CSE 152: Linear Algebra II

Lecturer: Hao Su

Scribed by Owen Jow on October 04, 2018

1 Pseudoinverse

Say we want to solve the matrix equation $\mathbf{Ax} = \mathbf{b}$.

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

We can look at this as solving the system of linear equations

$$\begin{gathered} a_{11}x_1 + \ldots + a_{1n}x_n = b_1 \\ \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n = b_m \end{gathered}$$

which can be overdetermined, exactly determined, or underdetermined, according to the number of equations m versus the number of unknowns n. By extension, the system can have zero, one, or infinitely many solutions.

We call $\mathbf{A}^\dagger$ the Moore-Penrose inverse, or **pseudoinverse**, of $\mathbf{A}$.

- If the system has no solutions, $\mathbf{x} = \mathbf{A}^\dagger\mathbf{b}$ is a least-squares solution [1] to $\mathbf{Ax} = \mathbf{b}$.
- If the system has one solution, $\mathbf{x} = \mathbf{A}^\dagger\mathbf{b} = \mathbf{A}^{-1}\mathbf{b}$ is the one solution to $\mathbf{Ax} = \mathbf{b}$.
- If the system has infinitely many solutions, $\mathbf{x} = \mathbf{A}^\dagger\mathbf{b}$ is the least-norm solution [2] to $\mathbf{Ax} = \mathbf{b}$.

Hence we can use the pseudoinverse to solve any matrix equation.

2 Matrix Rank

The rank of a matrix $\mathbf{A}$ tells us the dimension of its image. If $\mathbf{A} \in \mathbb{R}^{m\times n}$ has rank k, then it maps all of $\mathbb{R}^n$ to a k-dimensional subspace. For example, if it were rank 1 it would map $\mathbb{R}^n$ to a line; if it were rank 2 it would map $\mathbb{R}^n$ to a plane.

[1] $\arg\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2^2$

[2] $\arg\min_{\mathbf{x}} \|\mathbf{x}\|_2$ s.t. $\mathbf{Ax} = \mathbf{b}$

To see this, we can consider $\mathbf{A}$ in terms of its column vectors. Then

$$\begin{bmatrix} | & \dots & | \\ \mathbf{a}_1 & \dots & \mathbf{a}_n \\ | & \dots & | \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n$$

i.e. each output is a linear combination of the column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Then, for instance if the matrix is rank 1, we are always linearly combining vectors which are scales of each other – which of course can only produce a line!

It's kind of definitional that the rank is the dimension of the column space, but still worth registering.

3 Eigenvalues and Eigenvectors

A few notes and definitions:

- An eigenvector of $\mathbf{A}$ is a vector which retains its direction when $\mathbf{A}$ is applied to it.
- The trace, determinant, and rank of $\mathbf{A}$ can all be defined in terms of its eigenvalues.
- The **spectrum** of $\mathbf{A}$ is the set of all of its eigenvalues.
- The **spectral radius** of $\mathbf{A}$ is its maximum eigenvalue according to magnitude.

3.1 Eigendecomposition

If an $n \times n$ matrix $\mathbf{A}$ has n linearly independent eigenvectors, it is diagonalizable and can be written in the form $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$, where $\mathbf{V}$ is a matrix with the eigenvectors $\mathbf{v}_i$ as columns and $\mathbf{D}$ is a diagonal matrix with the eigenvalues λ_i as entries (ordered in the same way as their corresponding eigenvectors).

Why? Because

$$\mathbf{AV} = \begin{bmatrix} | & \dots & | \\ \mathbf{A}\mathbf{v}_1 & \dots & \mathbf{A}\mathbf{v}_n \\ | & \dots & | \end{bmatrix} = \begin{bmatrix} | & \dots & | \\ \lambda_1\mathbf{v}_1 & \dots & \lambda_n\mathbf{v}_n \\ | & \dots & | \end{bmatrix} = \begin{bmatrix} | & \dots & | \\ \mathbf{v}_1 & \dots & \mathbf{v}_n \\ | & \dots & | \end{bmatrix} \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix} = \mathbf{VD}$$

and therefore $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$.

(Spectral theorem.) If a real $n \times n$ matrix is symmetric, it is furthermore *orthogonally diagonalizable* and has n mutually orthogonal eigenvectors with real eigenvalues. We can thus construct an orthogonal matrix $\mathbf{V}$ from these eigenvectors, and have the result that

$$\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1} = \mathbf{V}\mathbf{D}\mathbf{V}^{\mathbf{T}}$$