Modelos matemáticos y numéricos

Proyecto #1: Reconocimiento de patrones

Instructor: JL Morales

Descomposición en valores singulares

Teorema 1 (Descomposición en valores singulares). *Si A es una matriz real de $m \times n$, entonces existen matrices ortogonales*

$$U = [u_1, \ldots, u_m] \in \mathbb{R}^{m \times m}, \quad V = [v_1, \ldots, v_n] \in \mathbb{R}^{n \times n}$$

tales que

$$U^T A V = \mathrm{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{m \times n}, \quad p = \min\{m, n\}$$

en donde $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$.

Prueba 1. *Consideremos la norma matricial inducida por la norma vectorial euclidiana; entonces es claro que existen 2 vectores unitarios $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ con la propiedad*

$$Ax = \sigma y, \quad \sigma = ||A||_2,$$

ahora podemos construir matrices ortogonales U, V a partir de los vectores x, y como sigue

$$V = [\, x \,|\, V_2 \,], \quad U = [\, y \,|\, U_2 \,].$$

De la construcción anterior es fácil probar que

$$U^T A V = \begin{bmatrix} \sigma & w^T \\ 0 & B \end{bmatrix} = A_1$$

en donde $w \in \mathbb{R}^{n-1}$ y $B \in \mathbb{R}^{(m-1)\times(n-1)}$; de la desigualdad

$$\left|\left| A_1 \begin{bmatrix} \sigma \\ w \end{bmatrix} \right|\right|_2^2 \geq (\sigma^2 + w^T w)^2$$

podemos concluir que $||A_1||_2^2 \geq \sigma^2 + w^T w$. Sin embargo $\sigma^2 = ||A||_2^2 = ||A_1||_2^2$, por lo tanto $w = 0$. Es decir que A_1 tiene la estructura diagonal por bloques

$$A_1 = \begin{bmatrix} \sigma & 0 \\ 0 & B \end{bmatrix}$$

Un argumento inductivo evidente completa la prueba del resultado. □

Propiedades

La DVS revela prácticamente toda la estructura algebraica de una matriz. Por ejemplo:

1. Los valores singulares de A son las longitudes de los semi ejes del hiperelipsoide

$$H = \{Ax : ||x||_2 = 1\}$$

definido por A.

2. Si definimos r como

$$\sigma_1 \geq \cdots \sigma_r > \sigma_{r+1} = \cdots = \sigma_p = 0,$$

entonces

   (a) $\text{rango}(A) = r$

   (b) $K(A) = \text{gen}\,\{v_{r+1}, \ldots, v_n\}$

   (c) $R(A) = \text{gen}\,\{u_1, \ldots, u_r\}$

3. La matriz A se puede representar como una suma de matrices de rango 1 (desarrollo DVS)

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T. \tag{1}$$

4. Diversas relaciones entre $||A||_\text{F}$ y $||A||_2$

   (a) $||A||_\text{F} = \sigma_1^2 + \cdots + \sigma_p^2, \quad p = \min\{m, n\}$

   (b) $||A||_2 = \sigma_1$

   (c) $\min\limits_{||x||_2=1} ||Ax||_2 = \sigma_n, \quad (m \geq n).$

Ejercicios

1. Investigar cómo se construye la DVS en forma numéricamente estable. ¿Cuál es la complejidad del método más difundido?

2. Considerar el problema de mínimos cuadrados $Ax = b$, en donde A es una matriz real de $m \times n$ de rango completo por columnas $(m > n)$. Obtener la expresión

$$\bar{x} = (A^T A)^{-1} A^T b$$

mediante argumentos de álgebra lineal.

3. Interpretar el método de análisis de componentes principales en el contexto de la DVS.

4. Investigar el significado e impacto práctico de la factorización de Schur (matrices reales y complejas).

Proyecto

Clasificar automáticamente dígitos manuscritos es un problema muy difundido en el área de reconocimiento de patrones. La aplicación más popular consiste en automatizar la lectura de los códigos postales en piezas de correo.

En este proyecto, los dígitos consisten de imágenes de 20×20 pixeles en tonos de gris; para fines prácticos las imágenes se tratan como vectores en $\mathbb{R}^{400}$, obtenidos apilando las 20 columnas de una imagen. Para cada dígito d, $d \in \{0, 1, \ldots, 9\}$, tenemos una muestra de 500 elementos organizada como una matriz de 400×500.

Consideremos un dígito particular, digamos $d \in \{0, 1, \ldots, 9\}$. Si obtenemos la DVS de A^d, entonces la propiedad (1)

$$A^d = \sum_{i=1}^{r} \sigma_i u_i v_i^T.$$

indica que cada columna de A^d, denotada como a_j, se puede expresar como

$$a_j = \sum_{i=1}^{m} (\sigma_i v_{ij}) u_i,$$

en donde v_{ij} es la j-ésima componente del vector v_i. La expresión anterior sugiere que cada columna se puede aproximar por los primeros K términos de la suma, es decir

$$a_j \approx \sum_{i=1}^{K} (\sigma_i v_{ij}) u_i.$$

Notar que los primeros vectores singulares de A^d contienen la *información dominante.*

Ahora podemos identificar un dígito desconocido, digamos z, resolviendo el siguiente problema de mínimos cuadrados (MC).

$$\text{minimizar} \quad ||z - U_K \alpha||_2^2,$$

en donde U_K es la submatriz de U formada por los primeros K vectores singulares. La solución del problema de MC es

$$\alpha = U_K^T z,$$

mientras que la norma del residuo es

$$||r(z)||_2 = ||z - U_K U_K^T z||_2 \tag{2}$$

Es inmediato concluir que si $||r(z)||_2$ es muy pequeña, entonces probablemente $z = d$.

El razonamiento anterior se puede formular como el siguiente algoritmo:

1. Inicio.
   - Calcular la DVS de las matrices $A^j, \quad j = 0, \ldots, 9.$
2. Elegir aleatoriamente z, un dígito de la muestra completa
   - resolver (2) para cada dígito
   - asignar z a la clase con menor error $||r(z)||_2$.
   - verificar si el dígito fue identificado correctamente.
3. Repetir 2. para 100 elecciones de z y obtener la tasa de éxito

Reporte técnico

Utilizar parte del archivo LaTeX de este documento para escribir un reporte en el que se discutan los resultados obtenidos. La estructura sugerida para el reporte es:

1. Introducción. El problema por resolver. Relevancia.
2. Fundamentos matemáticos. El teorema de existencia de la DVS + la correspondiente prueba.
3. Fundamentos numéricos. La realización práctica de la DVS mediante un algoritmo numéricamente estable.
4. Experimentos. Reportar la tasa de éxito para diferentes valores de K.
5. Conclusiones.