From SVD to PCA

Xin Li

April 6, 2009

Recall the SVD

- Given a matrix $\mathbf{X}_{m \times n}$ of rank n.

Recall the SVD

- Given a matrix $\mathbf{X}_{m\times n}$ of rank n.
- We can find orthogonal matrices $\mathbf{U}_{m\times n}$ and $\mathbf{V}_{n\times n}$,

Recall the SVD

- Given a matrix $\mathbf{X}_{m \times n}$ of rank n.
- We can find orthogonal matrices $\mathbf{U}_{m \times n}$ and $\mathbf{V}_{n \times n}$,
- and diagonal matrix $\mathbf{D}_{n \times n}$:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

Recall the SVD

- Given a matrix $\mathbf{X}_{m\times n}$ of rank n.
- We can find orthogonal matrices $\mathbf{U}_{m\times n}$ and $\mathbf{V}_{n\times n}$,
- and diagonal matrix $\mathbf{D}_{n\times n}$:

$$\mathbf{X} = \mathbf{UDV}^T$$

- Visualize it as:

$$\begin{bmatrix} \\ \mathbf{X} \\ \\ \end{bmatrix}_{m\times n} = \begin{bmatrix} \\ \mathbf{U} \\ \\ \end{bmatrix}_{m\times n} \begin{bmatrix} \diagdown \end{bmatrix}_{n\times n} \begin{bmatrix} \mathbf{V}^T \end{bmatrix}_{n\times n}$$

Recall the SVD

- Given a matrix $\mathbf{X}_{m\times n}$ of rank n.
- We can find orthogonal matrices $\mathbf{U}_{m\times n}$ and $\mathbf{V}_{n\times n}$,
- and diagonal matrix $\mathbf{D}_{n\times n}$:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

- Visualize it as:

$$\begin{bmatrix} \\ \mathbf{X} \\ \\ \end{bmatrix}_{m\times n} = \begin{bmatrix} \\ \mathbf{U} \\ \\ \end{bmatrix}_{m\times n} \begin{bmatrix} \diagdown \end{bmatrix}_{n\times n} \begin{bmatrix} \mathbf{V}^T \end{bmatrix}_{n\times n}$$

- Now we drive SVD when the rank of $\mathbf{X}$ is $r < n$.

Deriving the SVD

- Find the eigen decomposition of $\mathbf{X}^T\mathbf{X}$:

$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\Lambda\mathbf{V}^T$$

where Λ is the diagonal matrix with the eigenvalues λ_i's of $\mathbf{X}^T\mathbf{X}$ on its diagonal.

Deriving the SVD

- Find the eigen decomposition of $\mathbf{X}^T\mathbf{X}$:

$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\Lambda\mathbf{V}^T$$

  where Λ is the diagonal matrix with the eigenvalues λ_i's of $\mathbf{X}^T\mathbf{X}$ on its diagonal.
- The matrix $\mathbf{V}$ is orthogonal whose columns $\mathbf{v}_i$'s are eigenvectors of $\mathbf{X}^T\mathbf{X}$: $\quad \mathbf{X}^T\mathbf{X}\mathbf{v}_i = \lambda_i\mathbf{v}_i, \quad i = 1:n$

Deriving the SVD

- Find the eigen decomposition of $\mathbf{X}^T\mathbf{X}$:

$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\Lambda\mathbf{V}^T$$

  where Λ is the diagonal matrix with the eigenvalues λ_i's of $\mathbf{X}^T\mathbf{X}$ on its diagonal.
- The matrix $\mathbf{V}$ is orthogonal whose columns $\mathbf{v}_i$'s are eigenvectors of $\mathbf{X}^T\mathbf{X}$: $\quad \mathbf{X}^T\mathbf{X}\mathbf{v}_i = \lambda_i\mathbf{v}_i, \quad i = 1 : n$
- It is also known that $\lambda_i \geq 0$ (Homework: Verify it!). If the rank of $\mathbf{X}$ is r ($r \leq n$), there are exactly r positive eigenvalues, say

$$\lambda_1, \lambda_2, ..., \lambda_r$$

Deriving the SVD

- Find the eigen decomposition of $\mathbf{X}^T\mathbf{X}$:
$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\Lambda\mathbf{V}^T$$
where Λ is the diagonal matrix with the eigenvalues λ_i's of $\mathbf{X}^T\mathbf{X}$ on its diagonal.
- The matrix $\mathbf{V}$ is orthogonal whose columns $\mathbf{v}_i$'s are eigenvectors of $\mathbf{X}^T\mathbf{X}$: $\quad \mathbf{X}^T\mathbf{X}\mathbf{v}_i = \lambda_i\mathbf{v}_i, \quad i = 1 : n$
- It is also known that $\lambda_i \geq 0$ (Homework: Verify it!). If the rank of $\mathbf{X}$ is r ($r \leq n$), there are exactly r positive eigenvalues, say
$$\lambda_1, \lambda_2, ..., \lambda_r$$
- For $i = 1 : r$, define $\sigma_i = \sqrt{\lambda_i}$ and $\mathbf{u}_i = X\mathbf{v}_i/\sigma_i$.

Deriving the SVD

- Find the eigen decomposition of $\mathbf{X}^T\mathbf{X}$:
$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\Lambda\mathbf{V}^T$$
where Λ is the diagonal matrix with the eigenvalues λ_i's of $\mathbf{X}^T\mathbf{X}$ on its diagonal.
- The matrix $\mathbf{V}$ is orthogonal whose columns $\mathbf{v}_i$'s are eigenvectors of $\mathbf{X}^T\mathbf{X}$: $\quad \mathbf{X}^T\mathbf{X}\mathbf{v}_i = \lambda_i\mathbf{v}_i, \quad i = 1:n$
- It is also known that $\lambda_i \geq 0$ (Homework: Verify it!). If the rank of $\mathbf{X}$ is r ($r \leq n$), there are exactly r positive eigenvalues, say
$$\lambda_1, \lambda_2, ..., \lambda_r$$
- For $i = 1:r$, define $\sigma_i = \sqrt{\lambda_i}$ and $\mathbf{u}_i = X\mathbf{v}_i/\sigma_i$.
- Then
$$\mathbf{X}^T\mathbf{u}_i = \sigma_i\mathbf{v}_i$$
$$\mathbf{X}\mathbf{v}_i = \sigma_i\mathbf{u}_i$$

Deriving the SVD, continued

- Put in matrix notation:

$$\mathbf{X}[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_r] = [\sigma_1\mathbf{u}_1\ \sigma_2\mathbf{u}_2\ \cdots\ \sigma_r\mathbf{u}_r]$$

$$\mathbf{X}^T[\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_r] = [\sigma_1\mathbf{v}_1\ \sigma_2\mathbf{v}_2\ \cdots\ \sigma_r\mathbf{v}_r]$$

Deriving the SVD, continued

- Put in matrix notation:

$$\mathbf{X}[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_r] = [\sigma_1\mathbf{u}_1\ \sigma_2\mathbf{u}_2\ \cdots\ \sigma_r\mathbf{u}_r]$$

$$\mathbf{X}^T[\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_r] = [\sigma_1\mathbf{v}_1\ \sigma_2\mathbf{v}_2\ \cdots\ \sigma_r\mathbf{v}_r]$$

- Note that $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$ is an orthogonal set in $\mathbb{R}^n$. We can extend it to an orthogonal basis.

Deriving the SVD, continued

- Put in matrix notation:

$$\mathbf{X}[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_r] = [\sigma_1\mathbf{u}_1\ \sigma_2\mathbf{u}_2\ \cdots\ \sigma_r\mathbf{u}_r]$$

$$\mathbf{X}^T[\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_r] = [\sigma_1\mathbf{v}_1\ \sigma_2\mathbf{v}_2\ \cdots\ \sigma_r\mathbf{v}_r]$$

- Note that $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$ is an orthogonal set in $\mathbb{R}^n$. We can extend it to an orthogonal basis.
- Solve $\quad [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_r]\mathbf{v} = \mathbf{0}$
  (What is the dimension of the solution space of this homogeneous system?)

Deriving the SVD, continued

- Put in matrix notation:

$$\mathbf{X}[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_r] = [\sigma_1\mathbf{u}_1\ \sigma_2\mathbf{u}_2\ \cdots\ \sigma_r\mathbf{u}_r]$$

$$\mathbf{X}^T[\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_r] = [\sigma_1\mathbf{v}_1\ \sigma_2\mathbf{v}_2\ \cdots\ \sigma_r\mathbf{v}_r]$$

- Note that $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$ is an orthogonal set in $\mathbb{R}^n$. We can extend it to an orthogonal basis.
- Solve $\qquad [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_r]\mathbf{v} = \mathbf{0}$
(What is the dimension of the solution space of this homogeneous system?)
- Answer: Since the coefficient matrix $\mathbf{V}_1 := [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_r]$ has rank r (Why?), the dimension of the solution space is $n - r$.

Deriving the SVD, continued

- Put in matrix notation:

$$\mathbf{X}[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_r] = [\sigma_1\mathbf{u}_1\ \sigma_2\mathbf{u}_2\ \cdots\ \sigma_r\mathbf{u}_r]$$

$$\mathbf{X}^T[\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_r] = [\sigma_1\mathbf{v}_1\ \sigma_2\mathbf{v}_2\ \cdots\ \sigma_r\mathbf{v}_r]$$

- Note that $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$ is an orthogonal set in $\mathbb{R}^n$. We can extend it to an orthogonal basis.
- Solve $\quad [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_r]\mathbf{v} = \mathbf{0}$
  (What is the dimension of the solution space of this homogeneous system?)
- Answer: Since the coefficient matrix $\mathbf{V}_1 := [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_r]$ has rank r (Why?), the dimension of the solution space is $n - r$.
- So, there are exactly $n - r$ orthogonal vectors $\mathbf{v}_{r+1}$, $\mathbf{v}_{r+2}$, ... , $\mathbf{v}_n$ such that $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_n$ form an orthogonal basis of $\mathbb{R}^n$.

Deriving the SVD, continued

- Use the extended basis:

$$\mathbf{X}[\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n] = [\sigma_1\mathbf{u}_1 \ \sigma_2\mathbf{u}_2 \ \cdots \ \sigma_r\mathbf{u}_r \ \mathbf{X}\mathbf{v}_{r+1} \ \cdots \ \mathbf{X}\mathbf{v}_n]$$

Deriving the SVD, continued

- Use the extended basis:

$$\mathbf{X}[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n] = [\sigma_1\mathbf{u}_1\ \sigma_2\mathbf{u}_2\ \cdots\ \sigma_r\mathbf{u}_r\ \mathbf{X}\mathbf{v}_{r+1}\ \cdots\ \mathbf{X}\mathbf{v}_n]$$

- Claim: $\mathbf{X}\mathbf{v}_j = \mathbf{0}$, $j = r+1 : n$ (Homework: Verify it!).

Deriving the SVD, continued

- Use the extended basis:

$$\mathbf{X}[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n] = [\sigma_1\mathbf{u}_1\ \sigma_2\mathbf{u}_2\ \cdots\ \sigma_r\mathbf{u}_r\ \mathbf{X}\mathbf{v}_{r+1}\ \cdots\ \mathbf{X}\mathbf{v}_n]$$

- Claim: $\mathbf{X}\mathbf{v}_j = \mathbf{0}$, $j = r+1:n$ (Homework: Verify it!).
- Thus,

$$\mathbf{X}[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n] = [\sigma_1\mathbf{u}_1\ \sigma_2\mathbf{u}_2\ \cdots\ \sigma_r\mathbf{u}_r\ \mathbf{0}\ \cdots\ \mathbf{0}]$$

Deriving the SVD, continued

- Use the extended basis:

$$\mathbf{X}[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n] = [\sigma_1\mathbf{u}_1\ \sigma_2\mathbf{u}_2\ \cdots\ \sigma_r\mathbf{u}_r\ \mathbf{X}\mathbf{v}_{r+1}\ \cdots\ \mathbf{X}\mathbf{v}_n]$$

- Claim: $\mathbf{X}\mathbf{v}_j = \mathbf{0}$, $j = r+1 : n$ (Homework: Verify it!).
- Thus,

$$\mathbf{X}[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n] = [\sigma_1\mathbf{u}_1\ \sigma_2\mathbf{u}_2\ \cdots\ \sigma_r\mathbf{u}_r\ \mathbf{0}\ \cdots\ \mathbf{0}]$$

- Let $\mathbf{V} := [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n]$ and $D = \mathrm{diag}(\sigma_1, ..., \sigma_r)$.

Deriving the SVD, continued

- Use the extended basis:

$$\mathbf{X}[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n] = [\sigma_1\mathbf{u}_1\ \sigma_2\mathbf{u}_2\ \cdots\ \sigma_r\mathbf{u}_r\ \mathbf{X}\mathbf{v}_{r+1}\ \cdots\ \mathbf{X}\mathbf{v}_n]$$

- Claim: $\mathbf{X}\mathbf{v}_j = \mathbf{0}$, $j = r+1 : n$ (Homework: Verify it!).
- Thus,

$$\mathbf{X}[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n] = [\sigma_1\mathbf{u}_1\ \sigma_2\mathbf{u}_2\ \cdots\ \sigma_r\mathbf{u}_r\ \mathbf{0}\ \cdots\ \mathbf{0}]$$

- Let $\mathbf{V} := [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n]$ and $D = \mathrm{diag}(\sigma_1, ..., \sigma_r)$.
- So, $\mathbf{XV} = [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_r][D \vdots \mathbf{0}_{r\times(n-r)}]$. Thus

$$\mathbf{X} = [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_r][D \vdots \mathbf{0}_{r\times(n-r)}]\mathbf{V}^T$$

Deriving the SVD, continued

- Finally, to make the $\mathbf{u}_i$ matrix a square matrix, we extend it an orthogonal basis: $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$.

Deriving the SVD, continued

- Finally, to make the $\mathbf{u}_i$ matrix a square matrix, we extend it an orthogonal basis: $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m$.
- Define $U := [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_m]$ and fill 0's under $[D \vdots \mathbf{0}_{r\times(n-r)}]$:

$$\mathbf{X} = \mathbf{U} \begin{bmatrix} D \vdots \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{m-r\times n} \end{bmatrix} \mathbf{V}^T$$

Deriving the SVD, continued

- Finally, to make the $\mathbf{u}_i$ matrix a square matrix, we extend it an orthogonal basis: $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m$.
- Define $U := [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_m]$ and fill 0's under $[D \vdots \mathbf{0}_{r\times(n-r)}]$:

$$\mathbf{X} = \mathbf{U} \begin{bmatrix} D \vdots \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{m-r\times n} \end{bmatrix} \mathbf{V}^T$$

- We are done!

Back to column vector form

- Put **U** and **V** back to column vector form in SVD:

$$\mathbf{X} = \mathbf{U} \left[\begin{array}{c} D \vdots \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{m-r\times n} \end{array} \right] \mathbf{V}^T$$

Back to column vector form

- Put **U** and **V** back to column vector form in SVD:

$$\mathbf{X} = \mathbf{U}\begin{bmatrix} D \vdots \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{m-r\times n} \end{bmatrix}\mathbf{V}^T$$

- we have

$$\mathbf{X} = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_m]\begin{bmatrix} D \vdots \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{m-r\times n} \end{bmatrix}\begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}$$

Back to column vector form

- Put **U** and **V** back to column vector form in SVD:

$$\mathbf{X} = \mathbf{U} \begin{bmatrix} D \vdots \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{m-r\times n} \end{bmatrix} \mathbf{V}^T$$

- we have

$$\mathbf{X} = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_m] \begin{bmatrix} D \vdots \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{m-r\times n} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}$$

- So,

$$\mathbf{X} = [\sigma_1 \mathbf{u}_1 \ \cdots \ \sigma_r \mathbf{u}_r \ \mathbf{0} \ \cdots \ \mathbf{0}] \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

Dimension reduction

- Assume we have a collection data of column vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ of $\mathbb{R}^n$.

Dimension reduction

- Assume we have a collection data of column vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ of $\mathbb{R}^n$.
- When n is large, we want "reduce" the dimension the data.

Dimension reduction

- Assume we have a collection data of column vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ of $\mathbb{R}^n$.
- When n is large, we want "reduce" the dimension the data.
- How could this be possible?

Dimension reduction

- Assume we have a collection data of column vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ of $\mathbb{R}^n$.
- When n is large, we want "reduce" the dimension the data.
- How could this be possible?
- This is possible in the case that all data vectors are actually coming from a subspace U of small dimension (say, r).

Dimension reduction

- Assume we have a collection data of column vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ of $\mathbb{R}^n$.
- When n is large, we want "reduce" the dimension the data.
- How could this be possible?
- This is possible in the case that all data vectors are actually coming from a subspace U of small dimension (say, r).
- Find an orthogonal set of vectors $\mathbf{u}_1, \ldots, \mathbf{u}_r$ that forms a basis for U!

Dimension reduction

- Assume we have a collection data of column vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ of $\mathbb{R}^n$.
- When n is large, we want "reduce" the dimension the data.
- How could this be possible?
- This is possible in the case that all data vectors are actually coming from a subspace U of small dimension (say, r).
- Find an orthogonal set of vectors $\mathbf{u}_1$, ... , $\mathbf{u}_r$ that forms a basis for U!
- If we can do so, then

$$\mathbf{a}_i = f_{i1}\mathbf{u}_1 + f_{i2}\mathbf{u}_2 + \cdots + f_{ir}\mathbf{u}_r, \quad i = 1 : m$$

for some coefficients f_{ij}'s.

Dimension reduction

- Assume we have a collection data of column vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ of $\mathbb{R}^n$.
- When n is large, we want "reduce" the dimension the data.
- How could this be possible?
- This is possible in the case that all data vectors are actually coming from a subspace U of small dimension (say, r).
- Find an orthogonal set of vectors $\mathbf{u}_1$, ... , $\mathbf{u}_r$ that forms a basis for U!
- If we can do so, then

$$\mathbf{a}_i = f_{i1}\mathbf{u}_1 + f_{i2}\mathbf{u}_2 + \cdots + f_{ir}\mathbf{u}_r, \quad i = 1 : m$$

  for some coefficients f_{ij}'s.
- Note that $f_{ij} = \mathbf{a}_i^T \mathbf{u}_j$, the projection of $\mathbf{a}_i$ along the direction of $\mathbf{u}_j$.

How to find the first principal component?

- Let us find $\mathbf{u}_1$.

How to find the first principal component?

- Let us find $\mathbf{u}_1$.
- A good choice of $\mathbf{u}_1$: the data vectors have the greatest total projection along the direction of $\mathbf{u}_1$:

$$\max_{\|w\|=1} \sum_{i=1}^{m} |\mathbf{a}_i^T \mathbf{w}|^2$$

How to find the first principal component?

- Let us find $\mathbf{u}_1$.
- A good choice of $\mathbf{u}_1$: the data vectors have the greatest total projection along the direction of $\mathbf{u}_1$:

$$\max_{\|w\|=1} \sum_{i=1}^{m} |\mathbf{a}_i^T \mathbf{w}|^2$$

- Note that

$$\sum_{i=1}^{m} |\mathbf{a}_i^T \mathbf{w}|^2 = \left\|\mathbf{A}^T \mathbf{w}\right\|^2$$

$$= (\mathbf{A}^T \mathbf{w})^T (\mathbf{A}^T \mathbf{w}) = \mathbf{w}^T \mathbf{A}\mathbf{A}^T \mathbf{w}$$

How to find the first principal component?

- Let us find $\mathbf{u}_1$.
- A good choice of $\mathbf{u}_1$: the data vectors have the greatest total projection along the direction of $\mathbf{u}_1$:

$$\max_{\|w\|=1} \sum_{i=1}^{m} |\mathbf{a}_i^T \mathbf{w}|^2$$

- Note that

$$\sum_{i=1}^{m} |\mathbf{a}_i^T \mathbf{w}|^2 = \left\| \mathbf{A}^T \mathbf{w} \right\|^2$$

$$= (\mathbf{A}^T \mathbf{w})^T (\mathbf{A}^T \mathbf{w}) = \mathbf{w}^T \mathbf{A} \mathbf{A}^T \mathbf{w}$$

- So,

$$\max_{\|w\|=1} \sum_{i=1}^{m} |\mathbf{a}_i \mathbf{w}|^2 = \max_{\|w\|=1} \mathbf{w}^T (\mathbf{A}\mathbf{A}^T) \mathbf{w} = \max_{w \neq 0} \frac{\mathbf{w}^T (\mathbf{A}\mathbf{A}^T) \mathbf{w}}{\mathbf{w}^T \mathbf{w}}$$

Here comes the SVD

- Let us use SVD of **A**:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

Here comes the SVD

- Let us use SVD of $\mathbf{A}$:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

- Then,

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}\mathbf{D}\mathbf{V}^T\mathbf{V}\mathbf{D}^T\mathbf{U}^T = \mathbf{U}\mathbf{D}\mathbf{D}^T\mathbf{U}^T$$

Here comes the SVD

- Let us use SVD of $\mathbf{A}$:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

- Then,

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}\mathbf{D}\mathbf{V}^T\mathbf{V}\mathbf{D}^T\mathbf{U}^T = \mathbf{U}\mathbf{D}\mathbf{D}^T\mathbf{U}^T$$

- and

$$\frac{\mathbf{w}^T(\mathbf{A}\mathbf{A}^T)\mathbf{w}}{\mathbf{w}^T\mathbf{w}} = \frac{(\mathbf{U}^T\mathbf{w})^T\mathbf{D}\mathbf{D}^T(\mathbf{U}^T\mathbf{w})}{(\mathbf{U}^T\mathbf{w})^T(\mathbf{U}^T\mathbf{w})}$$

Here comes the SVD

- Let us use SVD of $\mathbf{A}$:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

- Then,

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}\mathbf{D}\mathbf{V}^T\mathbf{V}\mathbf{D}^T\mathbf{U}^T = \mathbf{U}\mathbf{D}\mathbf{D}^T\mathbf{U}^T$$

- and

$$\frac{\mathbf{w}^T(\mathbf{A}\mathbf{A}^T)\mathbf{w}}{\mathbf{w}^T\mathbf{w}} = \frac{(\mathbf{U}^T\mathbf{w})^T\mathbf{D}\mathbf{D}^T(\mathbf{U}^T\mathbf{w})}{(\mathbf{U}^T\mathbf{w})^T(\mathbf{U}^T\mathbf{w})}$$

- Let $\mathbf{x} := \mathbf{U}^T\mathbf{w}$ and notice that there are only r non-zero entries σ_1, ..., σ_r in $\mathbf{D}$.

Here comes the SVD

- Let us use SVD of $\mathbf{A}$:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

- Then,

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}\mathbf{D}\mathbf{V}^T\mathbf{V}\mathbf{D}^T\mathbf{U}^T = \mathbf{U}\mathbf{D}\mathbf{D}^T\mathbf{U}^T$$

- and

$$\frac{\mathbf{w}^T(\mathbf{A}\mathbf{A}^T)\mathbf{w}}{\mathbf{w}^T\mathbf{w}} = \frac{(\mathbf{U}^T\mathbf{w})^T\mathbf{D}\mathbf{D}^T(\mathbf{U}^T\mathbf{w})}{(\mathbf{U}^T\mathbf{w})^T(\mathbf{U}^T\mathbf{w})}$$

- Let $\mathbf{x} := \mathbf{U}^T\mathbf{w}$ and notice that there are only r non-zero entries σ_1, ..., σ_r in $\mathbf{D}$.
- We have

$$\frac{\mathbf{w}^T(\mathbf{A}\mathbf{A}^T)\mathbf{w}}{\mathbf{w}^T\mathbf{w}} = \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \cdots \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \cdots + x_m^2}$$

Finally...

- We have

$$\max_{w \neq 0} \frac{\mathbf{w}^T(\mathbf{A}\mathbf{A}^T)\mathbf{w}}{\mathbf{w}^T\mathbf{w}} = \max_{x \neq 0} \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \cdots \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \cdots + x_m^2}$$

Finally...

- We have

$$\max_{w\neq 0} \frac{\mathbf{w}^T(\mathbf{A}\mathbf{A}^T)\mathbf{w}}{\mathbf{w}^T\mathbf{w}} = \max_{x\neq 0} \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \cdots \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \cdots + x_m^2}$$

- Assume $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Then

$$\max_{x\neq 0} \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \cdots \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \cdots + x_m^2} = \sigma_1^2 = \lambda_1$$

This is the largest eigenvalue of A.

Finally...

- We have

$$\max_{w \neq 0} \frac{\mathbf{w}^T(\mathbf{A}\mathbf{A}^T)\mathbf{w}}{\mathbf{w}^T\mathbf{w}} = \max_{x \neq 0} \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \cdots \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \cdots + x_m^2}$$

- Assume $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Then

$$\max_{x \neq 0} \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \cdots \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \cdots + x_m^2} = \sigma_1^2 = \lambda_1$$

  This is the largest eigenvalue of A.
- A vector $\mathbf{x}$ that makes the maximum is $x_1 = 1$ and $x_i = 0$ for $i = 2 : m$

Finally...

- We have

$$\max_{w \neq 0} \frac{\mathbf{w}^T(\mathbf{A}\mathbf{A}^T)\mathbf{w}}{\mathbf{w}^T\mathbf{w}} = \max_{x \neq 0} \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \cdots \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \cdots + x_m^2}$$

- Assume $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Then

$$\max_{x \neq 0} \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \cdots \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \cdots + x_m^2} = \sigma_1^2 = \lambda_1$$

  This is the largest eigenvalue of A.
- A vector $\mathbf{x}$ that makes the maximum is $x_1 = 1$ and $x_i = 0$ for $i = 2 : m$
- This corresponds to $\mathbf{w} = \mathbf{U}\mathbf{x} = \mathbf{u}_1$.

Finally...

- We have

$$\max_{w \neq 0} \frac{\mathbf{w}^T(\mathbf{A}\mathbf{A}^T)\mathbf{w}}{\mathbf{w}^T\mathbf{w}} = \max_{x \neq 0} \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \cdots \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \cdots + x_m^2}$$

- Assume $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Then

$$\max_{x \neq 0} \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \cdots \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \cdots + x_m^2} = \sigma_1^2 = \lambda_1$$

  This is the largest eigenvalue of A.
- A vector $\mathbf{x}$ that makes the maximum is $x_1 = 1$ and $x_i = 0$ for $i = 2 : m$
- This corresponds to $\mathbf{w} = \mathbf{U}\mathbf{x} = \mathbf{u}_1$.
- So, the first principal component is indeed achieved by the first eigenvector $\mathbf{u}_1$ of $\mathbf{A}\mathbf{A}^T$

How about the 2nd, 3rd, ... principal components?

- Using the same idea, the second principal component is the direction: be orthogonal to the first and maximize the total projection:

$$\max_{\|w\|=1, w^T \mathbf{u}_1 = 0} \sum_{i=1}^{m} |\mathbf{a}_i^T \mathbf{w}|^2$$

How about the 2nd, 3rd, ... principal components?

- Using the same idea, the second principal component is the direction: be orthogonal to the first and maximize the total projection:

$$\max_{\|w\|=1, w^T\mathbf{u}_1=0} \sum_{i=1}^{m} |\mathbf{a}_i^T \mathbf{w}|^2$$

- This is the same to

$$\max_{w\neq 0, w^T\mathbf{u}_1=0} \frac{\mathbf{w}^T(\mathbf{A}\mathbf{A}^T)\mathbf{w}}{\mathbf{w}^T\mathbf{w}} = \max_{x\neq 0, x^T\mathbf{U}^T\mathbf{u}_1=0} \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \cdots \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \cdots + x_m^2}$$

$$= \max_{x\neq 0, x_1=0} \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \cdots \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \cdots + x_m^2} = \sigma_2^2 = \lambda_2$$

This is the largest eigenvalue of $\mathbf{A}\mathbf{A}^T$.

Finding the 2nd, 3rd, ... principal components

- A vector $\mathbf{x}$ that makes the maximum is $x_2 = 1$ and $x_i = 0$ for $i = 1, 3 : m$

Finding the 2nd, 3rd, ... principal components

- A vector $\mathbf{x}$ that makes the maximum is $x_2 = 1$ and $x_i = 0$ for $i = 1, 3 : m$
- This corresponds to $\mathbf{w} = \mathbf{U}\mathbf{x} = \mathbf{u}_2$.

Finding the 2nd, 3rd, ... principal components

- A vector $\mathbf{x}$ that makes the maximum is $x_2 = 1$ and $x_i = 0$ for $i = 1, 3 : m$
- This corresponds to $\mathbf{w} = \mathbf{U}\mathbf{x} = \mathbf{u}_2$.
- So, the second principal component is indeed achieved by the second eigenvector $\mathbf{u}_2$ of $\mathbf{A}\mathbf{A}^T$.

Finding the 2nd, 3rd, ... principal components

- A vector $\mathbf{x}$ that makes the maximum is $x_2 = 1$ and $x_i = 0$ for $i = 1, 3 : m$
- This corresponds to $\mathbf{w} = \mathbf{U}\mathbf{x} = \mathbf{u}_2$.
- So, the second principal component is indeed achieved by the second eigenvector $\mathbf{u}_2$ of $\mathbf{A}\mathbf{A}^T$.
- Now you see the pattern:
  $\mathbf{u}_i$ is the ith principal component.

Sample code

- Let us randomly choose 20 points on a line in space (3-dimension).

Sample code

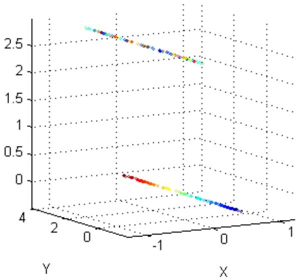
- Let us randomly choose 20 points on a line in space (3-dimension).
- p=100;r=rand(1,p);x=2*r-1;y=r+2;z=3-r; scatter3(x,y,z,3*ones(1,p),1:p);colorbar;view(-30,10);

Sample code

- Let us randomly choose 20 points on a line in space (3-dimension).
- ```
  p=100;r=rand(1,p);x=2*r-1;y=r+2;z=3-r;
  scatter3(x,y,z,3*ones(1,p),1:p);colorbar;view(-30,10);
  ```
- We want to find the first principal component.
  ```

Sample code

- Let us randomly choose 20 points on a line in space (3-dimension).
- p=100;r=rand(1,p);x=2*r-1;y=r+2;z=3-r; scatter3(x,y,z,3*ones(1,p),1:p);colorbar;view(-30,10);
- We want to find the first principal component.
- First, we center the data by remove the mean of all points.

Sample code

- Let us randomly choose 20 points on a line in space (3-dimension).
- p=100;r=rand(1,p);x=2*r-1;y=r+2;z=3-r;
  scatter3(x,y,z,3*ones(1,p),1:p);colorbar;view(-30,10);
- We want to find the first principal component.
- First, we center the data by remove the mean of all points.
- x=x-mean(x);y=y-mean(y);z=z-mean(z);hold on;scatter3(x,y,z,3*ones(1,p),1:p);

Sample code

- Let us randomly choose 20 points on a line in space (3-dimension).
- p=100;r=rand(1,p);x=2*r-1;y=r+2;z=3-r; scatter3(x,y,z,3*ones(1,p),1:p);colorbar;view(-30,10);
- We want to find the first principal component.
- First, we center the data by remove the mean of all points.
- x=x-mean(x);y=y-mean(y);z=z-mean(z);hold on;scatter3(x,y,z,3*ones(1,p),1:p);
- Now, we use PCA to find the principal direction of these points

Sample code

- Let us randomly choose 20 points on a line in space (3-dimension).
- p=100;r=rand(1,p);x=2*r-1;y=r+2;z=3-r; scatter3(x,y,z,3*ones(1,p),1:p);colorbar;view(-30,10);
- We want to find the first principal component.
- First, we center the data by remove the mean of all points.
- x=x-mean(x);y=y-mean(y);z=z-mean(z);hold on;scatter3(x,y,z,3*ones(1,p),1:p);
- Now, we use PCA to find the principal direction of these points
- A=[x;y;z];[U,S,V]=svds(A);scatter3(x,y,z,3*ones(1,p),U(:,1)'*A);

Sample code

- Let us randomly choose 20 points on a line in space (3-dimension).
- p=100;r=rand(1,p);x=2*r-1;y=r+2;z=3-r; scatter3(x,y,z,3*ones(1,p),1:p);colorbar;view(-30,10);
- We want to find the first principal component.
- First, we center the data by remove the mean of all points.
- x=x-mean(x);y=y-mean(y);z=z-mean(z);hold on;scatter3(x,y,z,3*ones(1,p),1:p);
- Now, we use PCA to find the principal direction of these points
- A=[x;y;z];[U,S,V]=svds(A);scatter3(x,y,z,3*ones(1,p),U(:,1)'*A);
- Color is according to x-projection

The m-file and plot

- ```
  p=100;r=rand(1,p);x=2*r-1;y=r+2;z=3-r;
  scatter3(x,y,z,3*ones(1,p),2*[1:p]/p-1);view(-30,10);
  x=x-mean(x);y=y-mean(y);z=z-mean(z);
  A=[x;y;z];[U,S,V]=svds(A);
  hold on;
  scatter3(x,y,z,3*ones(1,p),U(:,1)'*A);
  ```
  ```

The m-file and plot

- p=100;r=rand(1,p);x=2*r-1;y=r+2;z=3-r;
  scatter3(x,y,z,3*ones(1,p),2*[1:p]/p-1);view(-30,10);
  x=x-mean(x);y=y-mean(y);z=z-mean(z);
  A=[x;y;z];[U,S,V]=svds(A);
  hold on;
  scatter3(x,y,z,3*ones(1,p),U(:,1)'*A);



Images as column vectors

- Assume we have a set of n images of the same size (say, $p \times q$): $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_n$.

Images as column vectors

- Assume we have a set of n images of the same size (say, $p \times q$): $\mathbf{I}_1$, $\mathbf{I}_2$, ... , $\mathbf{I}_n$.
- We transform the $p \times q$ image matrix $\mathbf{I}_i$ to a column vector of size $m := pq$ by stacking its columns. For simplicity, we use the same notation to denote the column vector.

Images as column vectors

- Assume we have a set of n images of the same size (say, $p \times q$): $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_n$.
- We transform the $p \times q$ image matrix $\mathbf{I}_i$ to a column vector of size $m := pq$ by stacking its columns. For simplicity, we use the same notation to denote the column vector.
- Now, as a standard step, we take out the mean image $\mathbf{I} := \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}_i$ from every image:

$$\tilde{\mathbf{I}}_i := \mathbf{I}_i - \mathbf{I}, \quad i = 1 : n$$

  This step guarantees $\tilde{\mathbf{I}}_i$ are centered at the origin (and hence contained in a true **subspace**).

Images as column vectors

- Assume we have a set of n images of the same size (say, $p \times q$): $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_n$.
- We transform the $p \times q$ image matrix $\mathbf{I}_i$ to a column vector of size $m := pq$ by stacking its columns. For simplicity, we use the same notation to denote the column vector.
- Now, as a standard step, we take out the mean image $\mathbf{I} := \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}_i$ from every image:

$$\tilde{\mathbf{I}}_i := \mathbf{I}_i - \mathbf{I}, \quad i = 1 : n$$

  This step guarantees $\tilde{\mathbf{I}}_i$ are centered at the origin (and hence contained in a true **subspace**).
- Form the $m \times n$ data matrix

$$X := [\tilde{\mathbf{I}}_1 \ \tilde{\mathbf{I}}_2 \ \cdots \ \tilde{\mathbf{I}}_n]$$