Mathematics for Machine Learning

Lecture 4
(02.05.2024)

Linear Algebra and Analytic Geometry

Mahammad Namazov

Table of contents

Inverse Matrices

Significance

Algorithmic approach

Row-Factor

Significance

- Applicable when you have square matrix:
  - Output vector will be in the same shape as input vector does;
  - In the best-case scenario no information loss;
  - Otherwise, we cannot recover the path
- What happens when we have non-square matrix?
  - We find pseudo-inverse by using Singular Value Decomposition;
  - Pseudo-inverse or something very similar to inverse;
- There are 3 ways of computing inverse:
  - Using Adjoint matrices (algorithmic approach);
  - With Row factor operations (i.e., Gaussian elimination);
  - Using libraries (e.g., np.linalg.inv(.))

Adjoint matrices for Inverse

- For simplicity we will apply algorithm for 3x3 matrix: ⟹ $A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}$
- For a given matrix A, there will be 9 minors: ⟹ $m_{2,2} = \begin{bmatrix} a_{1,1} & a_{1,3} \\ a_{3,1} & a_{3,3} \end{bmatrix}$
- Compute cofactors using minors: ⟹ $c_{i,j} = (-1)^{i+j} \det(m_{i,j})$
- Build a matrix C with these elements: ⟹ $C = \begin{bmatrix} c_{1,1} & c_{1,2} & c_{1,3} \\ c_{2,1} & c_{2,2} & c_{2,3} \\ c_{3,1} & c_{3,2} & c_{3,3} \end{bmatrix} = [Adj(A)]^T$
  - Transpose of C will be Adjoint matrix of A:
- Use determinant of A and C to compute A^{-1} ⟹ $A^{-1} = \frac{1}{\det(A)} C^T$

Inverse with RF

- Significant operations to optimize matrix representation;

$$\boldsymbol{w} = A\boldsymbol{v} \Rightarrow \boldsymbol{v} = A^{-1}\boldsymbol{w}$$

- Thus:

$$A^{-1}A = I$$

- Steps for applying Row Factor operations to get inverse of the given matrix A:
  - Build the matrix with shape of $m \times k$, where $k = 2n$ and n is number of columns of A;
  - The first n columns will be A's columns, the rest of the matrix will be $I_{m\times n}$, where $m = n$;
  - Apply RF operations, until the left side becomes Identity Matrix;
  - Then, the right hand-side will be inverse matrix of A;

$$(A|I) = \left(\begin{array}{ccc|ccc} a_{1,1} & \dots & a_{1,n} & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,n} & 0 & \dots & 1 \end{array}\right) \xrightarrow{\text{Apply RF until}} \left(\begin{array}{ccc|ccc} 1 & \dots & 0 & a'_{1,1} & \dots & a'_{1,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & a'_{n,1} & \dots & a'_{n,n} \end{array}\right) = (A^{-1}|I)$$

Spaces and Rank

Column Space

Null Space

Row Space

Rank

Matrix Representation

- To do linear transformation from $V \in \mathbb{R}^n$ into $W \in \mathbb{R}^m$, we need a matrix M:

$$M_{m\times n}: V \to W$$

- Each column is **supposed** to represent the new place of basis vectors (Why supposed to?);
- Number of columns tells us what is the dimensions of input vector space;
- Number of rows tells us what is the dimensions of output vector space;
- There are 3 possible scenarios to analyze:

$m = n$

- **Best:** When $\det(M) \neq 0$
- No information loss
- Data recovery is possible

$m < n$

- **Best:** When m independent columns
- Dimensions will be reduced
- Data recovery is possible (approximately)

$m > n$

- **Best:** When n independent columns
- Dimensions will be expanded
- Data recovery is possible (approximately)

Column Space

- When you have a matrix, its columns represents the new ”place” for basis vectors;

- Any vector that **can be transformed** by the given matrix will end up with linear combination of columns of matrix:

$$A_{m\times n}\boldsymbol{v}_{n\times 1}=\begin{bmatrix}\vdots & \vdots & \dots & \vdots & \vdots \\ \boldsymbol{a_1} & \boldsymbol{a_2} & \dots & \boldsymbol{a_{n-1}} & \boldsymbol{a_n} \\ \vdots & \vdots & \dots & \vdots & \vdots\end{bmatrix}\begin{bmatrix}v_{1,1}\\ \vdots \\ v_{n,1}\end{bmatrix}=\boxed{v_{1,1}\boldsymbol{a_1}+\cdots+v_{n,1}\boldsymbol{a_n}=\boldsymbol{w}_{m\times 1}}$$

- Column space of a matrix is span of its column vectors:

$$Col(A)=span(\boldsymbol{a_1},\boldsymbol{a_2},\dots,\boldsymbol{a_n}),\boldsymbol{a_i}\in\mathbb{R}^m,v_i\in\mathbb{R}$$

- ***Takeaway:*** Column space is a subspace of **destination vector space**

- Subspace is a space itself, thus it should have some basis vectors:
  - Use Gaussian Elimination
  - Pivot columns will form the set of basis vectors for column space

Null Space

- After applying a linear transformation, there are some vectors that are transformed into zero vector in the destination vector space;
- The space that consists of those vectors is called null space;
- Notice, we are not analyzing the destination but input space.
- Thus, Null Space of linear transformation will be a subspace of the input space:

$$Null(A) = \{\boldsymbol{x} \in \mathbb{R}^n | A\boldsymbol{x} = \boldsymbol{0}\}$$

- How to get basis vectors of this space?
  - We already collected pivot columns;
  - Using non-pivot columns, we can get basis vectors of Null space;

- **Reminder:** $A: V \to W$
- A has m rows and n columns;
- $V \subseteq \mathbb{R}^n$ and $W \subseteq \mathbb{R}^m$

Row space

- Let's analyze linear transformation as system of equations;
- Column space defines where the unknown variables land, that is why number of columns equal to number of variables;
- But they must follow an order and be in organized representations. In engineering language, these variables must satisfy some constraints;
- These constraints are defined by rows of matrices, which spans the row space of linear transformation.
- Independent rows represents maximum possibility of constraints to be satisfied;
- ***Takeaway:*** Number of basis vectors of row space always equals to number of basis vectors of column space

Rank

- A number represents number of the maximum possible independent columns (or rows);
- Row rank represents number of independent rows;
- Column rank represents number of independent columns;
- Takeaway: Row rank always equals to column rank;
- Full rank, is when number of independent columns (or rows) equals to $\min(m, n)$;
  - Square matrix: $\min(m, n) = m = n$;
  - Rectangular matrix, where $m > n \Rightarrow \min(m, n) = n$;
  - Rectangular matrix, where $m < n \Rightarrow \min(m, n) = m$;

System of Linear Equations

Representation

Scenarios

Solutions

Problem Definition

- Assuming we have $\boldsymbol{w}$ in some vector space W, and transformation matrix $A_{m\times n}$

$$A\boldsymbol{v} = \boldsymbol{w}$$

- What $\boldsymbol{v}$ was transformed into $\boldsymbol{w}$?

- We can find the vector $\boldsymbol{v}$ by solving several linear equations (i.e., System of LE):

- There are several scenarios;
  - **Reminder:** You will see the insight not the rules!

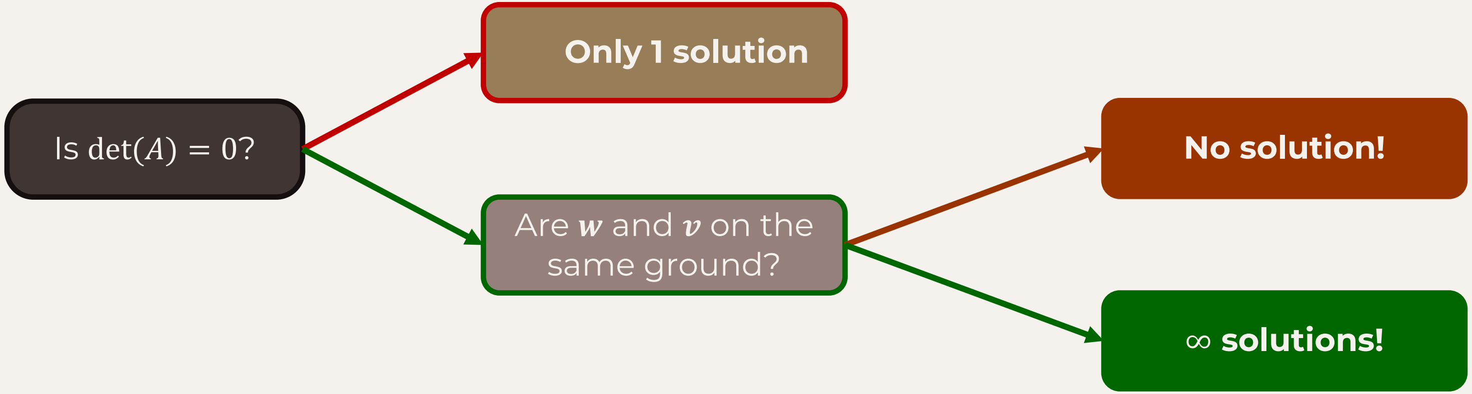
$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix} \begin{bmatrix} v_{1,1} \\ \vdots \\ v_{1,n} \end{bmatrix} = \begin{bmatrix} w_{1,1} \\ \vdots \\ w_{n,1} \end{bmatrix}$$

$$\Longleftrightarrow$$

$$\begin{gathered} a_{1,1}v_{1,1} + a_{1,1}v_{2,1} + \cdots + a_{1,n}v_{n,1} = w_{1,1} \\ a_{2,1}v_{1,1} + a_{2,2}v_{2,1} + \cdots + a_{2,n}v_{n,1} = w_{2,1} \\ \vdots \\ a_{m,1}v_{1,1} + a_{m,2}v_{2,1} + \cdots + a_{m,n}v_{n,1} = w_{n,1} \end{gathered}$$

A is a square matrix

- Starting with this scenario ($m = n$) will be very helpful to imagine
- Think in that way: **Number of equations is equal to number of variables**



Overview

- Non-zero Determinant: Linear Independent columns, full rank
- Having zero Determinant: Linear Dependent columns

$\rho \neq \rho_{full}$

- $\boldsymbol{v}$ and $\boldsymbol{w}$ share same ground: They are linear dependent
- Ground can be:
  - Line for 2D
  - Line or Plane for 3D
  - Line or Plane or Volume for 4D and so on
- Let’s continue!

A is a rectangular matrix (m>n)

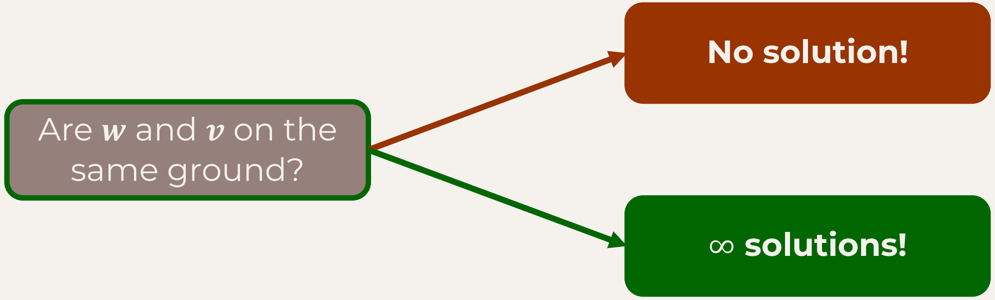
- Let's generalize the scenario for $A\boldsymbol{v} = \boldsymbol{w}$:
  - We generalize the ideal case, it might be different
- When number of equations is greater than number of variables:
- Most likely that you have redundant $k < m$ equations;
  - i.e., having or not having does not matter for you
- After removing them you will have 2 scenarios:

$m - k = n$:
We have a Paradise!

$\boldsymbol{m - k < n}$:
We have an Insufficiency

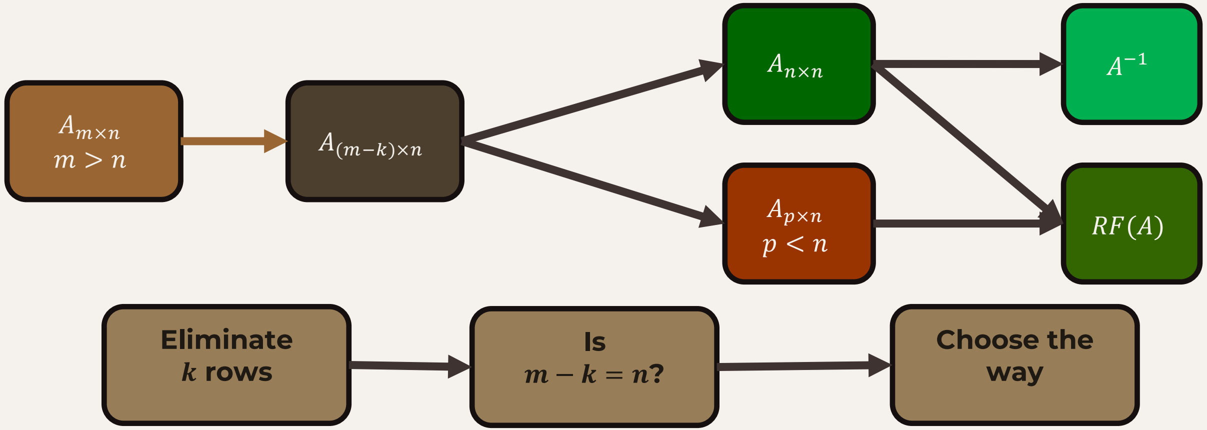
A is rectangular matrix (m<n)

- For paradise scenario, we can find **the** solution;
- In the other scenario, you will have insufficient constraints to find exact solution;
- The most possible rank that you will have will be new value of rows (it is the minimum);
- Dropping these linear dependent rows, we get A in the shape of $m{\times}n$



To sum up …

- We talk generally, rather than specific solutions;
- Combining what we have said earlier in a map:



Computation Strategy

- For the problem that is given in the form of $A\boldsymbol{v} = \boldsymbol{w}$, we investigate:
  - What $\boldsymbol{v}$ has transformed into $\boldsymbol{w}$, with the given matrix A

- In case matrix is in a square form, you have the following steps to do:
  - Check the determinant, if it is nonzero then use one of these:
    - ✓ Use inverse matrix:
    $$A^{-1}A\boldsymbol{v} = A^{-1}\boldsymbol{w} \Rightarrow \boldsymbol{v} = A^{-1}\boldsymbol{w}$$
    - ✓ Apply row factor operations:
    $$(A|\boldsymbol{w}) \Rightarrow (M|\boldsymbol{w}')$$
  - If the determinant is zero, it means you have k linear dependent rows, then follow these steps:
    - ✓ Remove k-1 linear dependent rows (Why?);
    - ✓ Use row factor operations;

- Otherwise, you have only one choice: Row factor operations

Analytic Geometry

Norms/Distances

Inner Product

Orthogonalization

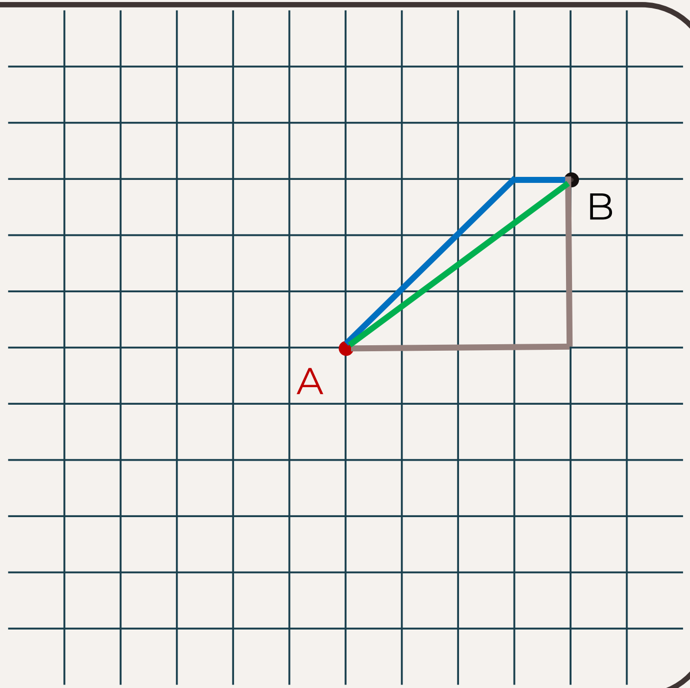
Norms and Distances

- When we talk about distance, it requires two points;
- Requirements for a metric to be a distance, which will be applied to the elements of the same space:
  - If distance between two points is zero, then these points must be the same;
    $$d(A,B) = 0 => A = B$$
  - There is unique distance between two points;
    $$d(A,B) = d(B,A)$$
  - If you want to go from A to B, through C instead of going directly, the resulting distance cannot be smaller than direct approach;
    $$d(A,B) \leq d(A,C) + d(C,B)$$

Types of Distances

- Assume you are given two points on 2-D plane and you have A and B points;
- What is the distance between A and B?

5	4	7
Euclidean	Chebyshev	Manhattan



Inner Product

- The most well-known and mostly used inner product is dot-product (i.e., scalar product)

$$\boldsymbol{a} \cdot \boldsymbol{b} = \|\boldsymbol{a}\|\|\boldsymbol{b}\| cos(\alpha), \forall \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$$

- However, it is not the only inner product;

- An operation between two vectors (in our case) can be inner product if following constraints are satisfied:
  - Commutativity: $(\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{b}, \boldsymbol{a})$
  - Distributivity: $(\boldsymbol{a}, \boldsymbol{b} + \gamma \boldsymbol{c}) = (\boldsymbol{a}, \boldsymbol{b}) + \boldsymbol{\gamma}(\boldsymbol{a}, \boldsymbol{c})$
  - Positive definiteness: $(\boldsymbol{a}, \boldsymbol{a}) \neq 0, if\ \boldsymbol{a} \neq \boldsymbol{0}$

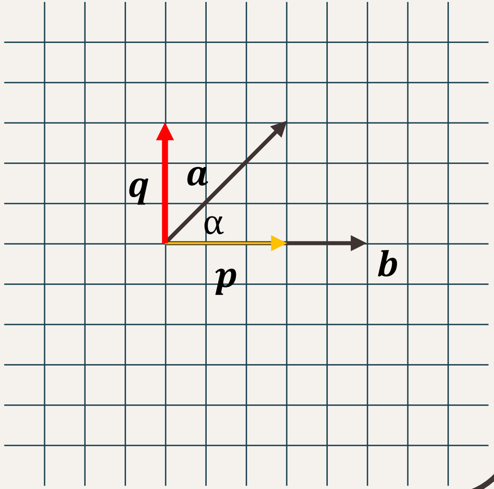
Projection

- From now on, we will use inner product representation as $(\boldsymbol{a}, \boldsymbol{b})$;
- Now assume these given vectors are in 2D-plane as expressed:
- Then projection of $\boldsymbol{a}$ onto $\boldsymbol{b}$ vector can be expressed using inner products:

$$\boldsymbol{p} = \frac{(\boldsymbol{a}, \boldsymbol{b})}{(\boldsymbol{b}, \boldsymbol{b})} \boldsymbol{b}\, cos(\alpha)$$

- We can also represent $\boldsymbol{q}$ using this representation:

$$\boldsymbol{q} = \boldsymbol{a} - \frac{(\boldsymbol{a}, \boldsymbol{b})}{(\boldsymbol{b}, \boldsymbol{b})} \boldsymbol{b}\, cos(\alpha)$$



Orthogonalization

- Assume you are in 3-D space, and you have 3 orthogonal vectors as basis vectors:

$$\boldsymbol{v} = a_1\boldsymbol{e_1} + a_2\boldsymbol{e_2} + a_3\boldsymbol{e_3}, \boldsymbol{e_i} \in \mathbb{R}^3, a_i \in \mathbb{R}$$

- Then for each a_i, we can say:

$$a_i = \frac{(\boldsymbol{v}, \boldsymbol{e_i})}{(\boldsymbol{e_i}, \boldsymbol{e_i})}$$

- When your basis vectors are not orthogonal, you cannot do that;
- Thus, orthogonalization brings a little bit simpler way of decomposition;

Gram-Schmidt Orthogonalization

- Now assume you are in n-dimensional space, where set of basis vectors are as follows:
$$B = \{e_1, e_2, \dots, e_n\}$$
- As we know any vector in the vector space can be written as a linear combination of basis vectors:
$$v = a_1\boldsymbol{e_1} + \dots + a_n\boldsymbol{e_n}$$
- We also know that if the basis vectors are orthogonal this decomposition coefficient cannot be easily computed;
- Remember this equation:
$$\boldsymbol{q} = \boldsymbol{a} - \frac{(\boldsymbol{a}, \boldsymbol{b})}{(\boldsymbol{b}, \boldsymbol{b})}\boldsymbol{b}\, cos(\alpha)$$

Conclusion

Summary

Takeaways

References

Summary

- Inverse matrices: “Hansel and Gretel” for Linear Algebra, but works only for square matrices that have non-zero determinant;
- When we have rectangular matrices, we can use pseudo-inverse, but it does not guarantee recovery of all details;
- In LSE, start with rectangular $A_{m \times n}$ $(m > n)$ go with removing independent rows;
- Three main constraints must be satisfied by a metric to be a distance;
- Three main axioms must be satisfied by an operation to be an inner product;
- Orthogonalization is significant for simpler decomposition

Takeaways

- Column space of a matrix, is a space that is spanned by columns of the matrix;
- Null space of a matrix consists all vectors are transformed into zero vector by the given matrix;
- Column Rank = Row rank = Rank
- When you have rectangular matrix where $m \neq n$, full rank will be $\min(m, n)$;
- There is a duality between dot product and projection;
- When we are given orthogonal basis vectors and a vector from the same space, each coefficient of each vector can be found using corresponding basis vector and vector itself;

References

- Further information to read:
  - Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for machine learning*. Cambridge University Press.
  - Chapter 2, Sections: 1, 2, 3, 6;
  - Chapter 3, Sections: all but 7 and 8;

The End

Thanks for your attention and patience!

Mahammad Namazov