Linear Algebra - Part II
Projection, Eigendecomposition, SVD

(Adapted from Punit Shah's slides)

2019

Brief Review from Part 1

- Matrix Multiplication is a linear tranformation.

- Symmetric Matrix:

$$\mathbf{A} = \mathbf{A}^T$$

- Orthogonal Matrix:

$$\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I} \quad \text{and} \quad \mathbf{A}^{-1} = \mathbf{A}^T$$

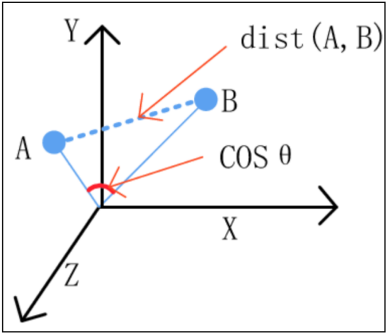
- L2 Norm:

$$||\mathbf{x}||_2 \; = \; \sqrt{\sum_i x_i^2}$$

Angle Between Vectors

- Dot product of two vectors can be written in terms of their L2 norms and the angle θ between them.

$$\mathbf{a}^T\mathbf{b} \;=\; ||\mathbf{a}||_2||\mathbf{b}||_2 \cos(\theta)$$



Cosine Similarity

- Cosine between two vectors is a measure of their similarity:

$$\cos(\theta) = \frac{\mathbf{a} \cdot \mathbf{b}}{||\mathbf{a}|| \ ||\mathbf{b}||}$$

- **Orthogonal Vectors**: Two vectors $\mathbf{a}$ and $\mathbf{b}$ are orthogonal to each other if $\mathbf{a} \cdot \mathbf{b} = 0$.

Vector Projection

- Given two vectors $\mathbf{a}$ and $\mathbf{b}$, let $\hat{\mathbf{b}} = \frac{\mathbf{b}}{||\mathbf{b}||}$ be the unit vector in the direction of $\mathbf{b}$.
- Then $\mathbf{a}_1 = a_1 \cdot \hat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{a}$ onto a straight line parallel to $\mathbf{b}$, where

$$a_1 = ||\mathbf{a}|| \cos(\theta) = \mathbf{a} \cdot \hat{\mathbf{b}} = \mathbf{a} \cdot \frac{\mathbf{b}}{||\mathbf{b}||}$$

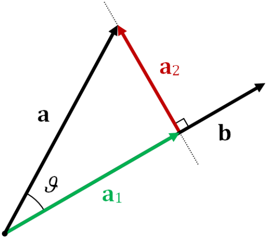


Image taken from wikipedia.

Diagonal Matrix

- Diagonal matrix has mostly zeros with non-zero entries only in the diagonal, e.g. identity matrix.
- A square diagonal matrix with diagonal elements given by entries of vector $\mathbf{v}$ is denoted:

$$\text{diag}(\mathbf{v})$$

- Multiplying vector $\mathbf{x}$ by a diagonal matrix is efficient:

$$\text{diag}(\mathbf{v})\mathbf{x} \;=\; \mathbf{v} \odot \mathbf{x}$$

$\odot$ is the entrywise product.

- Inverting a square diagonal matrix is efficient:

$$\text{diag}(\mathbf{v})^{-1} \;=\; \text{diag}\Big([\frac{1}{v_1}, \ldots, \frac{1}{v_n}]^T\Big)$$

Determinant

- Determinant of a square matrix is a mapping to a scalar.

$$\det(\mathbf{A}) \quad \text{or} \quad |\mathbf{A}|$$

- Measures how much multiplication by the matrix expands or contracts the space.

- Determinant of product is the product of determinants:

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$$

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

List of Equivalencies

The following are all equivalent:

- $\mathbf{Ax} = \mathbf{b}$ has a **unique** solution (for every b with correct dimension).
- $\mathbf{Ax} = \mathbf{0}$ has a unique, trivial solution: $\mathbf{x} = \mathbf{0}$.
- Columns of $\mathbf{A}$ are linearly independent.
- $\mathbf{A}$ is invertible, i.e. $\mathbf{A}^{-1}$ exists.
- $\det(\mathbf{A}) \neq 0$

$x^{'}$

Zero Determinant

If $\det(\mathbf{A}) = 0$, then:

- $\mathbf{A}$ is linearly dependent.
- $\mathbf{Ax} = \mathbf{b}$ has no solution or infinitely many solutions.
- $\mathbf{Ax} = \mathbf{0}$ has a non-zero solution.

Matrix Decomposition

- We can decompose an integer into its prime factors, e.g. $12 = 2 \times 2 \times 3$.

- Similarly, matrices can be decomposed into factors to learn universal properties:

$$\mathbf{A} = \mathbf{V}\mathrm{diag}(\boldsymbol{\lambda})\mathbf{V}^{-1}$$

- Unlike integers, matrix factorization is not unique:

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

Eigenvectors

- An eigenvector of a square matrix $\mathbf{A}$ is a nonzero vector $\mathbf{v}$ such that multiplication by $\mathbf{A}$ only changes the scale of $\mathbf{v}$.

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

- The scalar λ is known as the **eigenvalue**.

- If $\mathbf{v}$ is an eigenvector of $\mathbf{A}$, so is any rescaled vector $s\mathbf{v}$. Moreover, $s\mathbf{v}$ still has the same eigenvalue. Thus, we constrain the eigenvector to be of unit length:

$$||\mathbf{v}|| = 1$$

Characteristic Polynomial(1)

- Eigenvalue equation of matrix $\mathbf{A}$:

$$\begin{aligned} \mathbf{A}\mathbf{v} &= \lambda\mathbf{v} \\ \lambda\mathbf{v} - \mathbf{A}\mathbf{v} &= \mathbf{0} \\ (\lambda\mathbf{I} - \mathbf{A})\mathbf{v} &= \mathbf{0} \end{aligned}$$

- If nonzero solution for $\mathbf{v}$ exists, then it must be the case that:

$$\det(\lambda\mathbf{I} - \mathbf{A}) = 0$$

- Unpacking the determinant as a function of λ, we get:

$$P_A(\lambda) = |\lambda\mathbf{I} - \mathbf{A}| = 1 \times \lambda^n + c_{n-1} \times \lambda^{n-1} + \ldots + c_0$$

- This is called the characterisitc polynomial of A.

Characteristic Polynomial(2)

- If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are roots of the characteristic polynomial, they are eigenvalues of **A** and we have:

$$P_A(\lambda) = \prod_{i=1}^{n} (\lambda - \lambda_i)$$

- $c_{n-1} = -\sum_{i=1}^{n} \lambda_i = -tr(A)$
- $c_0 = (-1)^n \prod_{i=1}^{n} \lambda_i = (-1)^n det(\mathbf{A})$
- Roots might be complex. If a root has multiplicity of $r_j > 1$, then the dimension of eigenspace for that eigenvalue might be less than r_j (or equal but never more). But one eigenvector is guaranteed.

Example

- Consider the matrix:

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

- The characteristic polynomial is:

$$\det(\lambda \mathbf{I} - \mathbf{A}) = \det \begin{bmatrix} \lambda - 2 & -1 \\ -1 & \lambda - 2 \end{bmatrix} = 3 - 4\lambda + \lambda^2 = 0$$

- It has roots $\lambda = 1$ and $\lambda = 3$ which are the two eigenvalues of $\mathbf{A}$.

- We can then solve for eigenvectors using $\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$:

$$\mathbf{v}_{\lambda=1} = [1, -1]^T \quad \text{and} \quad \mathbf{v}_{\lambda=3} = [1, 1]^T$$

Eigendecomposition

- Suppose that $n \times n$ matrix $\mathbf{A}$ has n linearly independent eigenvectors $\{\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(n)}\}$ with eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$.
- Concatenate eigenvectors (as columns) to form matrix $\mathbf{V}$.
- Concatenate eigenvalues to form vector $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_n]^T$.
- The **eigendecomposition** of $\mathbf{A}$ is given by:

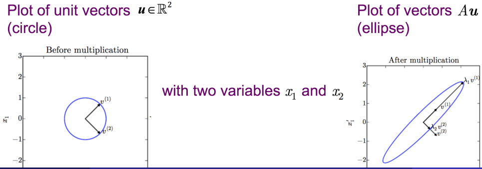
$$\mathbf{AV} = \mathbf{V}diag(\lambda) \implies \mathbf{A} = \mathbf{V}\text{diag}(\boldsymbol{\lambda})\mathbf{V}^{-1}$$

Symmetric Matrices

- Every symmetric (hermitian) matrix of dimension n has a set of (not necessarily unique) n orthogonal eigenvectors. Furthermore, All eigenvalues are real.
- Every real symmetric matrix **A** can be decomposed into real-valued eigenvectors and eigenvalues:

$$\mathbf{A} = \mathbf{Q} \boldsymbol{\Lambda} \mathbf{Q}^T$$

- **Q** is an orthogonal matrix of the eigenvectors of **A**, and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues.
- We can think of **A** as scaling space by λ_i in direction $\mathbf{v}^{(i)}$.



Eigendecomposition is not Unique

- Decomposition is not unique when two eigenvalues are the same.
- By convention, order entries of $\mathbf{\Lambda}$ in descending order. Then, eigendecomposition is unique if all eigenvalues are unique.
- If any eigenvalue is zero, then the matrix is **singular**. Because if $\mathbf{v}$ is the corresponding eigenvector we have: $\mathbf{Av} = 0\mathbf{v} = 0$.

Positive Definite Matrix

- If a symmetric matrix A has the property:

$$\mathbf{x}^T \mathbf{A} \mathbf{x} > 0 \quad \text{for any nonzero vector } \mathbf{x}$$

  Then A is called **positive definite**.

- If the above inequality is not strict then A is called **positive semidefinite**.

- For positive (semi)definite matrices all eigenvalues are positive(non negative).

Singular Value Decomposition (SVD)

- If $\mathbf{A}$ is not square, eigendecomposition is undefined.
- **SVD** is a decomposition of the form:

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

- SVD is more general than eigendecomposition.
- Every real matrix has a SVD.

SVD Definition (1)

- Write $\mathbf{A}$ as a product of three matrices: $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$.
- If $\mathbf{A}$ is $m \times n$, then $\mathbf{U}$ is $m \times m$, $\mathbf{D}$ is $m \times n$, and $\mathbf{V}$ is $n \times n$.
- $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices, and $\mathbf{D}$ is a diagonal matrix (not necessarily square).
- Diagonal entries of $\mathbf{D}$ are called **singular values** of $\mathbf{A}$.
- Columns of $\mathbf{U}$ are the **left singular vectors**, and columns of $\mathbf{V}$ are the **right singular vectors**.

SVD Definition (2)

- SVD can be interpreted in terms of eigendecompostion.
- Left singular vectors of $\mathbf{A}$ are the eigenvectors of $\mathbf{AA}^T$.
- Right singular vectors of $\mathbf{A}$ are the eigenvectors of $\mathbf{A}^T\mathbf{A}$.
- Nonzero singular values of $\mathbf{A}$ are square roots of eigenvalues of $\mathbf{A}^T\mathbf{A}$ and $\mathbf{AA}^T$.
- Numbers on the diagonal of D are sorted largest to smallest and are positive (This makes SVD unique)

SVD Optimality

- We can write $\mathbf{A} = \mathbf{UDV}^T$ in this form: $\mathbf{A} = \sum_{i=1}^{n} d_i \mathbf{u}_i \mathbf{v}_i^T$
- Instead of n we can sum up to r: $\mathbf{A}_r = \sum_{i=1}^{r} d_i \mathbf{u}_i \mathbf{v}_i^T$
- This is called a low rank approximation of A.
- $\mathbf{A}_r$ is the best approximation of rank r by many norms:
    - When considering vector norm, it is optimal. Which means $\mathbf{A}_r$ is a linear transformation that captures as much energy as possible.
    - When considering Frobenius norm, it is optimal which means $\mathbf{A}_r$ is projection of A on the best(closest) r dimensional subspace.