Matemáticas para ciencia de datos

Olivia Gutú

Semana 4
Descomposición en valores singulares y análisis en componentes principales

Descomposición en valores singulares

SVD

Toda matriz (real) tiene la forma:

$$A = U\Sigma V^T$$

U y V son matrices ortogonales y Σ matriz diagonal de valores singulares

SVD y Teorema espectral

La SVD existe para *cualquier matriz*: ¡no tiene que ser cuadrada!

SVD: idea principal

Sea A una matriz de $m \times n$. La idea es (1) encontrar (recordar que rank $A =$ rank A^T):

- un conjunto ortonormal de vectores $\mathrm{v}_1, \mathrm{v}_2, \cdots, \mathrm{v}_\mathrm{r}$ de $\mathrm{im}(A^T)$
- un conjunto ortonormal de vectores $\mathrm{u}_1, \mathrm{u}_2, \cdots, \mathrm{u}_\mathrm{r}$ de $\mathrm{im}(A)$

tal que:

$$\begin{aligned} A\begin{pmatrix}\mathrm{v}_1 & \mathrm{v}_2 & \cdots & \mathrm{v}_\mathrm{r}\end{pmatrix} &= \begin{pmatrix}\sigma_1\mathrm{u}_1 & \sigma_2\mathrm{u}_2 & \cdots & \sigma_r\mathrm{u}_\mathrm{r}\end{pmatrix} \\ &= \begin{pmatrix}\mathrm{u}_1 & \mathrm{u}_2 & \cdots & \mathrm{u}_\mathrm{r}\end{pmatrix}\begin{pmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{pmatrix}\end{aligned}$$

con

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$$

SVD: idea principal

(2) completar las bases de $\mathrm{im}(A^T)$ con el $\ker(A)$ y $\mathrm{im}(A)$ con el $\ker(A^T)$ (se completa la matriz diagonal con ceros). (1) y (2) implican $AV = U\Sigma$:

$$
\begin{aligned}
V &= \begin{pmatrix} \mathrm{v}_1 & \mathrm{v}_2 & \cdots & \mathrm{v}_\mathrm{r} & \mathrm{v}_{\mathrm{r}+1} \cdots & \mathrm{v}_\mathrm{n} \end{pmatrix} \\
U &= \begin{pmatrix} \mathrm{u}_1 & \mathrm{u}_2 & \cdots & \mathrm{u}_\mathrm{r} & \mathrm{u}_{\mathrm{r}+1} & \cdots & \mathrm{u}_\mathrm{m} \end{pmatrix} \\
\Sigma &= \left(\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & 0 \\ & & \ddots & & \\ & & & \sigma_r & \\ \hline & & 0 & & 0 \end{array}\right)
\end{aligned}
$$

por tanto:

$$A = U\Sigma V^T$$

SVD: ¿cómo funciona?

Notar que:

$$\text{rank } A^T A = \text{rank } AA^T = \text{rank } A = \text{rank } A^T$$

y

el rango de la matriz $A^T A$ es el número de autovalores distintos de cero (contando multiplicidad)

SVD: ¿cómo funciona?

- $\lambda_1, \lambda_2 \dots \lambda_r$ autovalores diferentes de cero de $A^T A$
- $v_1, v_2, \dots v_r$ conjunto ortonormal de autovectores de $A^T A$ correspondientes a los autovalores diferentes de cero.

$$\sigma_i := \sqrt{\lambda_i} \qquad \text{la raíz positiva}$$

Esto existe por el teorema espectral ya que $A^T A$ es simétrica.

SVD: ¿cómo funciona?

Para $i = 1, 2, \dots r$:

- se tiene $A^T A v_i = \sigma_i^2 v_i$
- notar que v_i está en $\mathrm{im}(A^T)$
- se define $u_i := \frac{Av_i}{\sigma_i}$ (luego $AA^T u_i = \sigma_i^2 u_i$ y por tanto u_i está en $\mathrm{im}(\mathrm{A})$)
- se cumple $u_i \cdot u_j = 0$ si $i \neq j$ y $u_i \cdot u_i = 1$

Después:

- completamos los últimos $n - r$ vectores v_{r+1} a v_n
- completamos los últimos $m - r$ vectores u_{r+1} a v_m
- los nuevos vectores pertenencen al $\ker A$ y $\ker A^T$, respectivamente

Forma reducida de la SVD

El corazón de la SVD es $AV_r = U_r\Sigma_r$:

$$A\begin{pmatrix} \mathrm{v}_1 & \cdots & \mathrm{v}_r \\ \multicolumn{3}{c}{\text{base del espacio}} \\ \multicolumn{3}{c}{\text{renglón de A}} \end{pmatrix} = \begin{pmatrix} \mathrm{u}_1 & \cdots & \mathrm{u}_r \\ \multicolumn{3}{c}{\text{base del espacio}} \\ \multicolumn{3}{c}{\text{columna de A}} \end{pmatrix}\begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{pmatrix}$$

$$\begin{aligned} \mathrm{im}A^T &= \mathrm{span}\{\text{renglones de } A\} \\ \mathrm{im}A &= \mathrm{span}\{\text{columnas de } A\} \end{aligned}$$

Dato importante para los científicos de datos

La matriz de rango k

$$A_k := \sigma_1 \mathbf{u_1} \mathbf{v_1}^T + \cdots + \sigma_k \mathbf{u_k} \mathbf{v_k}^T$$

es la que *mejor aproxima* a A con rango k:

Eckart-Young

Si B es una matriz de rango k entonces

$$\|A - A_k\| \leq \|A - B\|.$$

Dato importante para los cientíticos de datos

Norma espectral $\|A\|_2 = \max \frac{\|Ax\|}{\|x\|} = \sigma_1$

Norma de Frobenius $\|A\|_F = \sqrt{\sigma_1^2 + \cdots \sigma_r^2}$

Norma nuclear $\|A\|_N = \sigma_1 + \sigma_2 + \cdots + \sigma_r$

Análisis en componentes principales (PCA)

PCA resulta de aplicar la SVD a una matriz A_{centered} de datos centrada con n columnas (*variables*) y m renglones (*samples*)

Componentes principales

$$\sigma_1 u_1, \sigma_2 u_2, \cdots, \sigma_r u_r \quad \in \mathbb{R}^m$$

Análisis en componentes principales (PCA)

Variación total

$$T = \frac{\sigma_1^2 + \cdots + \sigma_r^2}{m-1}$$

Los σ_i^2 son los autovalores distintos de cero de la matriz de covarianza

$$A_{\text{centered}}^T A_{\text{centered}}$$

La componente principal $\sigma_i u_i$ *explica* la fracción σ_i^2/T de la variación total. Cada valor singular hace su mejor captura: σ_1 explica más que σ_2 y así sucesivamente.