Dimensionality Reduction

Hung Le

University of Victoria

March 23, 2019

Motivation

Matrices are everywhere:

- Graphs: Web or Social Network.
- Pairwise interaction between two types of entities: movie rating, image-tags
- Spatial representation: images.

Motivation

Matrices are everywhere:

- Graphs: Web or Social Network.
- Pairwise interaction between two types of entities: movie rating, image-tags
- Spatial representation: images.

In many applications, the raw matrix can be summarized by a "narrower" matrices.

- "Narrower" means the output matrices have much less number of columns/rows.
- "Can be summarized " means we can almost recover the original matrix from the summarized matrices.

Motivation

Matrices are everywhere:

- Graphs: Web or Social Network.
- Pairwise interaction between two types of entities: movie rating, image-tags
- Spatial representation: images.

In many applications, the raw matrix can be summarized by a "narrower" matrices.

- "Narrower" means the output matrices have much less number of columns/rows.
- "Can be summarized " means we can almost recover the original matrix from the summarized matrices.

Dimensionality reduction: find the "narrower" matrices of the original matrix.

Eigenvectors and Eigenvalues of Symmetric Matrices

$\mathbf{e}$ is an eigenvector corresponding to an eigenvalue λ of a square matrix M iff:

$$M\mathbf{e} = \lambda\mathbf{e} \tag{1}$$

Eigenvectors and Eigenvalues of Symmetric Matrices

$\mathbf{e}$ is an eigenvector corresponding to an eigenvalue λ of a square matrix M iff:

$$M\mathbf{e} = \lambda\mathbf{e} \tag{1}$$

Fact 1 If $\mathbf{e}$ is an eigenvector of M, for any constant $c \neq 0$, $c\mathbf{e}$ is also an eigenvector of M (with the same eigenvalue) $\Rightarrow$ We often require that $||\mathbf{e}||_2 = 1$.

Fact 2 If $M \in \mathbb{R}^{n\times n}$ is real and symmetric, then M has n *real* eigenvectors and eigenvalues $\Rightarrow$ In this lecture, M is real, symmetric.

Fact 3 We can order

$$\begin{array}{cccc} \lambda_1 \geq & \lambda_2 \geq & \ldots \geq & \lambda_n \\ \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \end{array} \tag{2}$$

We can make $\mathbf{e}_j, \mathbf{e}_j$ orthogonal, i.e, $\mathbf{e}_i^T\mathbf{e}_j = 0$ for all $i \neq j$.

Eigenvectors and Eigenvalues of Symmetric Matrices

$\mathbf{e}$ is an eigenvector corresponding to an eigenvalue λ of a square matrix M iff:

$$M\mathbf{e} = \lambda\mathbf{e} \tag{1}$$

Fact 1 If $\mathbf{e}$ is an eigenvector of M, for any constant $c \neq 0$, $c\mathbf{e}$ is also an eigenvector of M (with the same eigenvalue) $\Rightarrow$ We often require that $||\mathbf{e}||_2 = 1$.

Fact 2 If $M \in \mathbb{R}^{n \times n}$ is real and symmetric, then M has n *real* eigenvectors and eigenvalues $\Rightarrow$ In this lecture, M is real, symmetric.

Fact 3 We can order

$$\begin{array}{cccc} \lambda_1 \geq & \lambda_2 \geq & \ldots \geq & \lambda_n \\ \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \end{array} \tag{2}$$

We can make $\mathbf{e}_j, \mathbf{e}_j$ orthogonal, i.e, $\mathbf{e}_i^T \mathbf{e}_j = 0$ for all $i \neq j$.

λ_1 and $\mathbf{e}_1$ called the principal eigenvalue the principal eigenvector, respectively

Eigenvectors and Eigenvalues of Symmetric Matrices - An Example

$$M = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \tag{3}$$

has:

$$\begin{aligned} \lambda_1 = 7 \qquad & \text{and} \qquad \mathbf{x}_1 = [\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}]^T \\ \lambda_2 = 2 \qquad & \text{and} \qquad \mathbf{x}_2 = [\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}}]^T \end{aligned} \tag{4}$$

Finding Eigenvalues by Solving Equations

Eigenvalues are the roots of the following equation (with variable λ):

$$\det(M - \lambda \mathbf{I}) = 0 \tag{5}$$

where $\mathbf{I}$ is an identity matrix, and $\det(X)$ is the determinant of an $n \times n$ matrix X.

Fact 4 $\det(M - \lambda \mathbf{I})$ is a degree-n polynomial with variable λ.

Finding Eigenvalues by Solving Equations

Eigenvalues are the roots of the following equation (with variable λ):

$$\det(M - \lambda \mathbf{I}) = 0 \tag{5}$$

where $\mathbf{I}$ is an identity matrix, and $\det(X)$ is the determinant of an $n \times n$ matrix X.

Fact 4 $\det(M - \lambda \mathbf{I})$ is a degree-n polynomial with variable λ.

Fact 5 If M is real and symmetric, Equation 5 has n real roots.

Finding Eigenvalues by Solving Equations

Eigenvalues are the roots of the following equation (with variable λ):

$$\det(M - \lambda \mathbf{I}) = 0 \tag{5}$$

where $\mathbf{I}$ is an identity matrix, and $\det(X)$ is the determinant of an $n \times n$ matrix X.

Fact 4 $\det(M - \lambda \mathbf{I})$ is a degree-n polynomial with variable λ.

Fact 5 If M is real and symmetric, Equation 5 has n real roots.

Fact 6 Computing the determinant of a matrix takes $O(n^3)$.

Finding Eigenvalues by Solving Equations - An example

$$M = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \tag{6}$$

We have:

$$\det(M - \lambda I) = \det(\begin{bmatrix} 3-\lambda & 2 \\ 2 & 6-\lambda \end{bmatrix}) = \lambda^2 - 9\lambda + 14 \tag{7}$$

Two roots are $\lambda_1 = 7$ and $\lambda_2 = 2$.

Finding Eigenvalues and Eigenvectors by Power Iteration

Finding principal eigenvector and value.

$$
\begin{array}{|l|}
\hline
\textsc{PowerIteration1}(M) \\
\qquad \mathbf{v}_0 \leftarrow \text{a random vector} \\
\qquad \textbf{for } t \leftarrow 1 \text{ to } k \\
\qquad\qquad \mathbf{v}_t = \frac{M\mathbf{v}_{t-1}}{||M\mathbf{v}_{t-1}||} \\
\qquad \text{return } \mathbf{v}_k, \mathbf{v}_k^T M \mathbf{v}_k. \\
\hline
\end{array}
$$

- Running time $O((m+n)k)$ where m is the number of non-zeros of M.
- $\mathbf{e}_1 \approx \mathbf{v}_k$ and $\lambda_1 \approx \mathbf{v}_k^T M \mathbf{v}_k$.

Power Iteration - An example

$$M = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \tag{8}$$

and $\mathbf{v}_0 = [1, 1]^T$. Then

$$M\mathbf{v}_0 = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \end{bmatrix} \tag{9}$$

Thus, $\mathbf{v}_1 = \frac{M\mathbf{v}_0}{||M\mathbf{v}_0||_2} = [0.530, 0.848]^T$. Repeat second time we get:

$$\mathbf{v}_2 = [0.471, 0.882]^T \tag{10}$$

The limiting vector is:

$$\mathbf{v}_k = [0.447, 0.894]^T \tag{11}$$

with $\lambda = \mathbf{v}_k^T M \mathbf{v}_k = 6.993$

Finding Eigenvalues and Eigenvectors by Power Iteration

Finding the second largest eigenvalues and vectors

$$
\begin{array}{l}
\textsc{PowerIteration2}(M) \\
\quad (\mathbf{e}_1, \lambda_1) \leftarrow \textsc{PowerIteration1}(M) \\
\quad \mathbf{v}_0 \leftarrow \text{a random vector} \\
\quad M_2 \leftarrow M - \lambda_1 \mathbf{e}_1 \mathbf{e}_1^T \\
\quad \textbf{for } t \leftarrow 1 \text{ to } k \\
\quad\quad \mathbf{v}_t = \frac{M_2 \mathbf{v}_{t-1}}{||M \mathbf{v}_{t-1}||} \\
\quad \text{return } \mathbf{v}_k, \mathbf{v}_k^T M_2 \mathbf{v}_k.
\end{array}
$$

- Running time $O((m+n)k)$ where m is the number of non-zeros of M.
- $\mathbf{e}_2 \approx \mathbf{v}_k$ and $\lambda_2 \approx \mathbf{v}_k^T M \mathbf{v}_k$.

The Matrix of Eigenvectors

$$E = \begin{bmatrix} | & | & & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n \\ | & | & & | \end{bmatrix} \tag{12}$$

Then:

$$E^T E = E E^T = \mathbf{I} \tag{13}$$

The Matrix of Eigenvectors - An Example

$$M = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \tag{14}$$

has:

$$\mathbf{x}_1 = [\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}]^T \qquad \mathbf{x}_2 = [\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}}]^T \tag{15}$$

Thus,

$$E = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix} \tag{16}$$

It is straightforward to verify that:

$$EE^T = E^T E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{17}$$

Principal Component Analysis- PCA

Given a set of points $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$ in R^d, find a direction where all the points line up best.

Principal Component Analysis- PCA

Given a set of points $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$ in R^d, find a direction where all the points line up best.

- A direction is a (unit) vector **w**.
- All the points line up best along **w** when:

$$\sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{w})^2 \tag{18}$$

is maximum.

Principal Component Analysis- PCA

Let:

$$X = \begin{bmatrix} — & \mathbf{x}_1^T & — \\ — & \mathbf{x}_2^T & — \\ & \vdots & \\ — & \mathbf{x}_n^T & — \end{bmatrix} \tag{19}$$

Then

$$\sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{w})^2 = ||X\mathbf{w}||_2^2 = \mathbf{w}^T X^T X \mathbf{w} \tag{20}$$

Thus, $\sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{w})^2$ is maximum when $\mathbf{w}$ is the principal eigenvector of $X^T X$.

Principal Component Analysis- PCA

Let:

$$X = \begin{bmatrix} — & \mathbf{x}_1^T & — \\ — & \mathbf{x}_2^T & — \\ & \vdots & \\ — & \mathbf{x}_n^T & — \end{bmatrix} \tag{19}$$

Then

$$\sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{w})^2 = ||X\mathbf{w}||_2^2 = \mathbf{w}^T X^T X \mathbf{w} \tag{20}$$

Thus, $\sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{w})^2$ is maximum when $\mathbf{w}$ is the principal eigenvector of $X^T X$. **Question:** what is $\mathbf{w}^T X^T X \mathbf{w}$ when $\mathbf{w}$ is the principal eigenvector of $X^T X$?

Principal Component Analysis- PCA

Let:

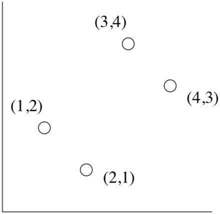
$$X = \begin{bmatrix} — & \mathbf{x}_1^T & — \\ — & \mathbf{x}_2^T & — \\ & \vdots & \\ — & \mathbf{x}_n^T & — \end{bmatrix} \tag{19}$$

Then

$$\sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{w})^2 = ||X\mathbf{w}||_2^2 = \mathbf{w}^T X^T X \mathbf{w} \tag{20}$$

Thus, $\sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{w})^2$ is maximum when $\mathbf{w}$ is the principal eigenvector of $X^T X$. **Question:** what is $\mathbf{w}^T X^T X \mathbf{w}$ when $\mathbf{w}$ is the principal eigenvector of $X^T X$?
Answer: $\lambda_1(X^T X)$.

PCA - An Example



$$X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} \quad \text{and} \quad X^T X = \begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix} \tag{21}$$

has $\lambda_1(X^T X) = 58$ with vector $\mathbf{x}_1 = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^T$

PCA - More Components

Given a set of points $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$ in R^d, find the second best direction where all the points line up best.

PCA - More Components

Given a set of points $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$ in R^d, find the second best direction where all the points line up best.
A direction $\mathbf{u}$ is the second best if

- $\mathbf{u}^T \mathbf{w} = 0$, here $\mathbf{w}$ is the best.
- All the points line up best along $\mathbf{u}$ among all directions orthogonal to w. That is

$$\sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{u})^2 \tag{22}$$

  is maximum among all directions orthogonal to w

PCA - More Components

Given a set of points $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$ in R^d, find the second best direction where all the points line up best.
A direction $\mathbf{u}$ is the second best if

- $\mathbf{u}^T \mathbf{w} = 0$, here $\mathbf{w}$ is the best.
- All the points line up best along $\mathbf{u}$ among all directions orthogonal to w. That is

$$\sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{u})^2 \tag{22}$$

  is maximum among all directions orthogonal to w

Fact: $\mathbf{u}$ is the second eigenvector of $X^T X$ corresponding to the second largest eigenvalue $\lambda_2(X^T X)$.

PCA - More Components

Given a set of points $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$ in R^d, find the *k*-th best direction where all the points line up best.
Fact: The *k*-the best direction is the eigenvector of $X^T X$ corresponding to the $k - th$ largest eigenvalue $\lambda_k(X^T X)$.

Using PCA for Dimensionality Reduction

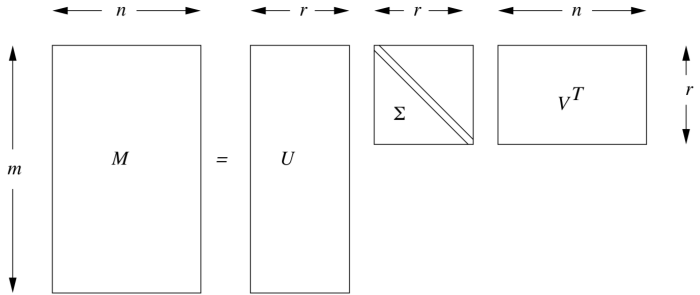
Given k eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k$ corresponding to k largest eigenvalues. We can then represent each new data point $\mathbf{x}_i \in \mathcal{D}$ as:

$$\begin{bmatrix} \mathbf{x}_i^T \mathbf{e}_1 \\ \mathbf{x}_i^T \mathbf{e}_2 \\ . \\ . \\ . \\ \mathbf{x}_i^T \mathbf{e}_k \end{bmatrix} \tag{23}$$

Singular Value Decomposition

Let M be a $m \times n$ matrix. Let r be the rank of M. A Singular Value Decomposition is a decomposition of M into three matrices U, Σ, V where:

- U is a column-orthogonal $m \times r$ matrix, i.e, $U^T U = \mathbf{I}_m$
- Σ is a diagonal $r \times r$ matrix. Elements on the diagonal of Σ are singular values and are *decreasingly* ordered.
- V is an column-orthogonal $n \times r$ matrix, i.e, $V^T V = \mathbf{I}_n$



SVD

$$
\underbrace{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 2 & 2 \end{bmatrix}}_{M} = \underbrace{\begin{bmatrix} .14 & 0 \\ .42 & 0 \\ .56 & 0 \\ .70 & 0 \\ 0 & .60 \\ 0 & .75 \\ 0 & .30 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 12.4 & 0 \\ 0 & 9.5 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} .58 & .58 & .58 & 0 & 0 \\ 0 & 0 & 0 & .71 & .71 \end{bmatrix}}_{V^{\mathrm{T}}}
$$

Understanding SVD

Think of U, Σ, V as a representation of *concepts* hidden in M.

	Matrix	Alien	Star Wars	Casablanca	Titanic
Joe	1	1	1	0	0
Jim	3	3	3	0	0
John	4	4	4	0	0
Jack	5	5	5	0	0
Jill	0	0	0	4	4
Jenny	0	0	0	5	5
Jane	0	0	0	2	2

Two concepts:

$$\texttt{ScienceFiction} = \{\texttt{TheMatrixAlien}, \texttt{StarWars}\}$$
$$\texttt{Romance} = \{\texttt{Casablanca}, \texttt{Titanic}\}$$

Dimensionality Reduction by SVD

A rank-k SVD approximation of M is the matrix $U_k \Sigma_k V_k^T$ where:

- U_k contains the first k columns of U.
- Σ_k contains k largest elements of Σ.
- V_k contains the first k columns of V.

Dimensionality Reduction by SVD - An Example

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} =$$

M'

$$\begin{bmatrix} .13 & .02 & -.01 \\ .41 & .07 & -.03 \\ .55 & .09 & -.04 \\ .68 & .11 & -.05 \\ .15 & -.59 & .65 \\ .07 & -.73 & -.67 \\ .07 & -.29 & .32 \end{bmatrix} \begin{bmatrix} 12.4 & 0 & 0 \\ 0 & 9.5 & 0 \\ 0 & 0 & 1.3 \end{bmatrix} \begin{bmatrix} .56 & .59 & .56 & .09 & .09 \\ .12 & -.02 & .12 & -.69 & -.69 \\ .40 & -.80 & .40 & .09 & .09 \end{bmatrix}$$

U $\qquad \Sigma$ $\qquad V^{\mathrm{T}}$

Dimensionality Reduction by SVD - An Example

A rank-2 approximation of M':

$$\begin{bmatrix} .13 & .02 \\ .41 & .07 \\ .55 & .09 \\ .68 & .11 \\ .15 & -.59 \\ .07 & -.73 \\ .07 & -.29 \end{bmatrix} \begin{bmatrix} 12.4 & 0 \\ 0 & 9.5 \end{bmatrix} \begin{bmatrix} .56 & .59 & .56 & .09 & .09 \\ .12 & -.02 & .12 & -.69 & -.69 \end{bmatrix}$$

$$= \begin{bmatrix} 0.93 & 0.95 & 0.93 & .014 & .014 \\ 2.93 & 2.99 & 2.93 & .000 & .000 \\ 3.92 & 4.01 & 3.92 & .026 & .026 \\ 4.84 & 4.96 & 4.84 & .040 & .040 \\ 0.37 & 1.21 & 0.37 & 4.04 & 4.04 \\ 0.35 & 0.65 & 0.35 & 4.87 & 4.87 \\ 0.16 & 0.57 & 0.16 & 1.98 & 1.98 \end{bmatrix}$$

Dimensionality Reduction by SVD - Why?

Theorem

Given M, a rank-k SVD approximation of M, denoted by $A_k = U_k \Sigma_k V_k^T$ has:

$$||M - A_k||_F \tag{24}$$

minimum among all possible rank-k matrices.

$||X||_F$ is the *Frobenius norm* of X:

$$||X||_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} X[i,j]^2} \tag{25}$$

How to choose k

Choose k so that at least 90% of *energy* of Σ is preserved.

$$\texttt{Energy}(\Sigma) = \sum_{i=1}^{r} \Sigma[i, i]^2 \tag{26}$$

Querying Concept

	Matrix	Alien	Star Wars	Casablanca	Titanic
Joe	1	1	1	0	0
Jim	3	3	3	0	0
John	4	4	4	0	0
Jack	5	5	5	0	0
Jill	0	0	0	4	4
Jenny	0	0	0	5	5
Jane	0	0	0	2	2

Suppose that a new person P has seen only one movie `the Matrix` and rated it 4. Recall:

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 2 & 2 \end{bmatrix}}_{M} = \underbrace{\begin{bmatrix} .14 & 0 \\ .42 & 0 \\ .56 & 0 \\ .70 & 0 \\ 0 & .60 \\ 0 & .75 \\ 0 & .30 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 12.4 & 0 \\ 0 & 9.5 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} .58 & .58 & .58 & 0 & 0 \\ 0 & 0 & 0 & .71 & .71 \end{bmatrix}}_{V^{\mathrm{T}}}$$

Querying Concept (Contt.)

Let $\mathbf{q}$ be the row representation of P, that is:

$$\mathbf{q} = \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{27}$$

We can determine the "concept space" of P by:

$$\mathbf{q}V = \begin{bmatrix} 2.32 & 0 \end{bmatrix} \tag{28}$$

Computing SVD

Recall

$$M = U\Sigma V^T \tag{29}$$

Hence,

$$M^T M = V\Sigma U^T U\Sigma V^T = V\Sigma^2 V^T \tag{30}$$

which implies:

$$M^T MV = V\Sigma^2 \tag{31}$$

Conclusion: V is the set of eigenvectors of $M^T M$. By the same argument, U is the set of eigenvectors of MM^T.
Question: What is Σ?