Universidad de Buenos Aires - Facultad de Ciencias Exactas y Naturales

Álgebra Lineal Computacional

2do Cuatrimestre 2024

Trabajo Práctico N° 2: Matrices Insumo-Producto.

Expresando $(\mathbf{I}-\mathbf{A})^{-1}$ como una suma infinita

Retomando la ecuación

$$\mathbf{p} = (\mathbf{I}-\mathbf{A})^{-1}\,\mathbf{d} \tag{1}$$

que daba la relación entre la demanda externa y lo que el sector produce. La inversa de la matriz $(\mathbf{I}-\mathbf{A})^{-1}$ se puede resolver de un análisis en sumas infinitas, o en series de Taylor que definen la siguiente igualdad:

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots \tag{2}$$

bajo la restricción de que $|x| < 1$. Lo podemos reescribir de manera más similar a nuestro problema de la forma:

$$(1-x)^{-1} = 1 + x + x^2 + x^3 + \ldots \quad \forall |x| < 1 \tag{3}$$

Luego, bajo ciertas condiciones de una matriz cuadrada $\mathbf{A}$ podemos tener:

$$(\mathbf{I}-\mathbf{A})^{-1} = \mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \ldots + \mathbf{A}^n + \ldots \tag{4}$$

Consigna 1. Probar que

$$(\mathbf{I}-\mathbf{A})^{-1} = \mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \ldots + \mathbf{A}^n + \ldots \Rightarrow \lim_{n\to\infty} \mathbf{A}^{n+1} = \mathbf{0} \tag{5}$$

la vuelta también vale, pero para la consigna dejamos una sola dirección.

Vamos a estudiar en la práctica la convergencia de esta suma infinita

Consigna 2. Comenzamos graficando, para las dos matrices $\mathbf{A}_1$ y $\mathbf{A}_2$, dos vectores definidos de la forma: $a_1(n) = \|\mathbf{A}_1^n\|_2$ y $a_2(n) = \|\mathbf{A}_2^n\|_2$. O sea, que cada uno de sus elementos representa la norma 2 de la matriz elevada a la potencia que corresponde a su índice (para esta parte se puede utilizar `scipy.linalg.norm`). Los vectores los definimos de largo $N = 250$.

$$\mathbf{A}_1 = \begin{pmatrix} 0{,}186 & 0{,}521 & 0{,}014 & 0{,}32 & 0{,}134 \\ 0{,}24 & 0{,}073 & 0{,}219 & 0{,}013 & 0{,}327 \\ 0{,}098 & 0{,}12 & 0{,}311 & 0{,}302 & 0{,}208 \\ 0{,}173 & 0{,}03 & 0{,}133 & 0{,}14 & 0{,}074 \\ 0{,}303 & 0{,}256 & 0{,}323 & 0{,}225 & 0{,}257 \end{pmatrix}$$

$$\mathbf{A}_2 = \begin{pmatrix} 0{,}186 & 0{,}521 & 0{,}014 & 0{,}32 & 0{,}134 \\ 0{,}24 & 0{,}073 & 0{,}219 & 0{,}013 & 0{,}327 \\ 0{,}098 & 0{,}12 & 0{,}311 & 0{,}302 & 0{,}208 \\ 0{,}173 & 0{,}03 & 0{,}133 & 0{,}14 & 0{,}074 \\ 0{,}003 & 0{,}256 & 0{,}323 & 0{,}225 & 0{,}257 \end{pmatrix}$$

Consigna 3. Mediante el método de la potencia, encontrar el mayor autovalor de las matrices $\mathbf{A}_1$ y $\mathbf{A}_2$. Implementar ese cálculo en la función `metodoPotencia` del archivo `funciones.py`. Dado que el método de la potencia es inicializado con un vector aleatorio x_0, vamos a hacer el cálculo del autovalor estadísticamente con el método de Monte Carlo. Este procedimiento repite la operación un número importante de veces (digamos 250). Se pide en la consigna transcribir en una tabla el promedio de los autovalores de las 250 iteraciones de Monte Carlo para cada matriz, poniendo además el desvío estandar encontrado.

Consigna 4.

(a) De forma similar a la consigna 2, graficar la serie de potencias $(\mathbf{I}-\mathbf{A})^{-1}$ de $\mathbf{A}_1$ y $\mathbf{A}_2$ (siguiendo la ecuación 4) para para $n = 10$ y $n = 100$.

(b) ¿Hay convergencia para ambas matrices? Argumentar la respuesta.

(c) ¿Podría establecer una regla por la cual se pueda asegurar convergencia de la serie infinita?

(d) En caso de existir convergencia, grafique el error obtenido como un vector $e(n) = \|(\mathbf{I}+\mathbf{A}+\mathbf{A}^2+\mathbf{A}^3+\ldots+\mathbf{A}^n)-(\mathbf{I}-\mathbf{A})^{-1}\|_2$.
Nota: calcular $(\mathbf{I}-\mathbf{A})^{-1}$ con la función `inversaLU` desarrollada por el grupo en el TP1.

Consigna 5. Retomar de las matrices insumo-producto de los paises P1 y P2 asignados en el TP1, y mediante el método de la potencia, encontrar el autovalor mayor de cada uno de las $\mathbf{A}^{ss}$ y $\mathbf{A}^{rr}$. Decidir si existe convergencia de la suma de potencias, a partir del argumento dado en la Consigna 4 (c).

Análisis en Componentes Principales

Consigna 6. Sea $n \geq 2$ y $\mathbf{E}_n \in \mathbb{R}^{n\times n}$ una matriz definida como $\mathbf{E}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{e}\,\mathbf{e}^t$ donde $\mathbf{e} \in \mathbb{R}^n$ es un vector columna de todos 1s y $\mathbf{I}_n$ es la matriz identidad de $n \times n$.

(a) Probar que para cualquier vector $\mathbf{z} \in \mathbb{R}^n$, el promedio de las componente del vector $\mathbf{E}_n\mathbf{z}$ es 0. Analizar $\mathbf{E}_n\mathbf{A}$ y describir qué efecto tiene $\mathbf{E}_n$ sobre una matriz $\mathbf{A}$ de $n \times n$.

(b) Calcular analíticamente todos los autovalores y autovectores de $\mathbf{E}_n$. Determinar la imagen y el núcleo de $\mathbf{E}_n$, y la dimensión de los mismos.
Sugerencia: considerar la base de vectores $\{\mathbf{e}, \mathbf{d}_2, \ldots, \mathbf{d}_n\}$ de $\mathbb{R}^n$ siendo $\{\mathbf{d}_2, \ldots, \mathbf{d}_n\}$ una base de $\langle \mathbf{e} \rangle^{\perp}$.

(c) Probar que $\mathbf{E}_n$ es una matriz de proyección ortogonal. ¿Sobre qué subespacio proyecta?

(d) Responder Verdadero o Falso, demostrando o dando un contraejemplo:

(i) $\mathbf{E}_n$ es una matriz ortogonal.
(ii) $\mathbf{E}_n$ es singular.
(iii) $\mathbf{E}_n$ es definida positiva.
(iv) traza$(\mathbf{E}_n) = n - 1$.

Consigna 7. Análisis en Componentes Principales (ACP) por el método de la potencia. Calcular la matriz de covarianzas $\mathbf{C}$ del país $P1$ a partir de las matriz $\mathbf{A}^{rr}$ desde el Excel de CEPAL. Para ello se debe:

- Normalizar la matriz de coeficientes $\mathbf{A}_{rr}$ de la forma:

$$\overline{\mathbf{A}^{rr}} = \mathbf{E}_{40}\, \mathbf{A}^{rr}$$

- Calcular la matriz de covarianzas como:

$$\mathbf{C} = \overline{\mathbf{A}^{rr}}^{\,t}\, \overline{\mathbf{A}^{rr}} / \,(40-1)$$

Encontrar los 2 primeros autovectores usando el método de la potencia de la forma siguiente.

Algoritmo (Deflación de Hotelling):

- Inicialmente multiplicar un vector aleatorio $\mathbf{x}_0$ de norma 1 a $\mathbf{C}$ y luego continuar iterando utilizando la siguiente regla:

$$\mathbf{x}_1 = \mathbf{C}\mathbf{x}_0 \tag{6}$$

$$\mathbf{x}_{k+1} = \mathbf{C}\mathbf{x}_k \tag{7}$$

  para encontrar el primer autovector (con máximo autovalor).

- En cada paso se debe volver a normalizar el siguiente $\mathbf{x}_k$.
- Continuar mientras $\|\mathbf{x}_{k+1}^t - \mathbf{x}_k\|_2 > (1-\epsilon)$, que es el criterio de parada para el algoritmo iterativo, con $\epsilon > 0$.
- Devolver $\mathbf{x}_{k+1}$ que es la aproximación al autovector buscado $\mathbf{v}_1$.
- Obtenemos, a partir del coeficiente de Rayleigh $\lambda_1 = \frac{\mathbf{v}_1^t\, \mathbf{C}\mathbf{v}_1}{\mathbf{v}_1^t\, \mathbf{v}_1}$
- Se procede recursivamente y se calcula la nueva matriz $\mathbf{C}'$ sobre la cual se va a iterar nuevamente:

$$\mathbf{C}' = \mathbf{C} - \lambda_1 \mathbf{v}_1 \mathbf{v}_1^t \tag{8}$$

  para encontrar el siguiente autovector $\mathbf{v}_2$ y luego λ_2.

Consigna 8. Utilizando los dos autovectores del Análisis en Componentes Principales de la consigna 5, proyectar las filas de $\mathbf{A}^{rr}$ en un scatter de 2 dimensiones. En este scatter identificar visualmente al menos dos clusters de 3 o más sectores que se encuentren a una distancia euclideana baja. *Nota: En el caso de que el país P1 asignado al grupo no tenga suficiente información pertinente para encontrar los clusters, probar con el país asignado P2.*

Matriz H

Consigna 9. En este caso utilizaremos la matriz $\mathbf{H}$ que representa la matriz input-output integrada verticalmente. Esta matriz se define mediante la ecuación

$$\mathbf{H} = \mathbf{A}\,(\mathbf{I} - \mathbf{A})^{-1} \tag{9}$$

Se dice que esta matriz tiene la particularidad de reforzar las relaciones entre los sectores que de ya por si tienen una alta relevancia en $\mathbf{A}$. Encontrar cuál es la relación de $\mathbf{H}$ con la suma de potencias.

Consigna 10. Repetir el análisis en Componentes Principales para las matrices $\mathbf{H}$, separando en clusters y analizando los sectores agrupados.

Consigna 11. Vamos a analizar el perfil de producción para sectores bien diferenciados.
Para ello,

- En el espacio de ACP en 2 dimensiones, calcular las distancias al origen de coordenadas de cada punto.
- Identificar el sector más lejano y el más cercano al origen.
- Graficar la producción (fila) en $\mathbf{A}^{rr}$ y $\mathbf{H}$ de estos dos sectores.
- Las curvas obtenidas para el caso más cercano y el más lejano al origen (de cada caso) se deberían mostrar diferentes. ¿Cuál sería la base de esa diferencia y lo que los hace estar tanto lejos como cerca del origen?

Entrega y lineamientos

La entrega se realizará a través del campus virtual de la materia con las siguientes fechas y formato:

- Fecha de entrega: hasta el viernes **8 de noviembre** a las 23:59 hs.
- Fecha de re-entrega: hasta el **26 de noviembre** a las 23:59 hs.
  Es condición obligatoria haber realizado un envío en la primera fecha **de entrega para tener la posibilidad de reentregar.**
- Formato: Jupyter Notebook del template-alumnos y archivo funciones.py.

Prestar especial atención a las siguientes indicaciones:

- El TP se realizará en grupos de tres personas. Se puede usar el foro 'Foro de Grupos de TP'.

  Importante: es indispensable realizar la inscripción previa del grupo para poder hacer el envío a través del campus. Los grupos o personas no inscriptas en grupos no estarán habilitadas en el formulario de carga del TP. No se aceptarán envíos por email.
- Leer el enunciado completo antes de comenzar a generar código y sacarse todas las dudas de cada ítem antes de implementar. Para obtener un código más legible y organizado, pensar de antemano qué funciones deberán implementarse y cuáles podrían reutilizarse.
- El código debe estar correctamente comentado. Cada función definida debe contener un encabezado donde se explique los parámetros que recibe y qué se espera que retorne. Además las secciones de código dentro de la función deben estar debidamente comentados. Los nombre de las variables deben ser explicativos.

- Las conclusiones y razonamientos que respondan los ejercicios, o cualquier experimentación agregada, debe estar debidamente explicada en bloques de texto de las notebooks (markdown cells), separado de los bloques de código. Aprovechen a utilizar código LaTeX si necesitan incluir fórmulas.
- Gráficos: deben contener título, etiquetas en cada eje y leyendas indicando qué es lo que se muestra.