Example 1

Let

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Then

$$\operatorname{span}\{v_1, v_2\} = \left\{ \begin{pmatrix} a \\ b \\ 0 \end{pmatrix} : a, b \in \mathbb{R} \right\}.$$

This is a plane in $\mathbb{R}^3$.

Subspaces

Def. A set $L \subseteq \mathbb{R}^n$ is a linear subspace if it is closed under linear combinations.

Equivalently,

$$x, y \in L, \qquad \alpha, \beta \in \mathbb{R}$$

implies

$$\alpha x + \beta y \in L.$$

The note writes this equivalently as

$$\operatorname{span}(L) = L.$$

An important fact is:

$$\boxed{\operatorname{span}(S) \text{ is always a subspace.}}$$

The proof is simple: a linear combination of linear combinations of vectors from S is still just a linear combination of vectors from S. This is the main observation behind Lemma 2.3.1 in the notes.

Hence,

$$\text{ColumnSpace}(A)$$

is always a subspace of $\mathbb{R}^m$.

If A has 10,000 columns, does that necessarily mean its outputs require 10,000 independent directions?

We want to distinguish between vectors that genuinely give us new directions and vectors that are redundant.

The vectors

$$v_1, \dots, v_k$$

are linearly independent if

$$\boxed{\text{No vector can be written as a linear combination of the others.}}$$

or

$$\alpha_1 v_1 + \cdots + \alpha_k v_k = 0$$

implies

$$\boxed{\alpha_1 = \cdots = \alpha_k = 0.}$$

Otherwise, they are linearly dependent.

Example:

Consider

$$v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

the set is linearly dependent.

so

$$\operatorname{span}\{v_1, v_2, v_3\} = \operatorname{span}\{v_1, v_2\}.$$

Thus v_3 contributes no new direction.

Matrix interpretation

If the columns of

$$A = \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix}$$

are linearly dependent, then some columns are redundant for describing the column space.

This is already beginning to sound like compression.

Basis

A basis of a subspace L is a nonredundant set of directions that generates the entire subspace. Mathmatically, it satisfies two properties:

The vectors span L. The vectors are linearly independent.

For example,

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

form the standard basis of $\mathbb{R}^3$.

Basis is not unique.

For example,

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

also form a basis of $\mathbb{R}^2$.

The important thing is:

Different bases may contain different vectors, but they always contain the same number of vectors.

Dimension

$$\boxed{\dim(L) = \text{number of vectors in any basis of } L.}$$

Examples:

$$\dim(\mathbb{R}^n) = n.$$

A line through the origin in $\mathbb{R}^3$ has dimension 1.

A plane through the origin has dimension 2.

The zero subspace

$$\{0\}$$

has dimension 0.

Rank Now return to our matrix.

Its column space

$$C(A)$$

is some subspace of $\mathbb{R}^m$. Define

$$\boxed{\operatorname{rank}(A) = \dim(C(A)).}$$

That is:

$$\boxed{\operatorname{rank}(A) = \text{number of independent directions among the columns of } A.}$$

Therefore,

$$\operatorname{rank}(A) \leq \min(m, n).$$

If

$$A \in \mathbb{R}^{1000\times 1000}$$

but

$$\operatorname{rank}(A) = 10,$$

then all 1000 columns actually lie in a 10-dimensional subspace.

So although A contains one million entries, its columns are generated using only ten independent directions.

This is where the compression problem becomes concrete.

Low Rank Means Factorization

Choose a basis of its column space:

$$u_1, \dots, u_r.$$

Put these basis vectors into a matrix

$$U = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{pmatrix} \in \mathbb{R}^{m\times r}.$$

Because every column a_j of A belongs to $C(A)$, it can be written as

$$a_j = v_{1j}u_1 + \cdots + v_{rj}u_r.$$

Collect these coefficients into vectors

$$v_j = \begin{pmatrix} v_{1j} \\ \vdots \\ v_{rj} \end{pmatrix} \in \mathbb{R}^r.$$

Then

$$a_j = Uv_j.$$

Now collect all these coefficient vectors into

$$V = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix} \in \mathbb{R}^{r \times n}.$$

Therefore,

$$\boxed{A = UV.}$$

This is extremely important:

$$\boxed{\operatorname{rank}(A) = r \quad \Longrightarrow \quad A = UV,}$$

with

$$U \in \mathbb{R}^{m \times r}, \qquad V \in \mathbb{R}^{r \times n}.$$

And conversely, if

$$A = UV$$

with these dimensions, then

$$\operatorname{rank}(A) \leq r.$$

So:

$$\boxed{\text{Low rank} \quad \Longleftrightarrow \quad \text{representation using a small intermediate dimension.}}$$

This is exactly the structure we wanted for network compression.