Linear Algebra II
SVD and PCA

Ming-Hsuan Kang

Projection onto Best-fit Subspaces

Let A be an $m \times n$ matrix with singular values $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ and a set of left singular vectors $\{\vec{v}_1, \cdots, \vec{v}_n\}$ and the corresponding right singular vectors $\{\vec{u}_1, \cdots, \vec{u}_m\}$.

Recall that for the row vectors $\vec{x}_1^t, \cdots, \vec{x}_m^t$ of A, $V_k = \mathrm{span}\{\vec{v}_1, \cdots, \vec{v}_k\}$ is a best-fit k-subspace of $\{\vec{x}_1, \cdots, \vec{x}_m\}$. Note that

$$\begin{pmatrix} \mathrm{proj}_{\vec{v}_i}(\vec{x}_1)^t \\ \vdots \\ \mathrm{proj}_{\vec{v}_i}(\vec{x}_m)^t \end{pmatrix} = \begin{pmatrix} \vec{x}_1^t \vec{v}_i \vec{v}_i^t \\ \vdots \\ \vec{x}_m^t \vec{v}_i \vec{v}_i^t \end{pmatrix} = \begin{pmatrix} \vec{x}_1^t \vec{v}_i \vec{v}_i^t \\ \vdots \\ \vec{x}_m^t \vec{v}_i \vec{v}_i^t \end{pmatrix} = A\vec{v}_i \vec{v}_i^t = \sqrt{\lambda_i}\vec{u}_i \vec{v}_i^t.$$

Therefore,

$$\begin{pmatrix} \mathrm{proj}_{V_k}(\vec{x}_1)^t \\ \vdots \\ \mathrm{proj}_{V_k}(\vec{x}_n)^t \end{pmatrix} = \sum_{i=1}^{k} \begin{pmatrix} \mathrm{proj}_{\vec{v}_i}(\vec{x}_1)^t \\ \vdots \\ \mathrm{proj}_{\vec{v}_i}(\vec{x}_n)^t \end{pmatrix} = \sum_{i=1}^{k} \sqrt{\lambda_i}\vec{u}_i \vec{v}_i^t.$$

Projection onto Best-fit Subspace

Theorem

Let A be an $m \times n$ matrix with singular values $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ and a set of left singular vectors $\{\vec{v}_1, \cdots, \vec{v}_n\}$ and the corresponding right singular vectors $\{\vec{u}_1, \cdots, \vec{u}_m\}$. Then for the row vectors $\vec{x}_1^t, \cdots, \vec{x}_m^t$ of A, the projection of these row vectors onto the best-fit k-subspace is given by

$$A(k) = \begin{pmatrix} \mathrm{proj}_{V_k}(\vec{x}_1)^t \\ \vdots \\ \mathrm{proj}_{V_k}(\vec{x}_m)^t \end{pmatrix} = \sum_{i=1}^{k} \sqrt{\lambda_i}\vec{u}_i\vec{v}_i^t.$$

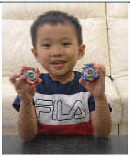
Note that $A(k)$ is an approximation of A and $A(r) = A$ where r is the rank of A. Moreover, we only need use $k(m+n)$ real numbers to characterize the matrix $A(r)$.

Image Compression

Give an image of size $m \times n$, let A_R, A_G and A_B be the RGB matrices of the image. For instance, the (i, j)-entry of A_R is the red number of the (i, j) pixel. In this case, we need use $3mn$ real numbers to store the image. Apply the previous theorem, we obtain three matrices $A_R(k)$, $A_G(k)$ and $A_B(k)$, which gives us a new image using only $3k(m + n)$ real numbers.

Example

The right is a picture of size $m \times n = 380 \times 314$. Consider the following different k.

k	10	30	50	80
C.R.	5.82%	17.4%	29.1%	46.5%
				

Here $C.R.$ is the compression rate which equals to $k(m+n)/mn$.

Best-fit affine Subspace

Given an $m \times n$ real matrix A, let $\{\vec{x}_1^t, \cdots, \vec{x}_m^t\}$ be the rows of A, which is a subset of $\mathbb{R}^n$. A k-dimensional subspace W shifted by some vector $\vec{w}_0$ is called an affine k-subspace, denoted by $\vec{w}_0 + W$.

A k-dimensional affine subspace $\vec{w}_0 + W$ in $\mathbb{R}^n$ is called a best-fit affine k-subspace of $\{\vec{x}_1, \cdots, \vec{x}_m\}$, if

$$\vec{w}_0 + W \in \underset{\vec{w}+W':dimW'=k}{\arg\min} \left\{ \sum_{i=1}^{m} \|\vec{(}x_i - \vec{w}) - \mathrm{proj}_{W'}(\vec{x}_i - \vec{w})\|^2 \right\}.$$

Note that

$$\|\vec{(}x_i - \vec{w}) - \mathrm{proj}_{W'}(\vec{x_i} - \vec{w})\|^2 = \|\mathrm{proj}_{W'^\perp}(\vec{x_i} - \vec{w})\|^2.$$

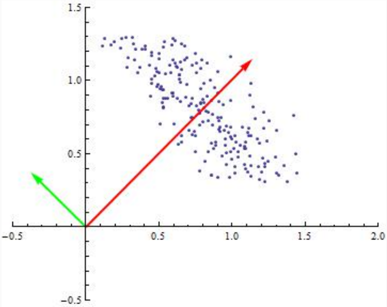
It is easy to see that when W' is fixed, $\sum_{i=1}^m \|\mathrm{proj}_{W'^\perp}(\vec{x_i}) - \mathrm{proj}_{W'^\perp}(\vec{w})\|^2$ is minimal when $\mathrm{proj}_{W'^\perp}(\vec{w})$ is the mean vector of $\{\mathrm{proj}_{W'^\perp}(\vec{x_i})\}_{i=1}^m$. Especially, one can choose $\vec{w}$ as the mean vector of $\vec{x_i}$. Then $\vec{w}$ always gives minimal value for all W'.

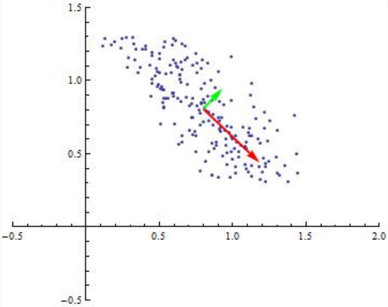
Theorem

Given an $m \times n$ real matrix A, let $\{\vec{x}_1^t, \cdots, \vec{x}_m^t\}$ be the rows of A and $\vec{x}_0^t$ be their mean vector. Let W be the best-fit k-subspace of $\{\vec{x}_1^t - \vec{x}_0^t, \cdots, \vec{x}_m^t - \vec{x}_0^t\}$. Then $\vec{x}_0 + W$ is the best-fit affine k-subspace.

Under the notation in the previous theorem, the i-th singular vectors of the matrix with rows $\{\vec{x}_1^t - \vec{x}_0^t, \cdots, \vec{x}_m^t - \vec{x}_0^t\}$ are called i-th principal component of $\{\vec{x}_1^t, \cdots, \vec{x}_m^t\}$.

SVD v.s. PCA





Left Singular Vectors

Principal Components

Example

Let $\{(3,2),(2,2),(0,4),(1,1),(1,1)\}$ be a set of points in $\mathbb{R}^2$. Find its first principal component.

The mean vector of the set is $(1,2)$ and the set of points shifted by $(-1,-2)$ becomes $\{(2,0),(-1,0),(-1,2),(0,-1),(0,-1)\}$. Let $A = \begin{pmatrix} 2 & 0 \\ -1 & 0 \\ -1 & 2 \\ 0 & -1 \\ 0 & -1 \end{pmatrix}$. Then $A^tA = \begin{pmatrix} 6 & -2 \\ -2 & 6 \end{pmatrix}$ which eigenvalues are 8 and 4. Let $(1,-1)^t$ be the eigenvector of A^tA corresponding to 8. Then $\frac{1}{\sqrt{2}}(1,-1)$ is the first principal component of the set.