Introduction to Machine Learning

Lecture 6: Linear Algebra
(Eigen Value Decomposition, Singular Value Decomposition)

Prof. Subrahmanya Swamy Peruru
Department of Electrical Engineering
IIT Kanpur

Recap

Matrix Representation of $T(\mathbf{x})$

Store the transformed basis vectors as columns of a matrix:

$$A = \begin{bmatrix} | & | & & | \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_m) \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n \times m}.$$

Then, any $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_m\mathbf{e}_m$

$$T(\mathbf{x}) = x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \cdots + x_mT(\mathbf{e}_m)$$

$$= \begin{bmatrix} | & | & & | \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_m) \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

$$\boxed{T(\mathbf{x}) = A\mathbf{x}}$$

Important remark: The matrix representation changes with the choice of basis

Column Space and Rank of a Matrix

Column space: The span of the columns (i.e., all linear combinations of the columns).

$$A = \begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_m \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n \times m}$$

$$\mathrm{col}(A) = \{A\mathbf{x}\colon \mathbf{x} \in \mathbb{R}^m\} = \mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} \subseteq \mathbb{R}^n$$

Rank: Dimension of the column space (i.e., the number of linearly independent columns)

Null Space

The null space of a matrix is the set of all **input vectors that get mapped to the zero** vector

$$\mathrm{null}(A) = \{\mathbf{x} \in \mathbb{R}^m \colon A\mathbf{x} = \mathbf{0}\} \subseteq \mathbb{R}^m.$$

Transformation of a linear combination of some vectors **is same as first applying the transformation** and **then taking their linear combination**

Recap

Orthogonal Matrices as Coordinate Transforms

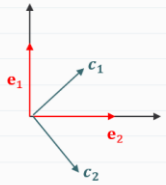
Orthogonal matrices are used for changing coordinates between orthonormal bases

Standard basis in $\mathbb{R}^2$:

$$\mathbf{e}_1 = \begin{bmatrix}1\\0\end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix}0\\1\end{bmatrix}.$$

New orthonormal basis:

$$\mathbf{c}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}, \quad \mathbf{c}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}.$$



Form the orthogonal matrix with the *new basis vectors as columns*:

$$Q = [\mathbf{c}_1 \quad \mathbf{c}_2] = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}, \qquad Q^\top Q = I.$$

Coordinate conversion rules:

$$[\mathbf{x}]_{\text{std}} = Q\,[\mathbf{x}]_{\text{new}} \quad \text{and} \quad [\mathbf{x}]_{\text{new}} = Q^\top [\mathbf{x}]_{\text{std}}$$

Eigenvalues and Eigenvectors

$\mathbf{x} \neq 0$ is an eigenvector if

$$A\mathbf{x} = \lambda\mathbf{x}.$$

Eigenvalues are the roots of the characteristic equation:

$$\det(A - \lambda I) = 0$$

Eigenvectors represent special directions that are preserved by a matrix, while eigenvalues indicate the corresponding scaling.

If $\mathbf{x}$ is an eigenvector, then any nonzero scalar multiple of $\mathbf{x}$ is also an eigenvector.

Recap

Diagonal Matrices

Definition:

A matrix is called *diagonal* if all its off-diagonal entries are zero.

$$D = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}.$$

Observations:

1. Each coordinate of the vector is scaled independently when multiplied by a diagonal matrix.

$$D\mathbf{x} = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} d_1 x_1 \\ d_2 x_2 \end{bmatrix}.$$

2. The diagonal entries are its eigen values, i.e., $\lambda_1 = \mathrm{d}_1,\ \lambda_2 = \mathrm{d}_2$

3. The standard basis vectors, $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, are its eigenvectors.

Exercise: Verify the above claims on eigen values and vectors.

These observations form the core idea of Eigenvalue Decomposition.

Eigenvalue Decomposition

If it is possible to **construct a basis using the eigenvectors** of a matrix,

then the **matrix is diagonal** when represented **in the eigenvector basis**.

The matrix (represented using standard basis vectors) may not be diagonal, but its new representation using the eigen basis will be diagonal!

Example: Symmetric Matrix

A symmetric matrix admits an **orthonormal basis consisting of its eigenvectors.**

We restrict attention to **symmetric matrices**.

Coordinates in the Eigen Basis

Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is an **orthonormal eigen basis** of a symmetric matrix A.

Then, any arbitrary vector $\mathbf{x} \in \mathbb{R}^n$ can be written as a linear combination of the eigen basis:

$$\mathbf{x} = \alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2 + \cdots + \alpha_n\mathbf{b}_n.$$

So the coordinates of $\mathbf{x}$ in the eigen basis are

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}.$$

Now apply A:

$$\begin{aligned} \mathbf{y} = A\mathbf{x} = & A(\alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2 \cdots + \alpha_n\mathbf{b}_n) \\ = & \alpha_1 A\mathbf{b}_1 + \alpha_2 A\mathbf{b}_2 + \cdots + \alpha_n A\mathbf{b}_n) \\ = & \alpha_1\lambda_1\mathbf{b}_1 + \alpha_2\lambda_2\mathbf{b}_2 \cdots + \alpha_n\lambda_n\mathbf{b}_n. \end{aligned}$$

So, in the *same eigen basis*,

$$[\mathbf{y}]_{\mathcal{B}} = \begin{bmatrix} \lambda_1\alpha_1 \\ \lambda_2\alpha_2 \\ \vdots \\ \lambda_n\alpha_n \end{bmatrix}.$$

Key idea: each coordinate is scaled independently.

Diagonal Form Appears Naturally

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}, [\mathbf{y}]_{\mathcal{B}} = \begin{bmatrix} \lambda_1\alpha_1 \\ \lambda_2\alpha_2 \\ \vdots \\ \lambda_n\alpha_n \end{bmatrix}$$

The coordinate relation can be written compactly as

$$[\mathbf{y}]_{\mathcal{B}} = \begin{bmatrix} \lambda_1\alpha_1 \\ \lambda_2\alpha_2 \\ \vdots \\ \lambda_n\alpha_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}.$$

Let

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$

Then

$$[\mathbf{y}]_{\mathcal{B}} = \Lambda[\mathbf{x}]_{\mathcal{B}}.$$

Remember that $\mathbf{y} = A\mathbf{x}$ implicitly means $\mathbf{x}$ and $\mathbf{y}$ are represented using standard basis:

$$[\mathbf{y}]_{\mathrm{std}} = A[\mathbf{x}]_{\mathrm{std}}$$

Eigen Decomposition: Relates A and Λ

Going back to standard basis gives Eigen Decomposition

Remember that we can use Orthogonal matrix to convert coordinates

$$Q = [\,\mathbf{b}_1\ \mathbf{b}_2\ \cdots\ \mathbf{b}_n\,].$$

Then:

$$[\mathbf{x}]_{\text{std}} = Q[\mathbf{x}]_{\mathcal{B}}, \qquad [\mathbf{x}]_{\mathcal{B}} = Q^T[\mathbf{x}]_{\text{std}}.$$

Relating A and Λ:

$$[\mathbf{y}]_{\mathcal{B}} = \Lambda[\mathbf{x}]_{\mathcal{B}}$$

$$[\mathbf{y}]_{\mathcal{B}} = \Lambda\, Q^T[\mathbf{x}]_{\text{std}}$$

Multiply on both sides by Q:

$$Q[\mathbf{y}]_{\mathcal{B}} = Q\Lambda\, Q^T[\mathbf{x}]_{\text{std}}$$

$$[\mathbf{y}]_{\text{std}} = Q\Lambda\, Q^T[\mathbf{x}]_{\text{std}}$$

But we know:

$$[\mathbf{y}]_{\text{std}} = A[\mathbf{x}]_{\text{std}}$$

$$\boxed{A = Q\Lambda Q^T}.$$

EVD for Symmetric Matrix

Every real symmetric matrix A can be written as

$$A = Q\Lambda Q^T,$$

where

Q is an orthogonal matrix whose columns are eigenvectors of A,

Λ is a diagonal matrix containing eigenvalues of A.

Change basis: $Q^T\mathbf{x}$ converts $\mathbf{x}$ into the eigenvector basis from the standard basis

Scale: Λ scales each eigen-direction independently

Change back: Q returns to the standard basis

This representation is called the Eigenvalue Decomposition (EVD).

Numerical Example

Consider the symmetric matrix

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Its eigenvalues are

$$\lambda_1 = 3, \qquad \lambda_2 = 1.$$

Corresponding (orthonormal) eigenvectors are

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Form

$$Q = [\mathbf{u}_1 \quad \mathbf{u}_2] = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad \Lambda = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then the diagonal form of A is

$$\boxed{A = Q\Lambda Q^T} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

What about other matrices?

EVD works only for square matrices (and even then, only when the matrix admits an eigenbasis)

Now consider a general matrix:

$$A \in \mathbb{R}^{m \times n}$$

Will the idea of basis change work for a general matrix?

YES! SVD generalises this idea.

$$\mathbf{y} = A\mathbf{x}$$

Input space: $\mathbf{x} \in \mathbb{R}^n$
Output space: $\mathbf{y} \in \mathbb{R}^m$

Input and output spaces are different:
A single common basis is not feasible the spaces.

We need two basis sets:
one basis for inputs
one basis for outputs

What SVD Provides

Singular Value Decomposition (SVD) provides:

An orthonormal basis for the **input space**:

$$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\} \subset \mathbb{R}^n$$

An orthonormal basis for the **output space**:

$$\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\} \subset \mathbb{R}^m$$

These are called:

Right singular vectors $(\mathbf{v}_i)$

Left singular vectors $(\mathbf{u}_i)$

Diagonalization in EVD comes from:

EVD: $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$

What about SVD?

Relation between Singular Vectors

The bases are related by:

$$\boxed{A\mathbf{v}_i = \sigma_i \mathbf{u}_i, \quad i = 1, \ldots, \min(m, n)}$$

where:

$\sigma_i \geq 0$ are singular values

usually ordered as $\sigma_1 \geq \sigma_2 \geq \cdots$

The rest of the story is similar to EVD.

- Express input vector $\mathbf{x}$ using the basis given by right singular vectors $\{\mathbf{v_i}\}$
- Express output vector $\mathbf{y}$ using the basis given by left singular vectors $\{\mathbf{u_i}\}$
- Then these coordinates are related by a diagonal matrix.

The SVD Formula and Interpretation

Putting everything together:

$$\boxed{A = U\Sigma V^{\top}}$$

Interpretation:

1. $V^{\top}$:change standard basis coordinates to right singular basis coordinates
2. Σ: diagonal scaling
3. U: change left singular basis coordinates to standard basis coordinates

Right singular vectors are eigen vectors of $A^T A$

Left singular vectors are eigen vectors of AA^T

Singular values are the square roots of their eigen values.

Important Consequences of SVD

Maximum stretch

The direction of maximum stretching corresponds to the largest singular value.

$$\max_{\|\mathbf{x}\|=1} \| A\mathbf{x} \| = \sigma_1 \quad \text{(achieved at } \mathbf{x} = \mathbf{v}_1\text{)}$$

Rank

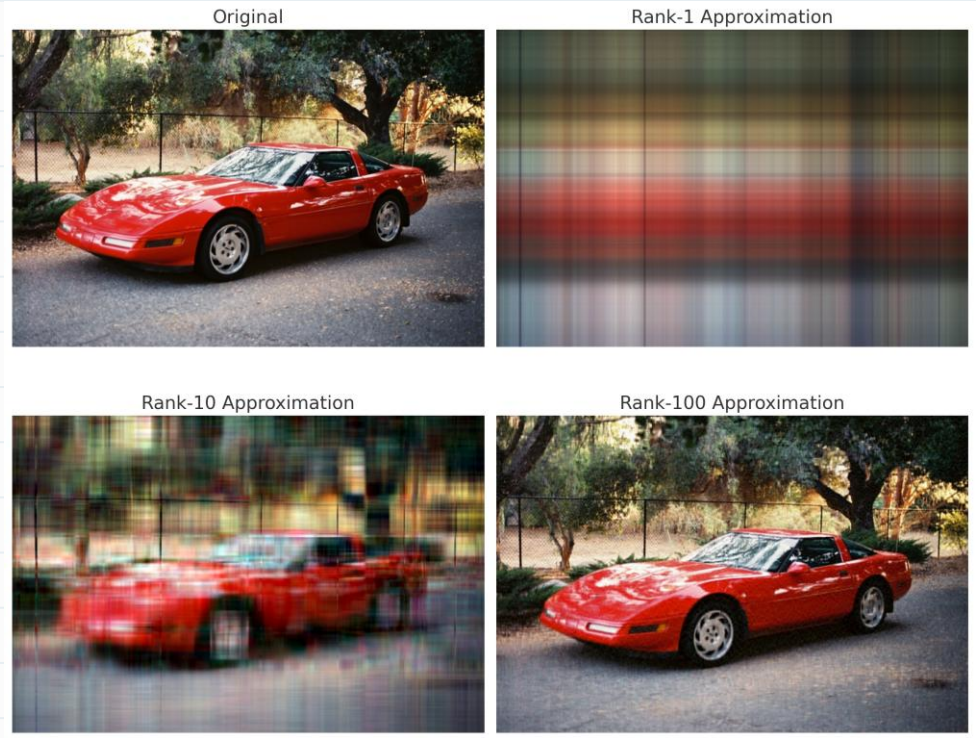
$$\text{Number of non–zero singular values} = \operatorname{rank}(A)$$

Low-rank approximation

Keeping top k singular values gives best rank-k approximation of A

$$A_k = \sum_{i=1}^{k} \sigma_i \, u_i v_i^\top$$

SVD based Image compression



SVD for solving $A\mathbf{x} = b$

Consider the system

$$A\mathbf{x} = \mathbf{b}, \qquad A \in \mathbb{R}^{m \times n},\ \mathbf{b} \in \mathbb{R}^{m}.$$

Exact solution. A solution exists if and only if $\mathbf{b}$ belongs to columns space of A

$$\mathbf{b} \in \mathcal{C}(A).$$

Approximate solution.

If $\mathbf{b} \notin \mathcal{C}(A)$**.** We find the best possible approximation

$$\mathbf{x}^{\star} = \underset{\mathbf{x}}{\operatorname{argmin}} \parallel A\mathbf{x} - \mathbf{b} \parallel_2.$$

$A\mathbf{x}^{\star}$ is the orthogonal projection of $\mathbf{b}$ onto $\mathcal{C}(A)$.

Assuming $\mathrm{A^{T}A}$ is invertible *(i.e., the columns of A are linearly independent)*,

$$\boxed{\mathbf{x}^{\star} = (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}\mathbf{b}.}$$

In the general case, when $A^{\mathsf{T}}A$ is not invertible, the solution is given by the SVD-based Moore–Penrose pseudoinverse $\mathbf{x}^{\star} = A^{\dagger}\mathbf{b}$.

Positive Semidefinite (PSD) Matrix

Definition. A real symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called *positive semidefinite (PSD)* if

$$\boxed{x^\top A x \geq 0 \quad \text{for all } x \in \mathbb{R}^n}$$

Equivalent views.

All eigenvalues of A are non-negative

$A = B^\top B$ for some matrix B

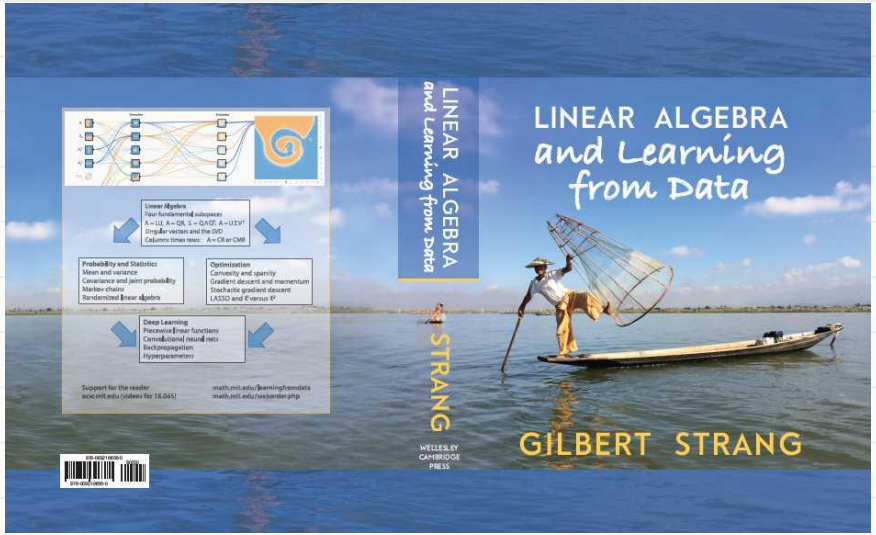
Why PSD matters in ML.

Covariance matrices are PSD

$M^\top M$ is always PSD for any matrix M

Reference Book

Read Chapter 1 from this book by G. Strang, "*Linear Algebra and Learning from Data*", MIT Press, 2019.



Thank You