Applied Linear Algebra in Data Analysis

Matrix Inverses

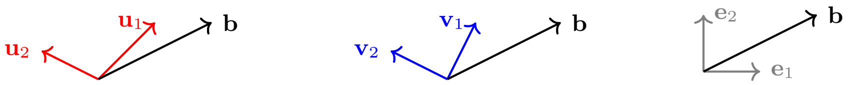
Sivakumar Balasubramanian

Department of Bioengineering
Christian Medical College, Bagayam
Vellore 632002

Representation of vectors in a basis

- Consider the vector space $\mathbb{R}^n$ with basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n\}$. Any vector in $\mathbf{b} \in \mathbb{R}^n$ can be representated as a linear combination of $\mathbf{v}_i$s,

$$\mathbf{b} = \sum_{i=1}^{n} \mathbf{v}_i \mathbf{a}_i = \mathbf{V}\mathbf{a}; \;\; \mathbf{a} \in \mathbb{R}^n, \;\; \mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{bmatrix} \in \mathbb{R}^{n \times n}$$



$\{\mathbf{v}_1, \mathbf{v}_2\}$, $\{\mathbf{u}_1, \mathbf{u}_2\}$ and $\{\mathbf{e}_1, \mathbf{e}_2\}$ are valid basis for $\mathbb{R}^2$, and the presentation for $\mathbf{b}$ in each one of them is different.

Matrix Inverse

- Consider the equation $\mathbf{A}\mathbf{x} = \mathbf{y}$, where $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
- Let us assume $\mathbf{A}$ is non-singular $\implies$ columns of $\mathbf{A}$ represent a basis for $\mathbb{R}^n$.
- What does $\mathbf{x}$ represent? It is the representation of $\mathbf{y}$ in the basis consisitng of the columns of $\mathbf{A}$.

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} \mathbf{a}_i x_i$$

$$\implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y} = \begin{bmatrix} \tilde{\mathbf{b}}_1^\top \\ \tilde{\mathbf{b}}_2^\top \\ \dots \\ \tilde{\mathbf{b}}_n^\top \end{bmatrix} \mathbf{y} = \begin{bmatrix} \tilde{\mathbf{b}}_1^\top \mathbf{y} \\ \tilde{\mathbf{b}}_2^\top \mathbf{y} \\ \dots \\ \tilde{\mathbf{b}}_n^\top \mathbf{y} \end{bmatrix}$$

Matrix Inverse

- $\mathbf{A}^{-1}$ is a matrix that allows change of basis to the columns of $\mathbf{A}$ from the standard basis!

Left Inverse

- Consider a rectangular matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. There exists no inverse $\mathbf{A}^{-1}$ for this matrix.

- But, there exist two matrices $\mathbf{B}, \mathbf{C} \in \mathbb{R}^{n \times m}$, such that,

$$\mathbf{CA} = \mathbf{I}_n \quad \text{or} \quad \mathbf{AB} = \mathbf{I}_m$$

- Both cannot be true for a rectangular matrix, only one can be true when the matrix is full rank.

- A rectangular matrix can only have either a left or a right inverse.

Right Inverse

- For $\mathbf{A} \in \mathbb{R}^{m \times n}$, $n > m$ with full rank, $\mathbf{AB} = \mathbf{I}_m \longrightarrow \mathbf{B}$ is the right inverse.

- Right inverse of $\mathbf{A}$ exists only if the rows of $\mathbf{A}$ are independent, i.e. $rank\,(\mathbf{A}) = m$ $\longrightarrow \mathbf{A}^\top \mathbf{x} = \mathbf{0} \implies \mathbf{x} = \mathbf{0}$

- $\mathbf{Ax} = \mathbf{b}$ can be solved for any $\mathbf{b}$. $\mathbf{x} = \mathbf{Bb} \implies \mathbf{A}\,(\mathbf{Bb}) = \mathbf{b}$.

- There are an infitnite number of $\mathbf{B}$s $\implies$ an infinite number of solutions $\mathbf{x}$.

Pseudo Inverse

- Consider a tall matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with independent columns. It turns out the Gram matrix $\mathbf{A}^\top \mathbf{A} \in \mathbb{R}^{n \times n}$ is invertible. If that is the case then,

$$\left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top \mathbf{A} = \mathbf{I}_n; \quad \left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top \text{ is a left inverse.}$$

- $\mathbf{A}^\dagger = \left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top$ is called the *pseudo inverse* or the *Moore-Penrose inverse*.

- For the case of a fat, wide matrix, we have $\mathbf{A}^\dagger = \mathbf{A}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1}$.

- When $\mathbf{A}$ is square and invertible, $\mathbf{A}^\dagger = \mathbf{A}^{-1}$.

Matrix Inverse and Pseudo Inverse through QR factorization

- Consider an invertible, square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$.

$$\mathbf{A} = \mathbf{QR} \implies \mathbf{A}^{-1} = (\mathbf{QR})^{-1} = \mathbf{R}^{-1}\mathbf{Q}^{-1} = \mathbf{R}^{-1}\mathbf{Q}^{\top}$$

where, $\mathbf{R}, \mathbf{Q} \in \mathbb{R}^{n\times n}$. $\mathbf{R}$ is upper triangular, and $\mathbf{Q}$ is an orthogonal matrix.

- In the case of a left invertible rectangular matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, we can factorize $\mathbf{A} = \mathbf{QR}$, with $\mathbf{Q} \in \mathbb{R}^{m\times n}$ and $\mathbf{R} \in \mathbb{R}^{m\times m}$.

$$\mathbf{A}^{\dagger} = \left(\mathbf{A}^{\top}\mathbf{A}\right)^{-1}\mathbf{A}^{\top} = \left(\mathbf{R}^{\top}\mathbf{Q}^{\top}\mathbf{Q}\mathbf{R}\right)^{-1}\mathbf{R}^{\top}\mathbf{Q}^{\top} = \left(\mathbf{R}^{\top}\mathbf{R}\right)^{-1}\mathbf{R}^{\top}\mathbf{Q}^{\top} = \mathbf{R}^{-1}\mathbf{Q}^{\top}$$

Matrix Inverse and Pseudo Inverse through QR factorization

- For a right invertible wide, fat matrix, we can find out the pseudo-inverse of $\mathbf{A}^\top$, and then take the transpose of the pseudo-inverse.

$$\mathbf{A}\mathbf{A}^\dagger = \mathbf{I} \implies \left(\mathbf{A}^\dagger\right)^\top \mathbf{A}^\top = \left(\mathbf{A}^\top\right)^\dagger \mathbf{A}^\top = \mathbf{I}$$

$$\mathbf{A}^\top = \mathbf{Q}\mathbf{R} \implies \left(\mathbf{A}^\top\right)^\dagger = \mathbf{R}^{-1}\mathbf{Q}^\top = \left(\mathbf{A}^\dagger\right)^\top \implies \mathbf{A}^\dagger = \mathbf{Q}\mathbf{R}^{-T}$$

What about when $\mathbf{A}$ is not full rank?

- There is no left or right inverse for $\mathbf{A} \in \mathbb{R}^{m \times n}$, when $rank(\mathbf{A}) = r < \min(m, n)$.

$$\nexists\ \mathbf{B} \in \mathbb{R}^{n \times m}, s.t.\ \mathbf{BA} = \mathbf{I}_m \text{ or } \mathbf{AB} = \mathbf{I}_n$$

- **$\mathbf{A}$ is tall**: First r columns of $\mathbf{A}$ are linear independent, then $\exists\ \mathbf{B} \in \mathbb{R}^{n \times m}$, $s.t.$

$$\mathbf{BA} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

- **$\mathbf{A}$ is fat**: First r rows of $\mathbf{A}$ are linear independent, then $\exists\ \mathbf{B} \in \mathbb{R}^{n \times m}$, $s.t.$

$$\mathbf{AB} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

What about when $\mathbf{A}$ is not full rank?

- What if we have a linear system of equations with a non-full rank matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$?

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

- $\mathbf{b} \in \mathcal{C}(\mathbf{A}) \implies$ There are infinitely many solutions to the above equation.

- $\mathbf{b} \notin \mathcal{C}(\mathbf{A}) \implies$ There is no solution to the above equation. But there are infinitely many solutions $\hat{\mathbf{x}}$ that minimize $\|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}\|_2$.

- One approach to solve the case wherre $\mathbf{b} \notin \mathcal{C}(\mathbf{A})$ is to formualte the problem as a regularized least squares problem,

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_2^2$$