayor parte de la varianza en los datos, es válido utilizar la gráfica de puntuaciones para evaluar la estructura de los datos y detectar conglomerados, valores atípicos y tendencias. Las agrupaciones de datos en la gráfica pudieran indicar dos o más distribuciones separadas en los datos. Si los datos siguen una distribución normal y no hay valores atípicos presentes, los puntos están distribuidos aleatoriamente alrededor de cero. En la gráfica de puntuaciones aquí presentada, el punto en la esquina inferior podría ser un valor atípico, por lo que se debe investigar dicho punto.

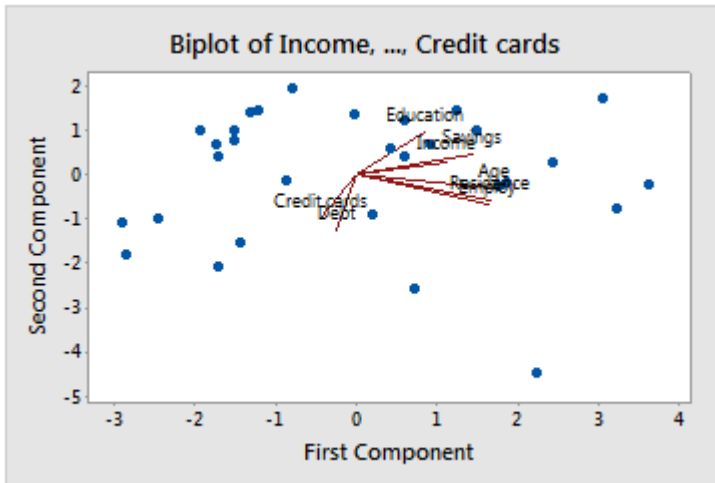
III.VII. Gráfica de Influencias



Fuente: (Minitab, 2019).

La gráfica de influencias grafica los coeficientes de cada variable para el primer componente versus los coeficientes para el segundo componente. La gráfica de influencias es utilizada para identificar cuáles variables tienen el mayor efecto en cada componente. Las influencias pueden ir de -1 a 1. Las influencias que se aproximan a -1 o 1 indican que la variable afecta considerablemente al componente. Las influencias cercanas a 0 indican que la variable tiene poca influencia en el componente. Evaluar las influencias también puede ayudar a caracterizar cada componente en términos de las variables.

III.VIII. Gráfica de Doble Proyección

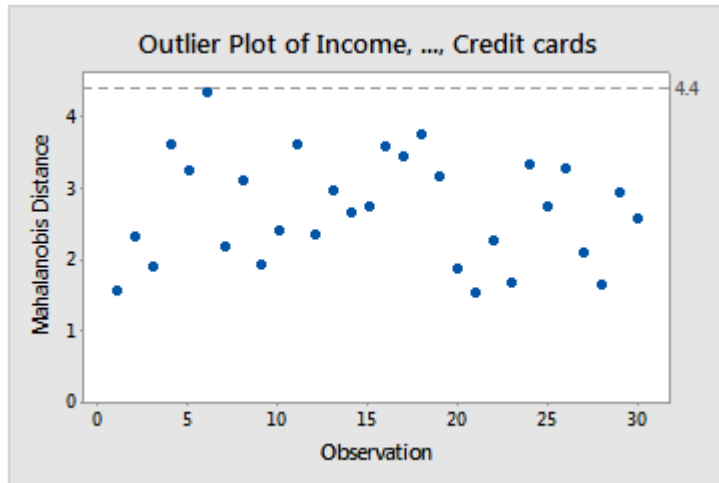


Fuente: (Minitab, 2019).

La gráfica de doble proyección sobrepone la gráfica de puntuaciones y la gráfica de influencias y es utilizada para evaluar la estructura de los datos y las influencias de los dos primeros componentes en una gráfica. Minitab grafica las puntuaciones del segundo componente principal versus las puntuaciones del primer componente principal, así como las influencias de ambos componentes. La gráfica de doble proyección aquí presentada muestra lo siguiente:

- “Edad”, “Residencia”, “Empleo” y “Ahorros” tienen influencias positivas grandes en el componente 1. Por lo tanto, este componente se centra en la estabilidad financiera a largo plazo de un solicitante.
- “Deudas” y “Tarj Crédito” tienen influencias negativas grandes en el componente 2. Por lo tanto, este componente se centra en el historial crediticio de un solicitante.
- El punto en la esquina inferior derecha podría ser un valor atípico. Se debe investigar este punto.

III.IX. Gráfica de Valores Atípicos



Fuente: (Minitab, 2019).

La gráfica de valores atípicos muestra la distancia de Mahalanobis para cada observación y una línea de referencia para identificar los valores atípicos. La distancia de Mahalanobis es la distancia entre cada punto de los datos y el centroide de un espacio multivariado (la media general). Examinar las distancias de Mahalanobis es un método más efectivo para detectar valores atípicos que examinar una variable a la vez, porque considera las diferentes escalas entre las variables y las correlaciones entre estas. La gráfica de valores atípicos es utilizada para identificar valores atípicos. Cualquier punto que se encuentre por encima de la línea de referencia es un valor atípico.

Los valores atípicos pueden afectar significativamente los resultados de su análisis. Por lo tanto, si se identifica un valor atípico en los datos, debe examinarse la observación para determinar por qué se trata de un valor poco común. Se debe corregir cualquier error de entrada de datos o de medición. También puede considerarse eliminar los datos que estén asociados a causas especiales y repetir el análisis. En los resultados aquí presentados no hay valores atípicos. Todos los puntos están por debajo de la línea de referencia.

III.X. Algunas Consideraciones Finales

Finalmente, conviene realizar algunas observaciones. Por un lado, debe especificarse que existen otros métodos para realizar PCA, como por ejemplo el algoritmo conocido como *Mínimos Cuadrados Parciales Iterativos No-Lineales* (NIPALS, por sus siglas en inglés); este método, según (Dunn, 2021, pág. 356), es "(...) un método secuencial para calcular los componentes principales. El cálculo puede finalizar antes, cuando el usuario considere que se han calculado suficientes componentes. La mayoría de los paquetes de computadora tienden a usar el algoritmo NIPALS, ya que tiene dos ventajas principales: maneja los datos faltantes y calcula los componentes secuencialmente. El propósito de considerar este algoritmo aquí es triple: brinda información adicional sobre lo que significan las cargas y las puntuaciones; muestra cómo cada componente es independiente (ortogonal a) los otros componentes, y muestra cómo el algoritmo puede manejar los datos faltantes. El algoritmo extrae cada componente secuencialmente, comenzando con el primer componente, la dirección de mayor varianza y luego el segundo componente, y así sucesivamente."

Por otro lado, el PCA parece tener alguna relación con la regresión lineal y, como se verá a continuación, esto no es mera apariencia. De lo expuesto por (Pearson, 1901, pág. 560), se verifica que el PCA puede plantearse matemáticamente de forma similar que planteamiento de una regresión lineal, específicamente que el PCA puede plantearse tomando a Y como variable dependiente en el sistema de puntos con coordenadas (X_n, Y_n) . Así se constata que su construcción matemática tiene una lógica muy similar, lo cual es así porque el PCA también se realiza mediante "la recta de mejor ajuste" (al igual que la regresión lineal), con la diferencia que el PCA asume que las distancias son ortogonales y la regresión lineal no adopta tal supuesto. Así, el PCA y la regresión lineal son metodologías estadísticas que usan rectas de mejor ajuste de diferentes tipos, por los supuestos implícitos respecto de las distancias. Desde una perspectiva restringida a la Ciencia de Datos y la Estadística, la diferencia fundamental entre regresión lineal y PCA es

que la regresión lineal busca la recta que minimice el error de la predicción respecto a una variable de respuesta, mientras que el PCA busca la recta que maximice la variación entre los puntos proyectados y, de esta forma, minimice la información pérdida.

IV. INTUICIÓN GEOMÉTRICA EN \mathbb{R}^3

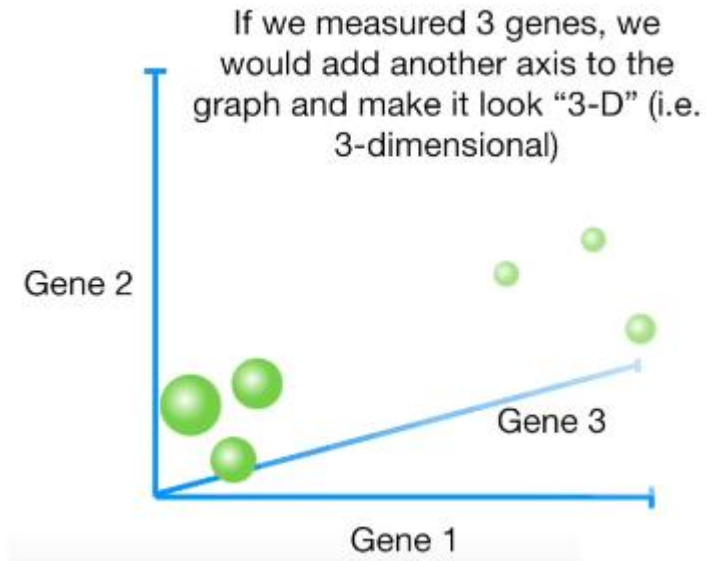
Los resultados obtenidos anteriormente pueden generalizarse al escenario de tres dimensiones, como se verá a continuación siguiendo con el ejemplo de (Starmer, 2018) basado en la transcripción genética suscitada en el estudio de los ratones. Para el caso tridimensional, se emplearán tres genes y se seguirán analizando seis ratones. El análisis que de ello debe hacerse es esencialmente el mismo

	Mouse 1	Mouse 2	Mouse 3	Mouse 4	Mouse 5	Mouse 6
Gene 1	10	11	8	3	2	1
Gene 2	6	4	5	3	2.8	1
Gene 3	12	9	10	2.5	1.3	2

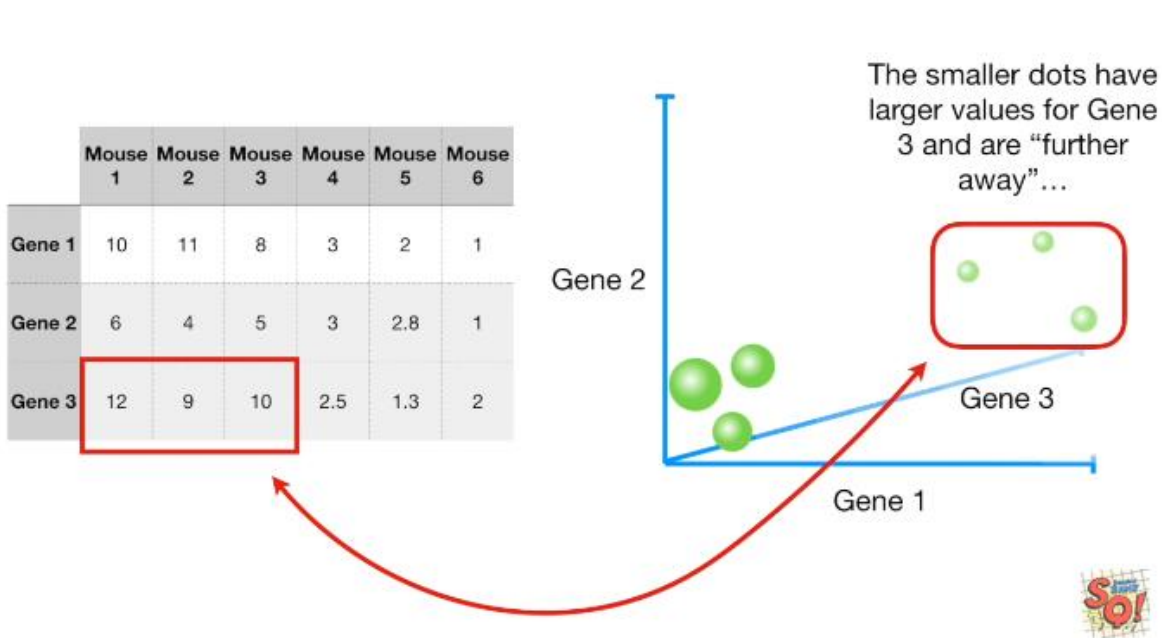
Fuente: (Starmer, 2018).

Como puede observarse, los ratones 1, 2 y 3 poseen valores de transcripción genética para el gen 3 que en un análisis preliminar indica que están más asociados

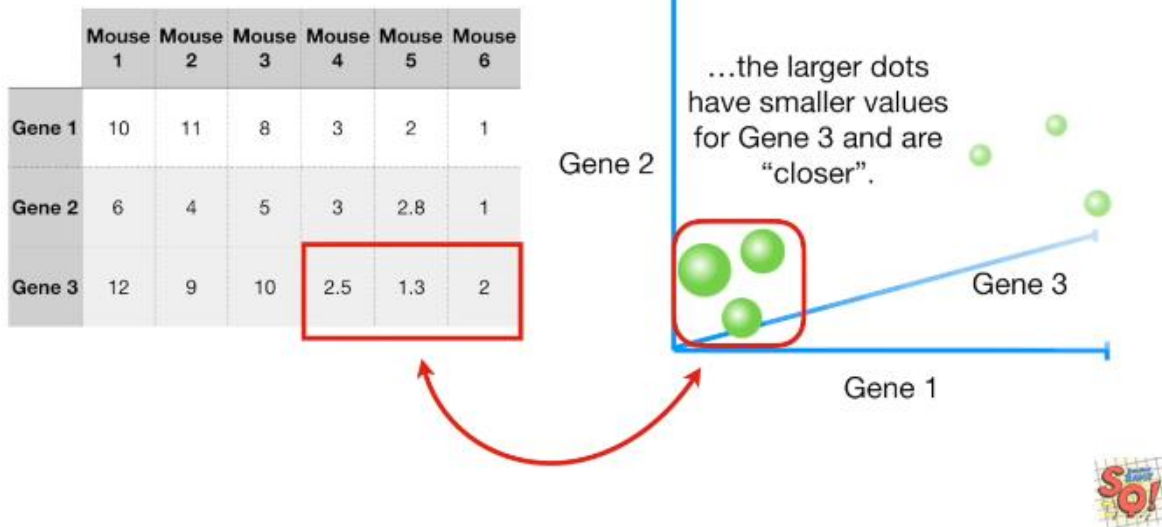
entre sí mismos que con los ratones 4, 5 y 6 (en relación al gen 3). Esto se aprecia geoméricamente como un agrupamiento en el eje cartesiano que representa al gen 3.



Fuente: (Starmer, 2018).



Fuente: (Starmer, 2018).

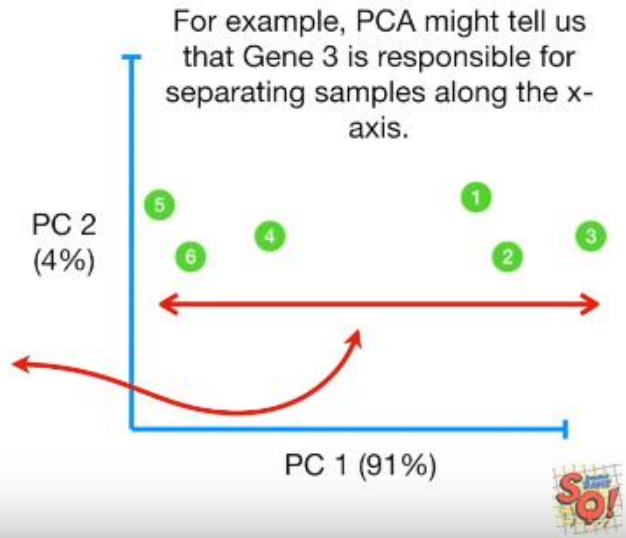


Fuente: (Starmar, 2018).

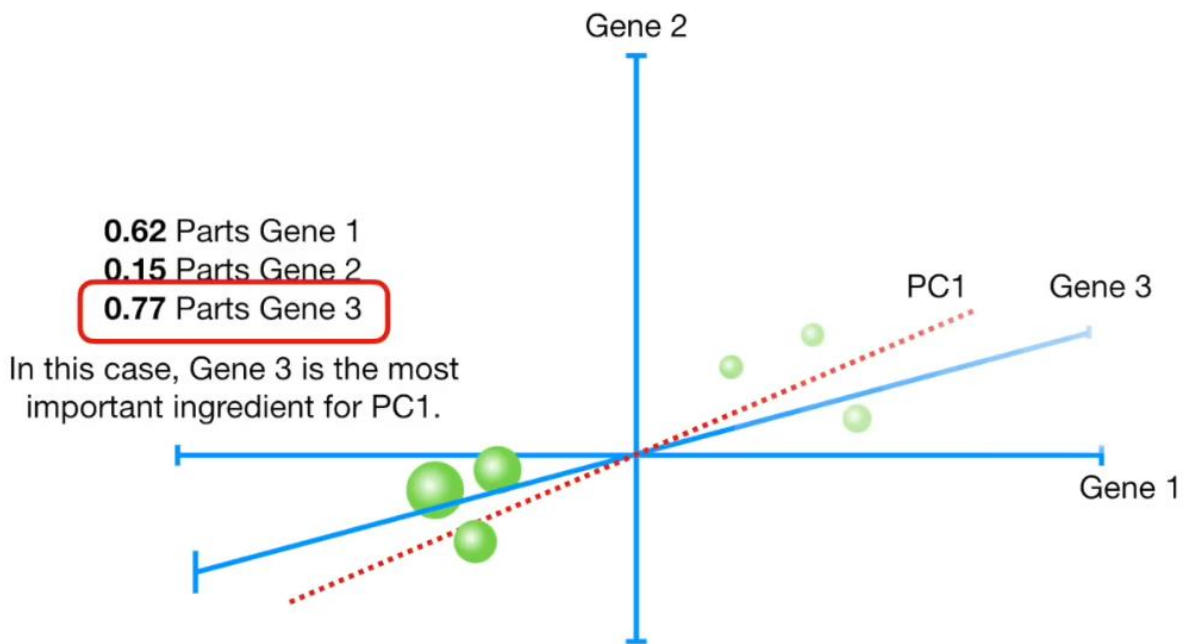
Así, lo que se busca es determinar cuál de las variables (en este caso los genes) que caracterizan a las unidades estadísticas de estudio (en este caso, los ratones) es la más valiosa para el agrupamiento (en inglés, *clustering*) de un determinado conjunto de datos. Por ejemplo, el PCA podría indicar que el gen 3 es el responsable de la existencia de tales agrupamientos dentro del espacio muestral (lo que se expresa en que se visualizan geoméricamente como "muestras separadas" dentro del plano en el que se están se observan).

Lo anterior es válido incluso en el escenario en que las variables sean más de tres y no puedan expresarse en un gráfico, tal como se muestra a continuación.

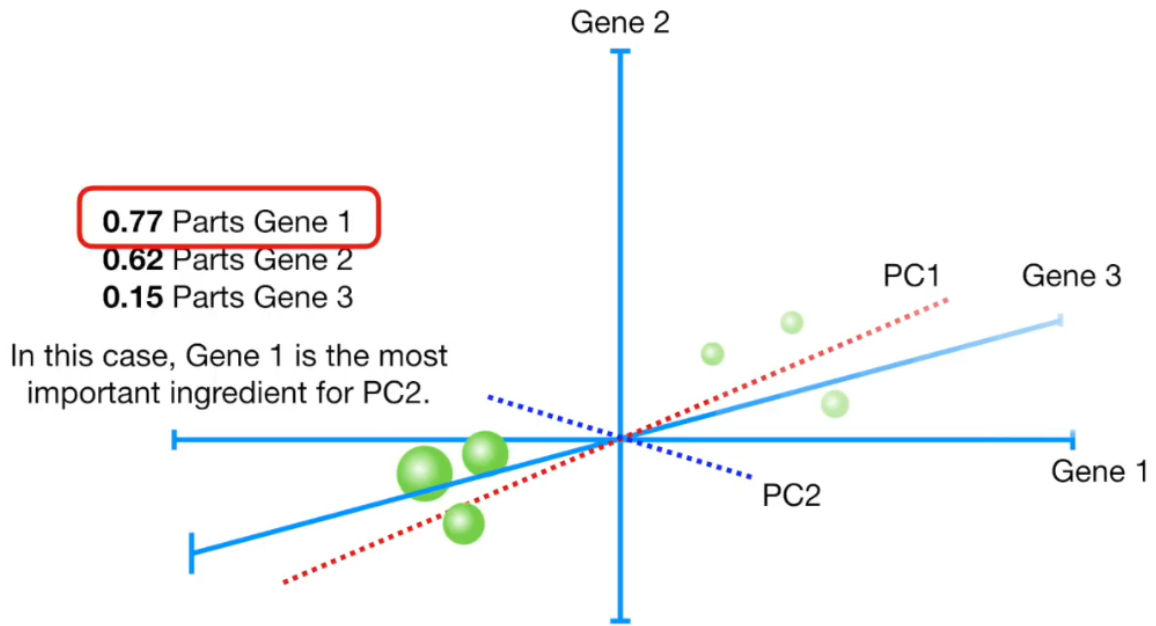
	Mouse 1	Mouse 2	Mouse 3	Mouse 4	Mouse 5	Mouse 6
Gene 1	10	11	8	3	2	1
Gene 2	6	4	5	3	2.8	1
Gene 3	12	9	10	2.5	1.3	2
Gene 4	5	7	6	2	4	7



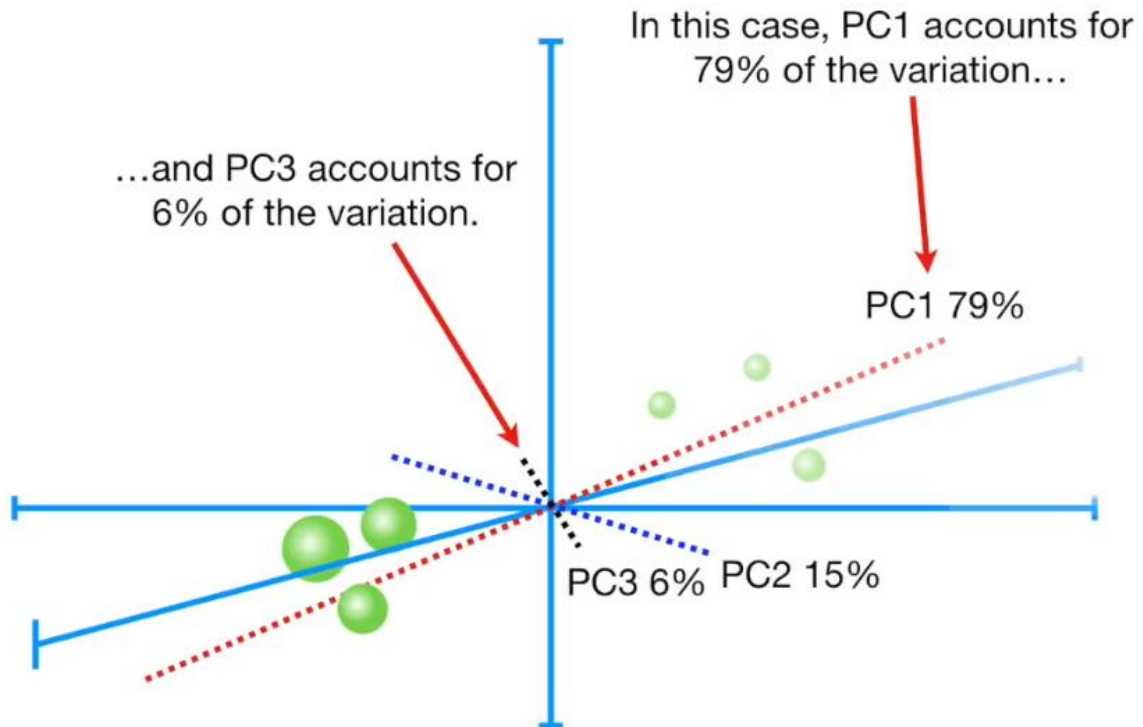
Fuente: (Starmer, 2018).



Fuente: (Starmer, 2018).



Fuente: (Starmer, 2018).



Fuente: (Starmer, 2018).

V. REFERENCIAS

- Adler, J. (2012). *R in a Nutshell* (Segunda ed.). Sebastopol, Crimea, Rusia: O'Reilly.
- Alger, N. (4 de Marzo de 2013). *Intuitively, what is the difference between Eigendecomposition and Singular Value Decomposition?* Obtenido de StackExchange Mathematics: <https://math.stackexchange.com/questions/320220/intuitively-what-is-the-difference-between-eigendecomposition-and-singular-valu>
- Bellman, R. (1972). *Dynamic Programming* (Sexta Impresión ed.). New Jersey: Princeton University Press.
- Dunn, K. G. (3 de Marzo de 2021). *Process Improvement Using Data*. Hamilton, Ontario, Canadá: Learning Chemical Engineering. Obtenido de 6.5. Principal Component Analysis (PCA) | 6. Latent Variable Modelling: <https://learnche.org/pid/PID.pdf?60da13>
- Jolliffe, I. (2002). *Principal Component Analysis*. New York: Springer-Verlag.
- Minitab. (18 de Abril de 2019). *Interpretar todos los estadísticos y gráficas para Análisis de componentes principales*. Obtenido de Soporte de Minitab 18: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/principal-components/interpret-the-results/all-statistics-and-graphs/>
- MIT. (23 de Febrero de 2021). *Linear transformations and their matrices*. Obtenido de Linear Algebra: https://ocw.mit.edu/courses/mathematics/18-06sc-linear-algebra-fall-2011/positive-definite-matrices-and-applications/linear-transformations-and-their-matrices/MIT18_06SCF11_Ses3.6sum.pdf
- Nabi, I. (2020). *Sobre los Estimadores de Bayes, el Análisis de Grupos y las Mixturas Gaussianas*. Documento inédito.
- Nabi, I. (3 de Abril de 2021). *¿Por qué se realiza un ajuste por re-escalamiento, normalización o estandarización sobre los datos en el contexto del aprendizaje automático?* Obtenido de El Blog de Isadore Nabi: <https://marxianstatistics.com/2021/04/03/por-que-se-realiza-un-ajuste-por-re-escalamiento-normalizacion-o-estandarizacion-sobre-los-datos-en-el-contexto-del-aprendizaje-automatico/>
- Nabi, I. (2 de Abril de 2021). *Una Interpretación Multidisciplinaria de los Espacios Característicos, Vectores Característicos y Valores Característicos*. Obtenido de El Blog de Isadore Nabi: <https://marxianstatistics.files.wordpress.com/2021/04/una-interpretacion-multidisciplinaria-de-los-espacios-caracteristicos-vectores-caracteristicos-y-valores-caracteristicos-isadore-nabi-1.pdf>
- Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, II(11), 559-572. Obtenido de <https://www.semanticscholar.org/paper/LIII.-On-lines->

and-planes-of-closest-fit-to-systems-
F.R.S./cac33f91e59f0a137b46176d74cee55c7010c3f8

Stack Exchange. (13 de Marzo de 2015). *Understanding proof of isometry implies isomorphism*.

Obtenido de Mathematics:

<https://math.stackexchange.com/questions/1188730/understanding-proof-of-isometry-implies-isomorphism/1188732>

Starmer, J. (2 de Abril de 2018). *Principal Component Analysis (PCA)*. Obtenido de

StatQuest: <https://www.youtube.com/watch?v=FgakZw6K1QQ>

Universidad Carlos III de Madrid. (7 de Noviembre de 2006). *Análisis de Componentes*

Principales. Obtenido de Proceso de extracción de factores:

<http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/AMult/tema3am.pdf>

Universitat de Girona. (24 de Enero de 2002). *Número de factores a conservar*. Obtenido de Análisis factorial:

[http://www3.udg.edu/dghha/cat/secciogeografia/prac/models/factorial\(5\).htm](http://www3.udg.edu/dghha/cat/secciogeografia/prac/models/factorial(5).htm)

Weisstein, E. (26 de Marzo de 2021). *Projection*. Obtenido de MathWorld - A Wolfram Web

Resource: <https://mathworld.wolfram.com/Projection.html>

Weisstein, E. (26 de Marzo de 2021). *Transformation*. Obtenido de MathWorld - A Wolfram

Web Resource: <https://mathworld.wolfram.com/Transformation.html>

Wikipedia. (4 de Noviembre de 2020). *Curse of dimensionality*. Obtenido de Numerical

Analysis: https://en.wikipedia.org/wiki/Curse_of_dimensionality

Wikipedia. (25 de Octubre de 2020). *Isomorfismo*. Obtenido de Álgebra:

<https://es.wikipedia.org/wiki/Isomorfismo>

Wikipedia. (26 de Marzo de 2021). *Isomorphism*. Obtenido de Equivalence (mathematics):

<https://en.wikipedia.org/wiki/Isomorphism>

Wikipedia. (22 de Marzo de 2021). *Transcripción genética*. Obtenido de Biosíntesis:

https://es.wikipedia.org/wiki/Transcripci%C3%B3n_gen%C3%A9tica