Linear Algebra Review

(Adapted from Punit Shah's slides)

Introduction to Machine Learning (CSC 311)

University of Toronto

Fall 2021

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 product of other matrices.

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

- Examples are Eigendecomposition, SVD, Schur decomposition, LU decomposition, ...

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:

$$||\mathbf{v}||_2 = 1$$

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:

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

- 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 $\mathbf{A}$ and we have $P_A(\lambda) = \prod_{i=1}^{n}(\lambda - \lambda_i)$.
- $c_{n-1} = -\sum_{i=1}^{n} \lambda_i = -tr(A)$. This means that the sum of eigenvalues equals to the trace of the matrix.
- $c_0 = (-1)^n \prod_{i=1}^{n} \lambda_i = (-1)^n det(\mathbf{A})$. The determinant is equal to the product of eigenvalues.
- Roots might be complex. If a root has multiplicity of $r_j > 1$ (This is called the algebraic dimension of eigenvalue), then the geometric dimension of eigenspace for that eigenvalue might be less than r_j (or equal but never more). But for every eigenvalue, 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 $\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$:

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

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 $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_n]^\top$.
- The **eigendecomposition** of $\mathbf{A}$ is given by:

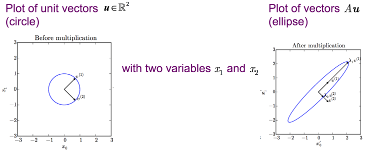
$$\mathbf{AV} = \mathbf{V}diag(\lambda) \implies \mathbf{A} = \mathbf{V}\text{diag}(\boldsymbol{\lambda})\mathbf{V}^{-1}$$

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 $\mathbf{A}$ can be decomposed into real-valued eigenvectors and eigenvalues:

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

- $\mathbf{Q}$ is an orthogonal matrix of the eigenvectors of $\mathbf{A}$, and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues.
- We can think of $\mathbf{A}$ 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 have multiplicity equal to one.
- If any eigenvalue is zero, then the matrix is **singular**. Because if $\mathbf{v}$ is the corresponding eigenvector we have: $\mathbf{A}\mathbf{v} = 0\mathbf{v} = 0$.

Positive Definite Matrix

- If a symmetric matrix A has the property:

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

  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 $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$.
- SVD is more general than eigendecomposition.
- Every real matrix has a SVD.

SVD Definition (1)

- Write $\mathbf{A}$ as a product of three matrices: $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$.
- 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 $\mathbf{A}\mathbf{A}^\top$.
- Right singular vectors of $\mathbf{A}$ are the eigenvectors of $\mathbf{A}^\top\mathbf{A}$.
- Nonzero singular values of $\mathbf{A}$ are square roots of eigenvalues of $\mathbf{A}^\top\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\top$.
- Numbers on the diagonal of D are sorted largest to smallest and are non-negative ($\mathbf{A}^\top\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\top$ are semipositive definite.).

Matrix norms

- We may define norms for matrices too. We can either treat a matrix as a vector, and define a norm based on an entrywise norm (example: Frobenius norm). Or we may use a vector norm to "induce" a norm on matrices.
- Frobenius norm:

$$\|A\|_F = \sqrt{\sum_{i,j} a_{i,j}^2}.$$

- Vector-induced (or operator, or spectral) norm:

$$\|A\|_2 = \sup_{\|x\|_2=1} \|Ax\|_2 .$$

SVD Optimality

- Given a matrix $\mathbf{A}$, SVD allows us to find its "best" (to be defined) rank-r approximation $\mathbf{A}_r$.
- We can write $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ as $\mathbf{A} = \sum_{i=1}^{n} d_i \mathbf{u}_i \mathbf{v}_i^\top$.
- For $r \leq n$, construct $\mathbf{A}_r = \sum_{i=1}^{r} d_i \mathbf{u}_i \mathbf{v}_i^\top$.
- The matrix $\mathbf{A}_r$ is a rank-r approximation of A. Moreover, it is the best approximation of rank r by many norms:
  - When considering the operator (or spectral) norm, it is optimal. This means that $\|A - A_r\|_2 \leq \|A - B\|_2$ for any rank r matrix B.
  - When considering Frobenius norm, it is optimal. This means that $\|A - A_r\|_F \leq \|A - B\|_F$ for any rank r matrix B. One way to interpret this inequality is that rows (or columns) of A_r are the projection of rows (or columns) of A on the best r dimensional subspace, in the sense that this projection minimizes the sum of squared distances.