



## Técnicas de Minería de Datos (problema de Componentes Principales)

Prof: Max Chacón

Un problema clásico de macro-economía es analizar los países según sus características socio-económicas. Lo fundamental del análisis es comparar los diferentes países según su desarrollo relativo y construir un indicador socio-económico que permita clasificarlos.

Un ejemplo clásico es el análisis realizado entre los países latinoamericanos, los datos correspondientes al año 1979 se muestran en la siguiente tabla.

País	X1 Tasa de mortalidad	X2 PNB	X3 Gasto Educación	X4 Habitantes por médico	X5 Calorías diarias por persona
Argentina	58	1.160	153	521	3.069
Bolivia	106	180	186	2.174	2.122
Brasil	105	420	65	1.341	2.731
Colombia	76	340	165	2.341	2.233
Chile	88	720	186	1.803	2.655
Ecuador	91	290	225	2.928	2.014
Méjico	69	670	243	1.726	2.689
Paraguay	84	260	144	1.811	2.835
Perú	110	450	207	1.917	2.428
Uruguay	43	820	260	1.032	3.167
Venezuela	49	980	183	1.115	2.641

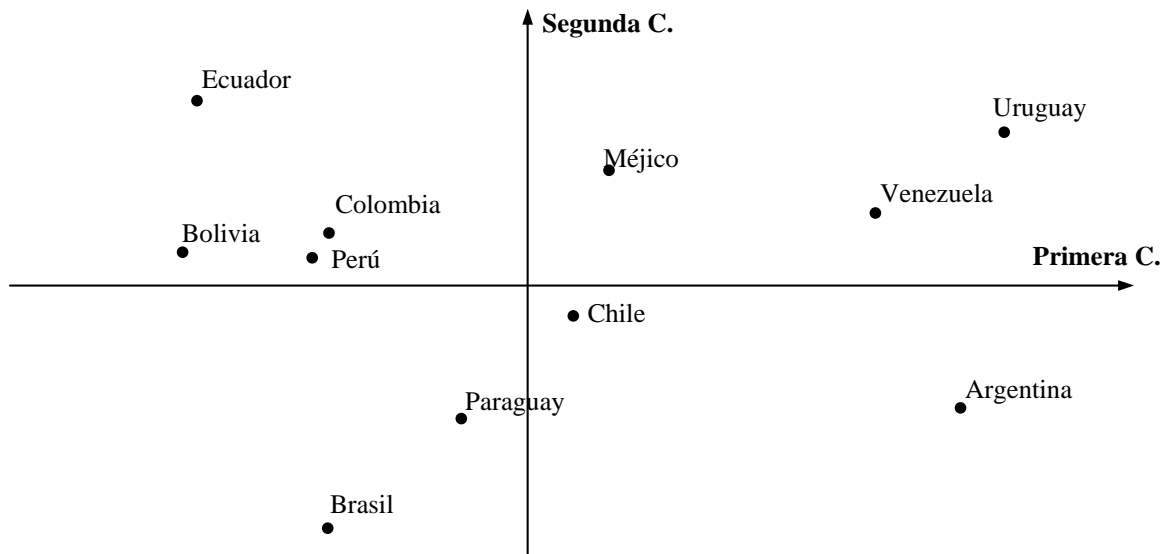
La matriz de correlación es la siguiente:

	X1	X2	X3	X4	X5
X1	1				
X2	-0,75	1			
X3	-0,39	0,19	1		
X4	0,68	-0,87	-0,03	1	
X5	-0,60	0,69	-0,04	-0,86	1

Un análisis de *Componentes Principales* arroja los siguientes resultados, para las dos primeras componentes:

- Primer valor propio normalizado: 0,65 Vector propio: [-0,47; 0,51; 0,12; -0,52; 0,48].
- Segundo valor propio normalizado: 0,23 Vector propio: [-0,28; 0,005; 0,89; 0,21; -0,29].

La distribución gráfica de los países en el plano de las dos primeras componentes es la siguiente:



Utilizando la interpretación geométrica de la transformación ubique en el plano, en forma aproximada, los ejes originales ( $X_i$ ).

Identifique la región del plano que va en el sentido del desarrollo socio-económico de los países. Determine cual es la relación de cada componente principal con las variables originales y cual de las componentes puede ser usada como indicador socio-económico.

Clasifique los países en forma aproximada en función de su desarrollo.

Indique el porcentaje de información no considerado en el análisis.