Métodos Cuantitativos para Ciencias Sociales y Negocios

Junghanss, Juan Cruz & Lopez Mondo, Ezequiel

Prof. Sergio Pernice

H.W. N°2

Comentarios sobre JN Nº2

Para la explicación de la notebook se empleó la función `np.random.seed(1)` para mantener los resultados iguales aunque se reinicie el código. De esta manera, las matrices y vectores representados contienen los mismos valores siempre.

Creamos dos vectores columna de dimensión 3x1 que, dada su aleatoriedad en los valores, son vectores linealmente independientes:

$$c_1 = \begin{bmatrix} -0.16 \\ 0.44 \\ 0.99 \end{bmatrix} \quad ; \quad c_2 = \begin{bmatrix} -0.39 \\ -0.70 \\ -0.81 \end{bmatrix} \quad ;$$

Para verificar su condición de linealidad, consideremos el coseno del ángulo entre vectores:

$$\cos(\theta) = \frac{c_1 \cdot c_2^T}{|c_1| \cdot |c_2|} = 0.4484$$

Al ser distinto de 1 o -1, vemos que son linealmente independientes.

Para obtener un tercer vector columna, utilizaremos escalares en el rango (-1,1) que generaran una combinación lineal con los vectores anteriores:

$$c_3 = d_1 \cdot c_1 + d_2 \cdot c_2 = d_1 \cdot \begin{bmatrix} -0.16 \\ 0.44 \\ 0.99 \end{bmatrix} + d_2 \cdot \begin{bmatrix} -0.39 \\ -0.70 \\ -0.81 \end{bmatrix} \Longrightarrow c_3 = \begin{bmatrix} 0.22 \\ -0.05 \\ 0.87 \end{bmatrix}$$

Debido a la definición de combinación lineal, entendemos que c_3 es linealmente dependiente de $c_1; c_2$

Podemos representar el sistema matricialmente, en este caso, con una matriz de columnas de tipo:

$$\boldsymbol{A} = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{bmatrix}$$

En nuestro caso, la matriz A sería:

$$A_{3x3} = \begin{bmatrix} c1_{11} & c2_{11} & c3_{11} \\ c1_{21} & c2_{21} & c3_{21} \\ c1_{31} & c2_{31} & c3_{31} \end{bmatrix} = \begin{bmatrix} -0.16 & -0.39 & 0.22 \\ 0.44 & -0.70 & -0.05 \\ 0.99 & -0.81 & 0.87 \end{bmatrix}$$

Una vez construida la matriz A podríamos, por ejemplo, calcular la cantidad de columnas linealmente independientes con la función `LA.matrix_rank(A)` que devuelve el rango de la matriz usando el método algebraico Descomposición de Valores Singulares (SVD en inglés). Para nuestro ejemplo, el rango es igual a 2, es decir, se comprueba nuevamente que tenemos dos columnas linealmente independientes, y no tres.

Hasta ahora, tenemos que c_1, c_2 son linealmente independientes, pero que c_3 es linealmente dependiente de $\{c_1, c_2\}$.

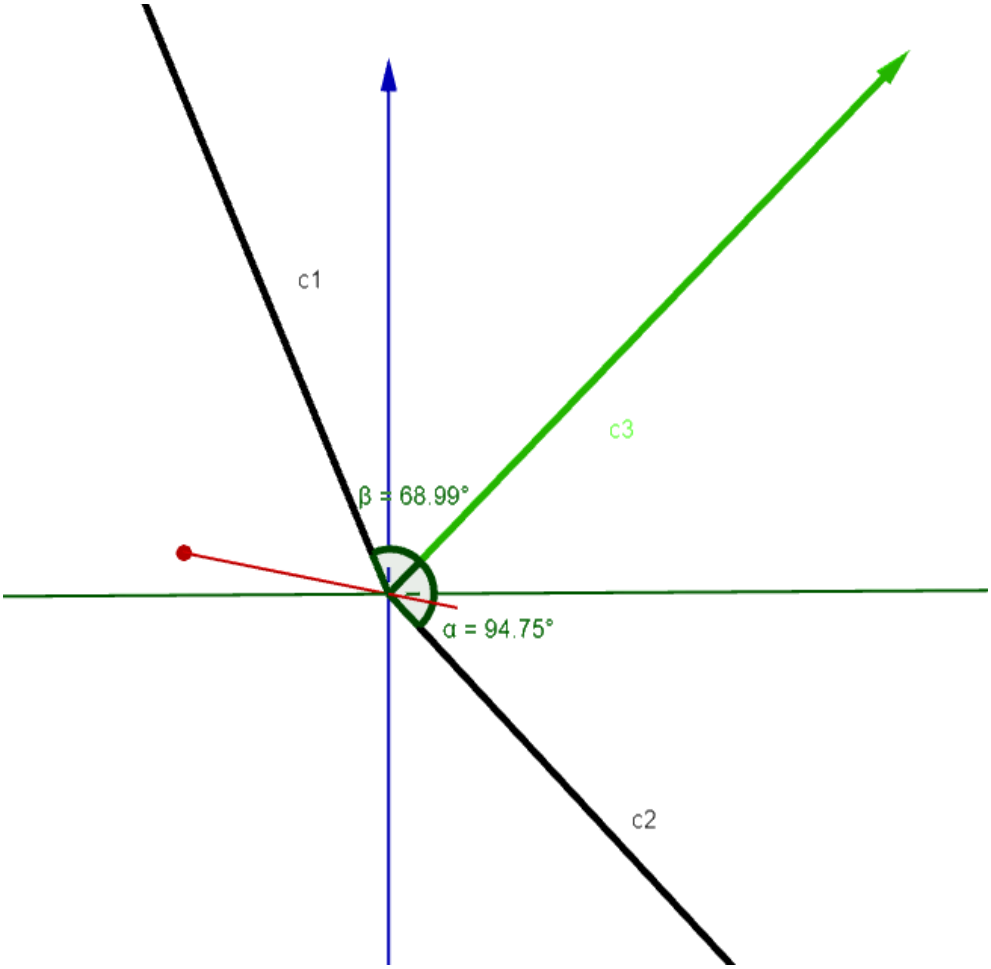
Sin embargo, esto último no significa que c_3 sea dependiente de c_1 y c_2 como vectores individuales, aunque este comprendido en el plano generado por c_1 y c_2. Resolviendo el coseno del ángulo entre cada par de vectores:

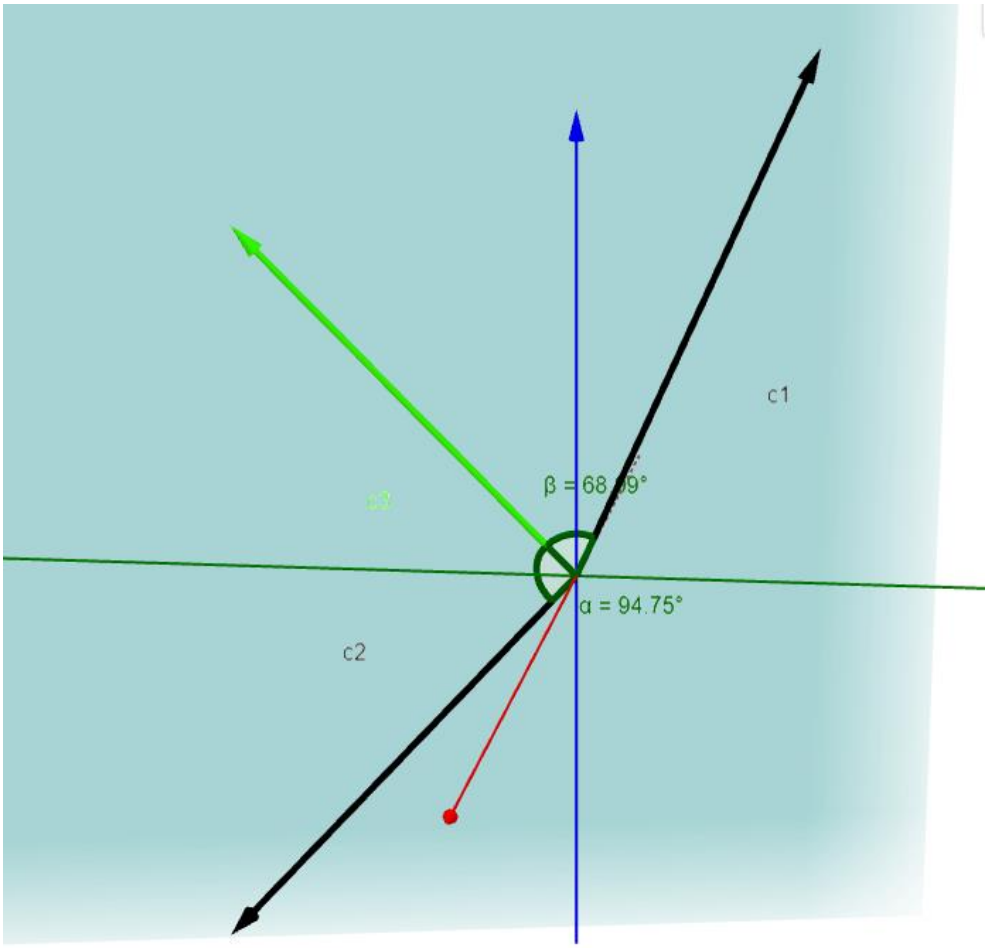
$$\cos(\theta) = \frac{c_1 \cdot c_2^T}{|c_1| \cdot |c_2|} = 0.4484$$

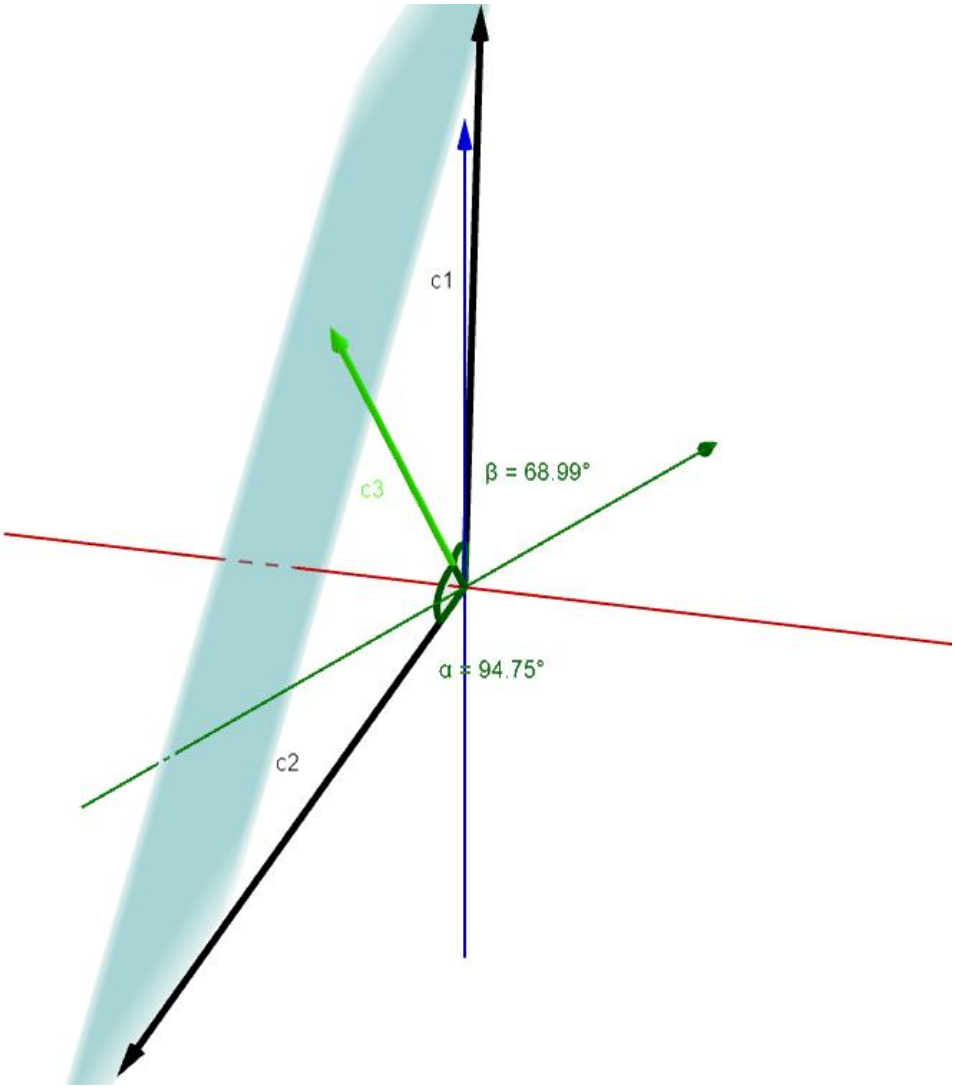
$$\cos(\theta) = \frac{c_1 \cdot c_3^T}{|c_1| \cdot |c_3|} = -0.9372$$

$$\cos(\theta) = \frac{c_2 \cdot c_3^T}{|c_2| \cdot |c_3|} = -0.7319$$

De todas maneras, el conjunto $\{c_1, c_2, c_3\}$ es linealmente dependiente, por lo que el plano que generan no es R^3




c1
β = 68.99°
c3
α = 94.75°
c2


c1
c3
β = 68.99°
α = 94.75°
c2

Por otra parte, de manera diferente, es posible determinar la condición de dependencia lineal en una matriz al calcular su determinante. Por medio de Numpy, `np.linalg.det(A)` nos devolverá el resultado del determinante. En nuestro ejemplo: $-0.00144 \approx 0$

Por último, podemos intentar también obtener, mediante el proceso de Gram-Schmidt, una base ortonormal del subespacio generado por $\{c_1, c_2, c_3\}$:

Para el vector c_1, su versor asociado será este mismo dividido por su norma:

$$n_1 = \frac{c_1}{|c_1|} = (-0.15\,,0.39\,,-0.90)$$

Para el vector c_2, un vector ortonormal podremos obtenerlo así:

$$n_2 = c_2 - Proyec_{n1}(c_2) = c_2 - \left(\frac{n_1 \cdot c_2^T}{n_1 \cdot n_1^T}\right) \cdot c_2$$

$$n_2 = (-0.39\,,-0.70\,,-0.81) - (-0.07\,,0.20\,,-0.46)$$
$$n_2 = (-0.31\,,-0.91\,,-0.34)$$

Finalmente, para el vector c_3:

$$n_3 = c_3 - Proyec_{n1}(c_3) - Proyec_{n2}(c_3) = c_3 - \left(\frac{n_1 \cdot c_3^T}{n_1 \cdot n_1^T}\right) \cdot c_3 - \left(\frac{n_2 \cdot c_3^T}{n_2 \cdot n_2^T}\right) \cdot c_3$$

$$n_3 = (0.22\,,-0.05\,,0.87) - (0.12\,,-0.33\,,0.77) - (0.09\,,0.28\,,0.10)$$
$$n_3 = (0\,,0\,,0)$$

A partir de este último resultado, podemos observar que al extraer la componente de c_3 que no está en el subespacio spanned por $\{n_1, n_2\}$ nos devuelve cero. Esto implica que c_3 es linealmente dependiente de c_1, c_2.

Esta última demostración nos lleva a comprender que para el sistema lineal $A_{3x3} \cdot X_{3x1} = b_{3x1}$ solo habrá una solución si y solo si b pertenece al mismo subespacio spanned por las columnas de A_{3x3}

Si generamos al azar un vector b, por probabilidad presuponemos que no tendrá solución. Por ejemplo:

$$b = \begin{bmatrix} 0.38 \\ 0.75 \\ 0.78 \end{bmatrix}$$

$$A_{3x3} \cdot X_{3x1} = b_{3x1} \Longrightarrow \begin{bmatrix} -0.16 & -0.39 & 0.22 \\ 0.44 & -0.70 & -0.05 \\ 0.99 & -0.81 & 0.87 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.38 \\ 0.75 \\ 0.78 \end{bmatrix}$$

Sabemos que la solución será: $A \cdot X = b \quad \Longrightarrow \quad A^{-1} \cdot A \cdot X = A^{-1} \cdot b \quad \Longrightarrow \quad I \cdot X = A^{-1} \cdot b$

$$X = A^{-1} \cdot b$$

Sin embargo, recordemos que $\det(A) = 0$, por lo que la matriz A no es invertible. Aunque esto lo sabemos de antemano, nos encontramos que si aplicamos la función de Numpy para resolver el sistema lineal `x = LA.solve(A, b)` obtendremos como resultado un vector de valores que tienden a infinito:

$$[-6.87729007e+14]$$
$$[-3.38536493e+14]$$
$$[-1.09601815e+15]$$

Esto sucede debido a que los elementos de las columnas de A no son literalmente cero y la función (que usa descomposición LU) continua sin arrojar un error, por lo que en lo que respecta al algebra lineal numérica de punto flotante (floating point) se termina encontrando una solución errónea. Incluso una pequeña perturbación en el vector b puede producir grandes errores en el vector solución X. Recordemos que los números de punto flotante son solo aproximaciones a números reales.

Como alternativa para detectar matrices no invertibles, se suele utilizar la función `np.linalg.cond(A)`, es decir, el número de condición de X que se define como la norma de X multiplicada por la norma de la inversa de X, que si devuelve un número mayor a ~1e15, sabremos que la matriz es esencialmente singular y no podemos usar descomposición LU, sino pseudoinversiones o descomposición de valores singulares. En nuestro caso el resultado de `np.linalg.cond(A)` es 2.468512216589437e+16

En síntesis, para fines computacionales, no hay una diferencia significativa entre una matriz que no es invertible (como nuestra matriz A, donde el número de condición es infinito) y una donde el número de condición es simplemente grande. Por este motivo obtuvimos una solución enorme. Al resolver un sistema de ecuaciones lineales, es importante observar el número de condición para estimar la precisión de la solución.

Referencia: página 187 (206 de 709) de Robert Johannson.

Naturalmente, cuando queremos verificar $A^T \cdot X - b = 0$ vemos que es distinto de cero el resultado:

$$(0.20\ ,0.29\ ,0.16)$$

Ahora bien, si generáramos un vector b que habita el subespacio generado por las columnas de A, al generar una combinación lineal de los vectores ortonormales a c_1, c_2 encontraríamos una solución:

$$b = d_1 \cdot n_1 + d_2 \cdot n_2$$

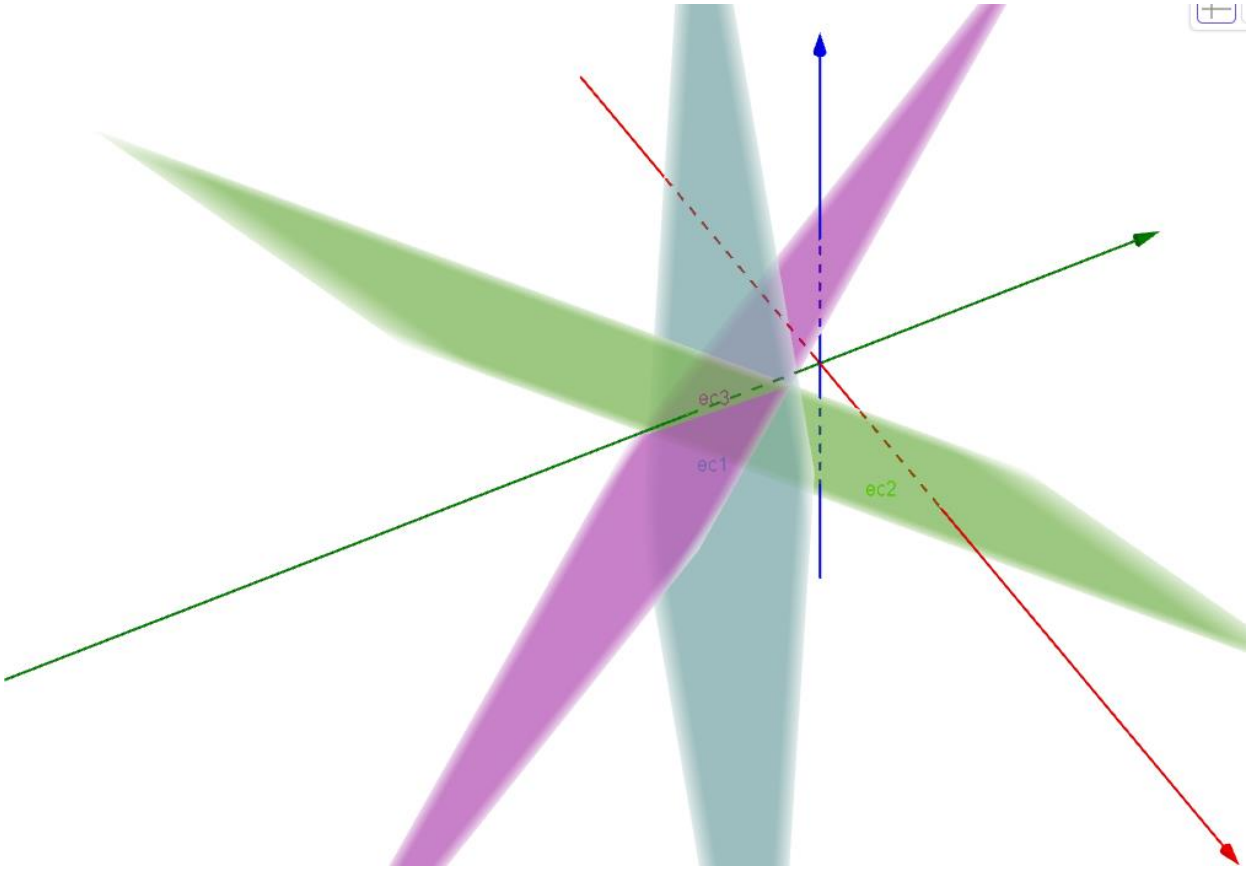
```
x = LA.solve(A, b)
```

$$X = (4.52\ ,1.35\ ,6.24)$$

Por último, podemos mostrar gráficamente cómo se ven los sistemas lineales con una solución única, una múltiple (infinitas soluciones) y sin solución. Recordemos antes las siguientes definiciones de los sistemas lineales de ecuaciones, que pueden servir de guía:

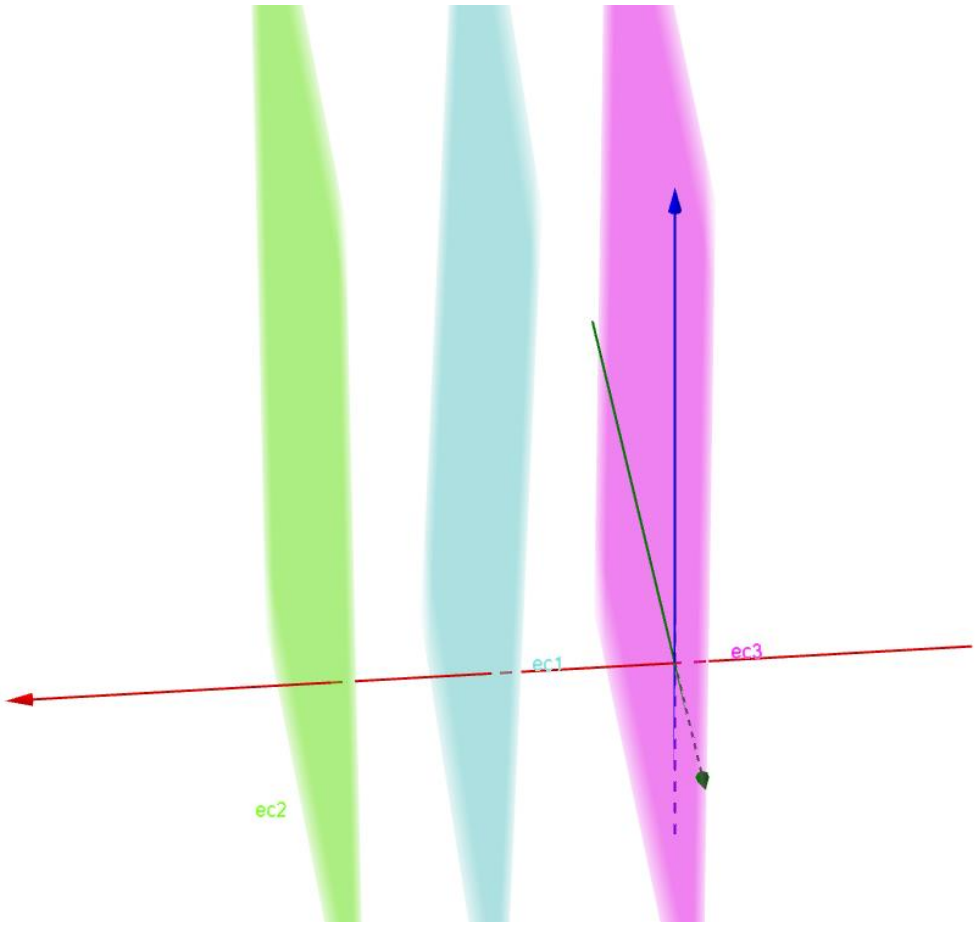
- Sistema Compatible Determinado: cuando posee una única solución.
- Sistema Compatible Indeterminado: cuando posee infinitas soluciones.
- Sistema Incompatible: cuando no posee solución.

Comencemos por el último ejemplo, donde creando un vector b que habita en el subespacio de las columnas de A, obtuvimos una solución.



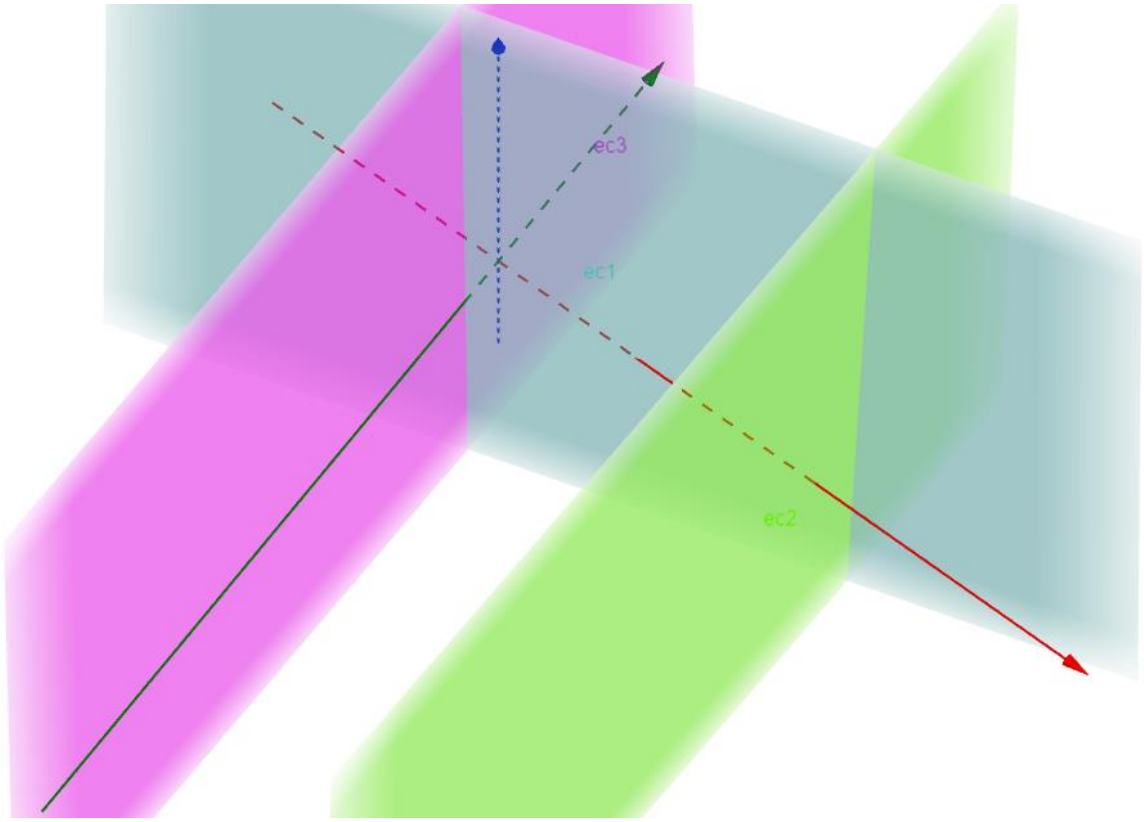
Sistema Compatible Determinado: podemos observar que los planos se intersecan.

Continuemos con un ejemplo para donde no hay una solución:



Sistema Incompatible: podemos observar que los planos no se intersecan en ningún punto.

Por último, vislumbremos un ejemplo donde hay múltiples soluciones:



Sistema Compatible Indeterminado: podemos observar que hay multiples soluciones.

Reinterpretación geométrica de Mínimos Cuadrados Ordinarios

Para abordar la reinterpretación de una regresión lineal múltiple realizada por mínimos cuadrados ordinarios se seleccionó una que corresponde al análisis del impacto de educación sobre el salario de un grupo de hombres en 1976 en EEUU a partir del siguiente modelo teórico:

$$\log(salario) = \beta_0 + \beta_1(educ) + \beta_2(exper) + \beta_3(exper2) + \beta_4(raza) + \beta_5(capital) + \beta_6(sur) + u$$

donde EDUC es la variable de años de educación, EXPER es la experiencia laboral y EXPER2 es la misma variable de experiencia laboral elevada al cuadrado para evaluar una tendencia cuadrática. RAZA es una variable binaria para raza negra, CAPITAL una variable binaria por vivir en un área metropolitana y SUR una variable binaria por vivir en el sur.

La cantidad de datos es de 3010, por lo que para nuestra reinterpretación geométrica tendremos una dimensión $\mathbb{R}^n$; donde $n = 3010$.

Los resultados de la salida de regresión fueron los siguientes:

```
reg l_salario educ exper raza capital sur
```

Source	SS	df	MS		
Model	165.205667	5	33.0411334	Number of obs	= 3010
Residual	427.435978	3004	.142288941	F(5, 3004)	= 232.21
				Prob > F	= 0.0000
Total	592.641645	3009	.196956346	R-squared	= 0.2788
				Adj R-squared	= 0.2776
				Root MSE	= .37721

l_salario	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.073807	.0035336	20.89	0.000	.0668784	.0807356
exper	.0393134	.0021955	17.91	0.000	.0350085	.0436183
raza	-.1882225	.0177678	-10.59	0.000	-.2230607	-.1533843
capital	.1647411	.0156919	10.50	0.000	.1339732	.195509
sur	-.1290528	.0152285	-8.47	0.000	-.1589122	-.0991935
_cons	4.913331	.0631212	77.84	0.000	4.789566	5.037096

- La ordenada al origen o intercepto resultó ser $\beta_0 = 4.913$
- El coeficiente $\beta_1 = 0.073$ para la variable educación, que implica un efecto marginal promedio sobre el logaritmo del salario de 7.4% por cada año de educación, manteniendo el resto de variables constantes.
- El coeficiente $\beta_2 = 0.039$ para la variable experiencia, que implica una variación promedio sobre la semielasticidad del salario de 4% por cada año adicional de experiencia laboral, ceteris paribus.
- El coeficiente $\beta_3 = -0.188$ para la variable cualitativa de raza, es decir, que para raza=1 verían una variación promedio salarial de -18%, manteniendo el resto de variables constantes respecto a quienes posean raza=0.
- El coeficiente $\beta_4 = 0.164$ para la variable ficticia de capital, que implica una variación promedio salarial de 16%, ceteris paribus, para quienes vivan en la capital frente a los que no.
- Por último, el coeficiente $\beta_5 = -0.129$ para la variable dummy de sur, es decir, que para quienes vivan en el sur verán un efecto marginal promedio en su salario de -13%, ceteris paribus, respecto a los que no vivan allí.

Considerando una visión geométrica del mismo problema, debemos entender que los coeficientes β_i son aquellos que minimizan el error cuadrático promedio, pero en realidad esta minimización implica la proyección del vector $Y_{salario}$ sobre el subespacio spanned por los vectores X_i, donde $i = 5\ variables\ exógenas$.

Nuestro modelo representado geométricamente implicaría vectores columna de (3010x1):

$$Y_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}} = \beta_0 \cdot \begin{bmatrix}\bar{1}\\\vdots\\\bar{1}\end{bmatrix} + \beta_1 \cdot X_1{}_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}} + \beta_2 \cdot X_2{}_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}} + \beta_3 \cdot X_3{}_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}} + \beta_4 \cdot X_4{}_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}} + \beta_5 \cdot X_5{}_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}} + u_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}}$$

Donde el error cometido en la predicción es:

$$u_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}} = Y_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}} - \hat{Y}_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}} = Y_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}} - \left[\beta_0 \cdot \begin{bmatrix}\bar{1}\\\vdots\\\bar{1}\end{bmatrix} + \beta_1 \cdot X_1{}_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}} + \beta_2 \cdot X_2{}_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}} + \beta_3 \cdot X_3{}_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}} + \beta_4 \cdot X_4{}_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}} + \beta_5 \cdot X_5{}_{\begin{bmatrix}1\\\vdots\\3010\end{bmatrix}}\right]$$

Expresando matricialmente el modelo obtenemos lo siguiente:

$$\begin{bmatrix}Y_1\\Y_2\\\vdots\\\vdots\\Y_{3010}\end{bmatrix}_{3010\times1} = \begin{bmatrix}1 & X_{1.1} & X_{2.1} & X_{3.1} & X_{4.1} & X_{5.1}\\1 & X_{1.2} & X_{2.2} & X_{3.1} & X_{4.2} & X_{5.2}\\\vdots & \vdots & \vdots & \dots & \vdots & \vdots\\\vdots & \vdots & \vdots & \dots & \vdots & \vdots\\1 & X_{1.3010} & X_{2.3010} & X_{3.3010} & X_{4.3010} & X_{5.3010}\end{bmatrix}_{3010x6} * \begin{bmatrix}\beta_0\\\beta_1\\\vdots\\\vdots\\\beta_5\end{bmatrix}_{6\times1} + \begin{bmatrix}u_1\\u_2\\\vdots\\\vdots\\u_{3010}\end{bmatrix}_{3010\times1}$$

Con:

$$Y = \begin{bmatrix}Y_1\\Y_2\\\vdots\\\vdots\\Y_{3010}\end{bmatrix} \quad X = \begin{bmatrix}1 & X_{1.1} & X_{2.1} & X_{3.1} & X_{4.1} & X_{5.1}\\1 & X_{1.2} & X_{2.2} & X_{3.1} & X_{4.2} & X_{5.2}\\\vdots & \vdots & \vdots & \dots & \vdots & \vdots\\\vdots & \vdots & \vdots & \dots & \vdots & \vdots\\1 & X_{1.3010} & X_{2.3010} & X_{3.3010} & X_{4.3010} & X_{5.3010}\end{bmatrix} \quad \beta = \begin{bmatrix}\beta_0\\\beta_1\\\vdots\\\vdots\\\beta_5\end{bmatrix} \quad u = \begin{bmatrix}u_1\\u_2\\\vdots\\\vdots\\u_{3010}\end{bmatrix}$$

Re-expresando:

$$Y = X * \beta + u$$

Donde β es el vector de los coeficientes, y u es la diferencia entre el valor predicho y el valor observado de Y.
El objetivo de la regresión lineal es minimizar la suma de los residuales al cuadrado:

$$\text{argmin } u^2 = \text{arg}min(Y - X\beta)'(Y - X\beta)$$

Diferenciando con respecto a β y resolviendo:

$$\begin{aligned}\frac{d}{d\beta}(Y - X\beta)'(Y - X\beta) &= -2X(Y - X\beta)\\ &= 2X'X\beta - 2X'Y \Rightarrow 0\\ X'X\hat{\beta} &= X'Y\\ (X'X)^{-1}X'X\hat{\beta} &= (X'X)^{-1}X'Y\\ \hat{\beta} &= (X'X)^{-1}X'Y\end{aligned}$$

Para obtener nuestra predicción de Y, es decir $\hat{Y}$, simplemente hay que multiplicar la matriz de los coeficientes por la matriz de las observaciones X:

$$\hat{Y} = X\left(X'X\right)^{-1}X'Y$$

Notemos como la derivación de $\hat{Y}$ es muy similar a la forma genérica de una matriz de proyección de forma:

$$P_v = v(v'v)^{-1}v'$$

Ambas matrices solo difieren en que $\hat{Y}$ incluye al vector de valores observados Y.

Entonces tenemos que:

$$\hat{Y} = X(X'X)^{-1}X'Y$$

Por lo tanto, los valores predichos de una regresión lineal son simplemente la proyección ortogonal de Y sobre el espacio definido por X.

Forma gráfica:

Para apreciar la conexión que hay entre una proyección lineal y su regresión equivalente primero tenemos que invertir la forma en la que típicamente pensamos sobre observaciones, variables y *datapoints*.
Consideremos una regresión hipotética de dos variables con solamente tres observaciones. Nuestros datos incluyen una constante "c" (un vector de unos), una variable independiente "X", y una variable dependiente "Y":

Y	X	C
2	3	1
3	1	1
2	1	1

Usualmente, representaríamos la relación entre las variables de forma geométrica, tratándolas como si fueran dimensiones.
Una forma alternativa de representar estos datos es tratar a cada observación (a cada fila) como una dimensión y después representar cada variable como un vector.
Consideremos, por ejemplo, la columna "Y". Este vector esencialmente nos da las coordenadas para un punto en un espacio tridimensional que corresponde a:

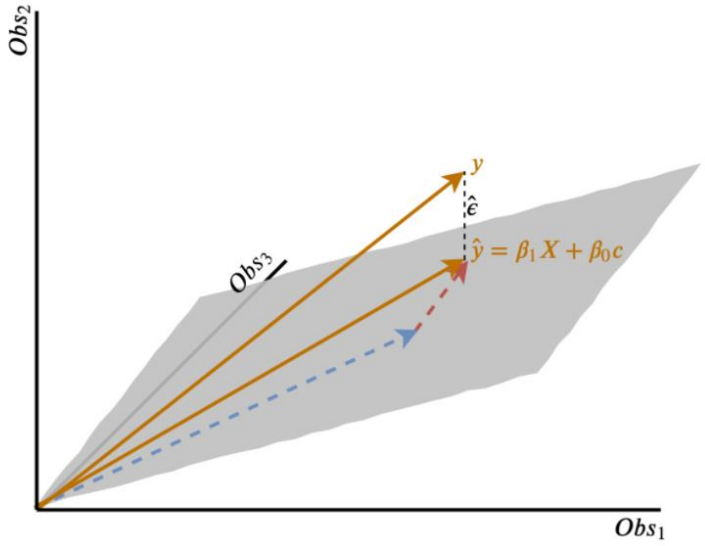
$$(x, y, z) = (2,3,2)$$

El conjunto de todos los puntos cubiertos por la combinación lineal de "X" y "c" es el denominado *span*. Para el caso especifico de este ejemplo, el span de las combinaciones lineales de X y c es un plano en el espacio tridimensional denominado *col(X, c)*.

Es importante tener en cuenta que, en nuestro ejemplo reducido de un espacio de tres dimensiones, hay puntos que no son alcanzables por ninguna combinación de X y c (cualquier punto por encima o por debajo del plano formado por X y c). Sabemos, particularmente, que el vector que representa a "Y" se encuentra por fuera de dicho plano. El problema entonces es encontrar un vector que se encuentre en el plano *col(X, c)* que se acerque lo más posible al vector Y que se encuentra fuera de dicho plano.

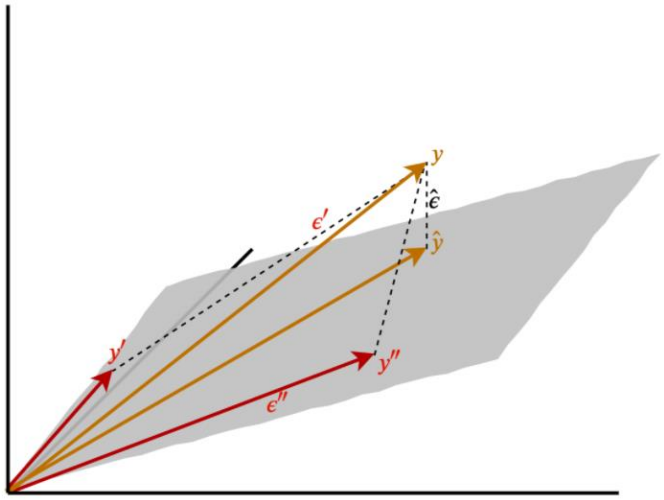
Sabemos que el vector que mejor se ajusta a la condición de "cercanía" con Y es aquel que surge de la proyección ortogonal del vector Y sobre el plano col(X,c). Esta es la distancia más corta posible entre el plano y el vector Y, y se la suele denominar como $\hat{Y}$ ($\hat{Y} = \beta_0 * c + \beta_1 * X$). Los coeficientes β_0 y β_1 son los regresores derivados de MCO (Mínimos Cuadrados Ordinarios). Esto es porque sabemos que la proyección ortogonal de Y sobre el plano minimiza el error entre nuestra predicción $\hat{Y}$ y los valores observados de Y. Esto es el mismo problema de minimización que resuelve la aplicación del proceso de MCO.

La siguiente figura grafica la situación hasta ahora descripta:



Consideremos cualquier otro vector en el plano, y la distancia entre dicho vector y el vector Y. Cualquier vector no-ortogonal tendría mayor magnitud que el vector $\hat{Y}$, y, por lo tanto, tendría un mayor error de predicción.

Por ejemplo, la siguiente figura grafica dos vectores alternativos sobre el plano *col(X,c)* junto con el vector $\hat{Y}$. Podemos observar que la distancia euclidiana entre $Y'e\ Y$, y entre $Y''e\ Y$, es mayor que la distancia entre el vector original $\hat{Y}$ e Y. Es decir $\hat{\epsilon} < \epsilon' < \epsilon''$.



Entonces, la proyección lineal y la regresión lineal pueden ser vistas, tanto algebraica como geométricamente, como la solución al mismo problema, a saber, la minimización de la distancia al cuadrado entre un vector observado Y y el valor predicho de dicho vector $\hat{Y}$. Esta demostración se generaliza para cualquier cantidad de dimensiones (observaciones), aunque evidentemente se vuelve más difícil de representar gráficamente. Similarmente, con más observaciones podríamos extender el numero de variables independientes de forma tal que X no sea solamente un vector-columna, sino una matriz de variables independientes (como en el caso de la regresión del logaritmo del salario mostrada anteriormente). Nuevamente, visualizar la forma de dicha matriz multidimensional en un grafio sería bastante dificultoso, o imposible.

Queda explicado entonces que proyectar un vector sobre un espacio de menores dimensiones implica encontrar la combinación lineal de vectores del espacio mencionado que minimiza la distancia euclidiana entre el espacio y el vector extra-dimensional. Los escalares que utilizamos para realizar dicho proceso son los coeficientes de regresión que obtenemos de la aplicación del proceso de MCO.