

Linear algebra

$\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \mathbb{R} \cdots \times \mathbb{R}$ is a the cartesian product of the group (from group theory) of real numbers, n times.

Vector Space

A vector space is a set V with with binary operations $+$, \cdot where $+$ is between elements (vectors) of V and \cdot is between elements of V and elements (scalars) of \mathbb{R} , such that :

- $(V, +)$ is an Abelian Group.
- $\forall \lambda \in \mathbb{R}, x, y \in V, \lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y$
- $\forall \lambda, \psi \in \mathbb{R}, x \in V, (\lambda + \psi) \cdot x = \lambda \cdot x + \psi \cdot x$
- $\forall \lambda, \psi \in \mathbb{R}, x \in V, \psi \cdot (\lambda \cdot x) = \psi \lambda \cdot x$
- $\forall x \in V, 1 \cdot x = x$.

Dimension

If for a vectors space V , the group $(V, +) \cong \mathbb{R}^n$, then $\dim(V) = n$ is the dimension of the vector space.

Thus the set of all matrices in $\mathbb{R}^{m \times n}$ over addition and scalar multiplication is also a vector space of dimension mn . (In case you still haven't got it, we are talking about image data) .

Subspaces

A vector subspace is just a subgroup of a vector space.

Affine spaces

For any subspace $U \subseteq V$, any coset $v_i + U$ of the supspace is called an affine space.

Affine space containing affine space

For an affine space $v_1 + U_1$ to be a subset of another affine space $x_2 + U_2$, we require $\forall u_1 \in U_1, \exists u_2 \in U_2, s.t. v_1 + u_1 = v_2 + u_2 \implies u_1 = (v_2 - v_1) + u_2$
 $\implies \exists u_2 \in U_2, s.t. 0 = (v_2 - v_1) + u_2 \iff u_2 = v_1 - v_2$
 $\implies (v_1 - v_2) \in U_2$
 $\implies \forall u_1 \in U_1, u_1 = -(v_2 - v_1) + u_2 \in U_2$ for some $u_2 \in U_2$.

Thus we arrive at $U_1 \subseteq U_2$.

This condition is equivalent to the bases of U_1 being in U_2 .

Solving a system of equations with rectangular matrices

Consider the equation $\mathbf{Ax} = \mathbf{c}$. To solve this, find a specific solution \mathbf{x}_0 and as many linearly independent solutions $\mathbf{x}_1, \mathbf{x}_2, \dots$ to $\mathbf{Ax} = \mathbf{0}$ as possible. Then the general solution is $\mathbf{x} = \mathbf{x}_0 + \lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \dots$

To find $\mathbf{x}_1, \mathbf{x}_2, \dots$ quickly, first do Gaussian elimination to get and then express the column vectors with multiple non zero entries, say the j^{th} column as a linear combination of columns with only 1 (and 0s) as entries. Then just put the coefficients as coordinates for \mathbf{x}_i and put -1 as the j^{th} coordinate.

Example : Suppose after row operations, we have arrived at this matrix :

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 1 & -4 \end{bmatrix}$$

Then using our method for the 2nd column, we get this vector :

$$\mathbf{x}_1 =$$

$$\begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

And doing it for the 5th column vector, we get $\mathbf{x}_2 =$

$$\begin{bmatrix} 3 \\ 0 \\ 9 \\ -4 \\ -1 \end{bmatrix}$$

This way you can solve the equation when $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a horizontal matrix ($m < n$). When it's a vertical matrix, we simply chop off the lower rows after Gaussian elimination. Another way to solve a system like this is to apply the Moore-Penrose pseudo-inverse on both sides, given by $(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$. Be warned though..computing $\mathbf{A}^T \mathbf{A}$ is not easy.

This solution is also the one we get when we do the least squares approximation.

Least squares

Consider \mathbf{A} to be a list of row vectors \mathbf{r}_i which are really points in some n dimensional space and \mathbf{c} to be a list of values closed to ones assigned to these points by some linear function, say $c_i \approx f(\mathbf{r}) = x_1 r_1 + x_2 r_2 + \dots = \mathbf{x} \cdot \mathbf{r}$. We want to find the \mathbf{x} that gives the best approximation and reduces the quantity $E(\mathbf{x}) = \sum_i (c_i - f(\mathbf{r}_i))^2$.

We consider $(\mathbf{Ax} - \mathbf{c})^2$ as a function of coordinates of \mathbf{x} and try to minimise it. It is easy to see that for the minima, we have $(\mathbf{Ax} - \mathbf{c})^T (\mathbf{Ax} - \mathbf{c}) = 0$. These equations written one below other give rise to $(\mathbf{Ax} - \mathbf{c})^T \mathbf{A} = \mathbf{0}$. Taking transpose on both sides, we get $\mathbf{A}^T (\mathbf{Ax} - \mathbf{c}) = \mathbf{0}$
 $\implies \mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{c}$
 $\implies \mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{c}$.

This method is rather computation heavy. In practice, we set up an iteration of the form : $\mathbf{x}_{(k+1)} = \mathbf{C} \mathbf{x}_k + \mathbf{d}$ that decreases $E(\mathbf{x}_k)$ in every iteration.

linear mapping

A **homomorphism** (function distributive over group operation (addition here)), say, $f : V \rightarrow W$, between vector spaces is called a linear mapping if it meets this extra criterion : $\forall \lambda \in \mathbb{R}, \mathbf{x} \in V, f(\lambda \mathbf{x}) = \lambda f(\mathbf{x})$.

Because of the linearity, we can represent such a function as a matrix.

Suppose $f : V \rightarrow W$ is a linear map with V, W expressed using the basis $B = \{\mathbf{b}_1, \mathbf{b}_2 \dots\}$ and $C = \{\mathbf{c}_1, \mathbf{c}_2 \dots\}$, then f can be expressed by a matrix with the i^{th} column vector given by the coordinates of $f(\mathbf{b}_i)$ as expressed using C .

Note that any element of V, W isn't simply a list of coordinates but a full physical vector that requires a basis to be expressed as coordinates in.

Rank and nullity

Since a linear map $f : V \rightarrow W$ is a homomorphism, according to the first isomorphism theorem, we have $V / \ker f \cong \text{im}(f)$ and thus $V \cong \text{im}(f) \times \ker f$, which tells us $\dim(V) = \dim(\text{im}(f)) + \dim(\ker f)$

Define $\dim(\text{im}(f)) = \text{rk}(A_f)$ as the rank of the matrix that represents f in some basis and define $\dim(\ker f) = \text{nul}(A_f)$ as the nullity (The kernel is also called the null space, since it gets mapped to 0). It's easy to see that the rank is number of linearly independent column vectors of A_f and thus $\text{im}(f)$ is also called the column space of A_f , which is generated by these linearly independent columns. And since $\dim(V)$ is basically the number of columns in A_f , the nullity is the the number of remaining columns.

Column rank and row rank are the same.

Define column rank to the number of independent columns and row rank to be number of independent rows for any matrix.

Since elementary row operations do not change the column rank or row rank (poof left as exercise), thus after Guassian elimination, these quantities will still be same as of the original matrix. And for a reduced row echolon form, both the column and row ranks are just the number of non zero entries in the marix, and thus they are the same.

Change of basis

If a vector space V is expressed using two diffent basis $B = \{\mathbf{b}_1, \mathbf{b}_2 \dots\}$ and $B' = \{\mathbf{b}'_1, \mathbf{b}'_2 \dots\}$.

Then to go from representation of a vector using B' to one using B , we have to simply multiply by

a matrix with column vectors being \mathbf{b}'_i expressed using B .

Suppose the matrices \mathbf{B}, \mathbf{B}' have column vectors as $\mathbf{b}_i, \mathbf{b}'_i$ expressed using a basis G , then the matrix for changing basis from B' to B is given by $\mathbf{B}^{-1}\mathbf{B}'$.

Now, consider a linear map $f : V \rightarrow W$ given by the matrix \mathbf{A} when expressed using basis B, C for V, W and the same map can also be given by the matrix \mathbf{A}' when expressed using basis B', C' for V, W respectively. Suppose \mathbf{S} is the matrix that changes the basis from B' to B and \mathbf{T} changes from C' to C . Then,

$$\mathbf{A}' = \mathbf{T}^{-1}\mathbf{A}\mathbf{S}$$

Expressing everything in some universal frame, we can write :

$$\mathbf{A}' = ((\mathbf{C}')^{-1}\mathbf{C}) \mathbf{A} (\mathbf{B}^{-1}\mathbf{B}')$$

(Note : Read right to left please)