Lecture 8 PCA and low rank approximation

References : James et al. Section 12.2

Assume that the data matrix X is represented via its thin SVD as

$$X = U_r \Sigma_r V_r^T$$

where $U_r \in \mathbb{R}^{m \times r}$ with orthonormal columns

$\Sigma_r \in \mathbb{R}^{r \times r}$ with diagonal entries that are the first r singular values of A

$V_r \in \mathbb{R}^{n \times r}$ with orthonormal columns.

the data matrix $X = [\vec{x}_1 \vdots \vec{x}_2 \vdots \cdots \vdots \vec{x}_n]$.

We can express each data vector $\vec{x}_i$ $(i = 1, \cdots, n)$ as

$$\vec{x}_i = \sum_{j=1}^{r} c_i^j \vec{u}_j$$

where $\vec{u}_j$ $(j = 1, \cdots, r)$ are the orthonormal columns of U_r (singular vectors), also called the feature vectors.

The scalars c_i^j are called the principal components of $\vec{x}_i$.

Note thet

$$\underset{(m\times1)}{\vec{x}_i} = \underset{(m\times n)}{X}\,\underset{(n\times1)}{\vec{e}_i} = U_r \Sigma_r V_r^T \vec{e}_i \equiv U_r \vec{c}_i \qquad i=1,\cdots,n$$

where $$\vec{c}_i = \underset{(r\times r)}{\Sigma_r}\,\underset{(r\times n)}{V_r^T}\,\underset{(n\times 1)}{\vec{e}_i} = \begin{bmatrix} \sigma_1 V_{i1} \\ \sigma_2 V_{i2} \\ \vdots \\ \sigma_r V_{ir} \end{bmatrix}_{r\times 1} \equiv [C_i^j]$$

Therefore, we have

$$\vec{x}_i = U_r \vec{c}_i \qquad i=1,\cdots,n$$

$$= \underset{(m\times r)}{[\vec{u}_1 \vdots \vec{u}_2 \vdots \cdots \vdots \vec{u}_r]} \begin{bmatrix} c_i^1 \\ c_i^2 \\ \vdots \\ c_i^r \end{bmatrix} \quad (r\times1)$$

$$\Rightarrow \quad \vec{x}_i = \sum_{j=1}^{r} c_i^j \vec{u}_j \qquad i=1,\cdots,n$$

We define the principal component metrix as

$$C = [C_i^j]_{r\times n} = [\vec{c}_1 \vdots \vec{c}_2 \vdots \cdots \vdots \vec{c}_n]_{(r\times n)}$$

Then the reletion between the principal compoment metrix and the deta metrix is

$$\underset{(m\times n)}{X} = \underset{(m\times r)}{U_r}\,\underset{(r\times n)}{C}$$

since U_r is has orthonormal columns, $U_r^T U_r = I$

$$\Rightarrow \quad C = U_r^T X$$

Example: See Jupyter notebook. (using X from previous lecture)

With $m = 3$, $n = 6$, $r = 2$.

$$\underset{(3\times 6)}{X} = \underset{(3\times 2)}{U_r}\ \underset{(2\times 2)}{\Sigma_r}\ \underset{(2\times 6)}{V_r^T}$$

The principal component matrix

$$\underset{(2\times 6)}{C} = \underset{(2\times 3)}{U_r^T}\ \underset{(3\times 6)}{X}$$

Verify: $\underset{(3\times 6)}{X} = \underset{(3\times 2)}{U_r}\ \underset{(2\times 6)}{C}$

Recall the full SVD of $A \in \mathbb{R}^{m\times n}$

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

$U \in \mathbb{R}^{m\times m}$ orthogonal

$V \in \mathbb{R}^{n\times n}$ orthogonal.

$$\Sigma = \begin{bmatrix} \hat{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{where } \hat{\Sigma} = \mathrm{diag}\{\sigma_1, \cdots, \sigma_r\}$$

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

$\mathrm{Rank}(A) = \mathrm{Rank}(\Sigma) = r$, the number of nonzero singular values

Now we write Σ as a sum of r matrices Σ_j

where $\Sigma_j = \mathrm{diag}(\{0, 0, \cdots, \sigma_j, 0, \cdots 0\})$

Then

$$A = \begin{bmatrix} u_1 & \cdots & u_n & \cdots & u_m \end{bmatrix} \Bigg(\underbrace{\begin{bmatrix} \sigma_1 & & \\ & 0 & \\ & & \ddots \\ & & & 0 \end{bmatrix} + \begin{bmatrix} 0 & & \\ & \sigma_2 & \\ & & \ddots \\ & & & 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}}_{r \text{ matrices}} \Bigg) \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}$$

$$= \sum_{j=1}^{r} \sigma_j u_j v_j^T$$

<u>Theorem</u>: Given $A = U \Sigma V^T$,

A is the sum of r rank-one matrices.

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^T$$

Here, $u_j v_j^T$ is the outer product of u_j, v_j

and $u_j v_j^T$ gives a rank-one matrix.

§ Low rank approximation.

$$A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^T \qquad k \leq r$$

then $$\|A - A_k\|_2 = \inf_{\substack{B \in \mathbb{R}^{m\times n} \\ \text{rank}(B) \leq k}} \|A - B\|_2 = \sigma_{k+1}$$

A_k gives the best approximation of the matrix A by matrices of rank lower than or equal to k in the least squares sense

See Jupyter notebook for an example in image compression.