REORGANIZACIÓN DE LA INFORMACIÓN EN MATRICES

29 de marzo de 2006

1. Matrices de rango uno

Una matriz genérica de rango uno se puede escribir en la forma

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \begin{bmatrix} b_1 & \dots & b_m \end{bmatrix} = \begin{bmatrix} a_1 b_1 & \dots & a_1 b_m \\ \vdots & & \vdots \\ a_n b_1 & \dots & a_n b_m \end{bmatrix},$$

es decir, las matrices de rango uno son aquéllas de la forma $(a_i\, b_j)$ donde los vectores $\mathbf{a} = (a_1, \dots, a_n)^\top$ y $\mathbf{b} = (b_1, \dots, b_m)^\top$ son no nulos, es decir, con el convenio de que los vectores son columnas, tenemos la matriz,

$$\mathbf{a}\,\mathbf{b}^\top.$$

El producto por una matriz de rango uno es muy simple

$$\mathbf{a}\,\mathbf{b}^\top \mathbf{x} = (\mathbf{b}, \mathbf{x})\,\mathbf{a}, \qquad (\mathbf{b}, \mathbf{x}) = \mathbf{b} \cdot \mathbf{x} = \mathbf{b}^\top \mathbf{x} = \sum_{j=1}^{m} b_j\, x_j.$$

De aquí se deduce trivialmente que

$$\mathcal{N}(\mathbf{a}\,\mathbf{b}^\top) = \{\mathbf{x} \in \mathbb{R}^m \,|\, (\mathbf{b}, \mathbf{x}) = 0\}$$

La descomposición de un vector en una base ortogonal es un ejemplo de empleo de matrices de rango uno. Si tenemos $\mathbf{p}_1, \dots, \mathbf{p}_n$ una base ortonormal de $\mathbb{R}^n$, entonces

$$\mathbf{x} = \sum_{j=1}^{n} (\mathbf{p}_j, \mathbf{x})\,\mathbf{p}_j = \sum_{j=1}^{n} \mathbf{p}_j\,\mathbf{p}_j^\top \mathbf{x}.$$

Esta operación se puede ver empleando una matriz que agrupe a los vectores de la base ortonormal

$$P = \left[\begin{array}{c|c|c} \mathbf{p}_1 & \dots & \mathbf{p}_n \end{array}\right].$$

Que la matriz P sea ortogonal significa que sus columnas son ortonormales, esto es, $P^\top P = I$. Por ser cuadrada (la ortogonalidad de matrices también se admite para matrices rectangulares cuyas columnas sean ortonormales), $P^\top = P^{-1}$ y por tanto, también tenemos que $P\,P^\top = I$, esto es, las filas de una matriz cuadrada ortogonal forman otra base ortonormal de $\mathbb{R}^n$.

El hecho de que $P\,P^\top = I$ es el mismo que nos permite escribir que para todo $\mathbf{x}$,

$$\mathbf{x} = P(P^\top \mathbf{x}) = \left[\begin{array}{c|c|c} \mathbf{p}_1 & \ldots & \mathbf{p}_n \end{array}\right] \begin{bmatrix} (\mathbf{p}_1, \mathbf{x}) \\ \vdots \\ (\mathbf{p}_n, \mathbf{x}) \end{bmatrix} = \sum_{j=1}^{n} (\mathbf{p}_j, \mathbf{x})\, \mathbf{p}_j.$$

En la expresión anterior hemos empleado las dos interpretaciones del producto matriz por vector, que en determinados contextos del mundo de la ingeniería se llaman *input–output* y *output–input*. Las expresiones anteriores también permiten escribir la siguiente descomposición de la matriz identidad como suma de matrices de rango uno

$$I = \sum_{j=1}^{n} \mathbf{p}_j\, \mathbf{p}_j^\top.$$

Regresemos momentáneamente a las matrices de rango uno. Si una matriz de rango uno cuadrada $(a_i\, b_j)$ es simétrica, entonces

$$\mathbf{b}\,\mathbf{a}^\top = \mathbf{a}\,\mathbf{b}^\top,$$

luego $\mathbf{b} = \lambda\,\mathbf{a}$ (basta multiplicar a derecha por $\mathbf{a}$), y la matriz se puede escribir como

$$\lambda\,\mathbf{a}\,\mathbf{a}^\top.$$

Obviamente,

$$(\lambda\,\mathbf{a}\,\mathbf{a}^\top \mathbf{x}, \mathbf{x}) = \lambda(\mathbf{a}^\top \mathbf{x})^\top \mathbf{a}^\top \mathbf{x} = \lambda |(\mathbf{a}, \mathbf{x})|^2.$$

Si $\lambda > 0$, entonces, la matriz es semidefinida positiva y si $\lambda < 0$, es semidefinida negativa. Como la matriz tiene rango uno, no puede ser indefinida, ya que solo tiene un valor propio no nulo.

Tomemos ahora números $\lambda_j \in \mathbb{R}$ y una base ortonormal de $\mathbb{R}^n$: $\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$. Consideremos entonces una matriz simétrica (es suma de matrices simétricas):

$$A := \sum_{j=1}^{n} \lambda_j \mathbf{p}_j\, \mathbf{p}_j^\top.$$

Multiplicar por esta matriz es simple,

$$A\mathbf{x} = \sum_{j=1}^{n} \lambda_j (\mathbf{p}_j, \mathbf{x})\, \mathbf{p}_j.$$

Notemos que, en particular,

$$A\mathbf{p}_i = \lambda_i \mathbf{p}_i$$

y que

$$A = P\Lambda P^\top, \qquad \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}, \qquad P = \left[\begin{array}{c|c|c} \mathbf{p}_1 & \ldots & \mathbf{p}_n \end{array}\right].$$

Así pues, $P^\top AP = \Lambda$ no es más que la diagonalización ortogonal de una matriz simétrica y las matrices

$$\sum_{j=1}^{n} \lambda_j \mathbf{p}_j\, \mathbf{p}_j^\top$$

cubren todo el conjunto de las matrices simétricas.

2. Compresión de información

Vamos a restringirnos en esta sección a matrices simétricas definidas positivas. Tomamos números

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$$

y una base ortonormal de $\mathbb{R}^n$, $\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$, de manera que la diagonalización ortogonal de A es equivalente a la descomposición

$$A = \sum_{j=1}^{n} \lambda_j \mathbf{p}_j\, \mathbf{p}_j^\top.$$

Notemos también que si

$$|\mathbf{x}| := \sqrt{(\mathbf{x}, \mathbf{x})} = \sqrt{\sum_{j=1}^{n} |x_j|^2},$$

entonces

$$P^\top \mathbf{x} = \begin{bmatrix} (\mathbf{p}_1, \mathbf{x}) \\ \vdots \\ (\mathbf{p}_n, \mathbf{x}) \end{bmatrix}$$

y

$$\sum_{j=1}^{n} |(\mathbf{p}_j, \mathbf{x})|^2 = |P^\top \mathbf{x}|^2 = (P^\top \mathbf{x})^\top (P^\top \mathbf{x}) = \mathbf{x}^\top P P^\top \mathbf{x} = \mathbf{x}^\top \mathbf{x} = |\mathbf{x}|^2.$$

De forma similar, se ve que

$$|A\mathbf{x}|^2 = \sum_{j=1}^{n} \lambda_j^2 |(\mathbf{p}_j, \mathbf{x})|^2.$$

Consideremos ahora las matrices

$$A_k := \sum_{j=0}^{k} \lambda_j \mathbf{p}_j\, \mathbf{p}_j^\top.$$

Las matrices A_k son simétricas y semidefinidas positivas. Su rango es k, ya que

$$A_k\mathbf{x} \in \{\alpha_1\mathbf{p}_1 + \ldots + \alpha_k\mathbf{p}_k \,|\, \alpha_j \in \mathbb{R}\},$$

a la vez que

$$A_k\mathbf{p}_j = \lambda_j\mathbf{p}_j, \qquad j = 1, \ldots, k.$$

El espacio nulo de A_k es

$$\mathcal{N}(A_k) = \{\alpha_{k+1}\mathbf{p}_{k+1} + \ldots + \alpha_n\mathbf{p}_n \,|\, \alpha_j \in \mathbb{R}\}.$$

Además,

$$A\mathbf{x} - A_k\mathbf{x} = \sum_{j=k+1}^{n} \lambda_j(\mathbf{p}_j, \mathbf{x})\,\mathbf{p}_j$$

luego

$$|A\mathbf{x} - A_k\mathbf{x}|^2 = \sum_{j=k+1}^{n} \lambda_j^2|(\mathbf{p}_j, \mathbf{x})|^2 \leq \lambda_{k+1}^2 \sum_{j=k+1}^{n} |(\mathbf{p}_j, \mathbf{x})|^2 \leq \lambda_{k+1}^2 \sum_{j=1}^{n} |(\mathbf{p}_j, \mathbf{x})|^2 = \lambda_{k+1}^2|\mathbf{x}|^2.$$

Merece la pena observar que en la inversa, el orden de los valores propios es el contrario

$$A^{-1} = \sum_{j=1}^{n}(1/\lambda_j)\mathbf{p}_j\,\mathbf{p}_j^\top.$$

Este resultado se deduce del hecho de que si $A = P\Lambda P^\top$, entonces $A^{-1} = P\Lambda^{-1}P^\top$. Al número

$$\mathrm{cond}(A) := \frac{\lambda_1}{\lambda_n}$$

(producto de los mayores valores propios de A y A^{-1}) se le llama condicionamiento espectral de A.

El mismo argumento se puede usar para cualquier tipo de matriz simétrica, ordenando los valores λ_j por módulo. Si ninguno de ellos es cero, se puede invertir la matriz y repetir los comentarios sobre la inversa.

3. Matrices no simétricas

Consideremos una matriz B cualquiera con n filas y m columnas. La matriz $B^\top B$ es $m \times m$, simétrica y semidefinida positiva. Así

$$B^\top B = \sum_{j=1}^{r} \sigma_j^2\mathbf{p}_j\,\mathbf{p}_j^\top,$$

donde

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0,$$

y los vectores $\mathbf{p}_1, \ldots, \mathbf{p}_r \in \mathbb{R}^m$ son ortonormales. El número r es el rango de $B^\top B$. Ahora bien, también es el rango de B. ¿Por qué? Para verlo, notemos primero que

$$B\mathbf{x} = \mathbf{0} \qquad \Longleftrightarrow \qquad B^\top B\mathbf{x} = \mathbf{0}.$$

Así, aprovechando que el número de columnas de B y de $B^\top B$ es el mismo, tenemos que

$$r = \mathrm{rango}(B^\top B) = m - \dim \mathcal{N}(B^\top B) = m - \dim \mathcal{N}(B) = \mathrm{rango}(B).$$

Definimos seguidamente los vectores

$$\mathbf{q}_j := (1/\sigma_j)B\mathbf{p}_j \in \mathbb{R}^n, \qquad j = 1, \ldots, r.$$

Notemos que

$$B^\top \mathbf{q}_j = (1/\sigma_j)B^\top B\mathbf{p}_j = \sigma_j \mathbf{p}_j, \qquad j = 1, \ldots, r.$$

Además estos vectores cumplen

$$(\mathbf{q}_i, \mathbf{q}_j) = \frac{1}{\sigma_i\,\sigma_j}(B\mathbf{p}_i, B\mathbf{p}_j) = \frac{1}{\sigma_i\,\sigma_j}\mathbf{p}_i^\top B^\top B\mathbf{p}_j = \frac{\sigma_j}{\sigma_i}\mathbf{p}_i^\top\,\mathbf{p}_j = \frac{\sigma_j}{\sigma_i}(\mathbf{p}_i, \mathbf{p}_j) = \delta_{ij},$$

luego son ortonormales. Así, como $\dim \mathcal{R}(B) = r$ y los vectores $\mathbf{q}_j$ son ortogonales y están en $\mathcal{R}(B)$ (por su definición), entonces

$$\mathcal{R}(B) = \{\beta_1\mathbf{q}_1 + \ldots + \beta_r\mathbf{q}_r \,|\, \beta_j \in \mathbb{R}\}.$$

Por tanto, para cualquier $\mathbf{x} \in \mathbb{R}^m$

$$B\mathbf{x} = \sum_{j=1}^r \xi_j\mathbf{q}_j, \qquad \xi_j \in \mathbb{R}$$

y se tiene que

$$\sum_{j=1}^r \sigma_j^2(\mathbf{p}_j, \mathbf{x})\,\mathbf{p}_j = B^\top B\mathbf{x} = \sum_{j=1}^r \xi_j B^\top \mathbf{q}_j = \sum_{j=1}^r \xi_j\,\sigma_j\,\mathbf{p}_j,$$

luego, por independencia lineal,

$$\xi_j = \sigma_j\,(\mathbf{p}_j, \mathbf{x}), \qquad j = 1, \ldots, r$$

y tenemos que

$$B\mathbf{x} = \sum_{j=1}^r \sigma_j(\mathbf{p}_j, \mathbf{x})\,\mathbf{q}_j = \sum_{j=1}^r \,\mathbf{q}_j\,\mathbf{p}_j^\top\mathbf{x}, \qquad \forall \mathbf{x} \in \mathbb{R}^m.$$

Esto equivale e que

$$B = \sum_{j=1}^r \sigma_j\mathbf{q}_j\mathbf{p}_j^\top,$$

esto es, hemos descompuesto B como una suma de matrices de rango unidad. Empaquetando la información en matrices

$$P := \left[\begin{array}{c|c|c} \mathbf{p}_1 & \dots & \mathbf{p}_r \end{array} \right], \qquad P^\top P = I_r$$

$$Q := \left[\begin{array}{c|c|c} \mathbf{q}_1 & \dots & \mathbf{q}_r \end{array} \right], \qquad Q^\top Q = I_r$$

(P es $m \times r$, mientras que Q es $n \times r$) y

$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

hemos factorizado B en la forma

$$B = Q\Sigma P^\top.$$

A esta factorización de B se le llama **descomposición en valores singulares**. Se suele abreviar con el nombre inglés SVD.

Hay una serie de resultados que se obtienen inmediatamente: para la matriz $B^\top$ se tiene inmediatamente

$$B^\top = \sum_{j=1}^{r} \sigma_j \mathbf{p}_j \, \mathbf{q}_j^\top = P\Sigma Q^\top,$$

y para $B\,B^\top$

$$B\,B^\top = \sum_{j=1}^{r} \sigma_j^2 \mathbf{q}_j \, \mathbf{q}_j^\top.$$

Hay una matriz interesante, que es

$$B^\dagger := \sum_{j=1}^{n} (1/\sigma_j) \, \mathbf{p}_j \, \mathbf{q}_j^\top = P\Sigma^{-1} Q^\top.$$

Esta matriz se denomina **pseudoinversa** de B.

La compresión de información de B se realiza recortando la descomposición en valores singulares

$$B_k := \sum_{j=1}^{k} \sigma_j \mathbf{q}_j \, \mathbf{p}_j^\top.$$

Notemos que

$$|B\mathbf{x}|^2 = \sum_{j=1}^{r} \sigma_j^2 |(\mathbf{p}_j, \mathbf{x})|^2$$

y que entonces

$$|B\mathbf{x} - B_k\mathbf{x}|^2 = \sum_{j=k+1}^{r} \sigma_j^2 |(\mathbf{p}_j, \mathbf{x})|^2 \leq \sigma_{k+1}^2 |\mathbf{x}|^2.$$

Hay otra versión de la descomposición en valores singulares, donde Σ tiene el mismo tamaño que B (en la esquina superior derecha de Σ se pone la matriz diagonal anterior) y P y Q son cuadradas (cada una de un tamaño). Las r primeras columnas de P y Q son como las dadas antes. Las restantes, se obtienen completando hasta una base ortonormal del espacio correspondiente.

Temas relacionados:

- problemas mal condicionados
- regularización
- filtros de Wiener
- análisis de componentes principales

by F.–J. Sayas
Mathematics, Ho!, March, 2006