Lecture 4-5: Why Do We Need More Linear Algebra? Network Compression as a Motivating Problem. Span, Inner product, and Norms.

Previously, we saw that a FNN of the form

$$x \longmapsto \sigma(Wx + b).$$

In practice, W is huge.

One simple idea: compress a matrix

Suppose

$$W \approx UV,$$

where

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

Then

$$Wx \approx U(Vx).$$

Instead of storing mn numbers, we store

$$mr + rn = r(m + n).$$

But now the obvious question is:

$$\boxed{\text{How do we know whether } UV \text{ is close to } W?}$$

Or more generally,

$$\boxed{\min_{\operatorname{rank}(B) \le r} \|W - B\|.}$$

Let

$$A \in \mathbb{R}^{m \times n}, \qquad x \in \mathbb{R}^n.$$

Then

$$Ax = x_1 a_1 + \cdots + x_n a_n.$$

Ax is a linear combination of the columns of A.

Therefore,

$$\text{Range}(A) = \{Ax : x \in \mathbb{R}^n\} = \text{span}\{a_1, \dots, a_n\}.$$

So

$$\boxed{\text{Range}(A) = \text{ColumnSpace}(A).}$$

This will become important when we discuss rank and low-dimensional representations.

Inner Products

For

$$x, y \in \mathbb{R}^n,$$

define

$$\boxed{\langle x, y \rangle = x^T y = \sum_{j=0}^{n-1} x_j y_j.}$$

Recall the basic properties:

$$\langle x, y \rangle = \langle y, x \rangle,$$
$$\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle.$$

$$\langle x, x \rangle = \sum_j x_j^2 \geq 0,$$

with equality only when $x = 0$.

Geometric property: for nonzero x, y,

$$\cos\theta = \frac{\langle x, y \rangle}{\sqrt{\langle x, x \rangle}\sqrt{\langle y, y \rangle}}.$$

or

$$\boxed{\cos\theta = \frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2}.}$$

To prove this, normalize the two vectors:

$$u = \frac{x}{\|x\|_2}, \qquad v = \frac{y}{\|y\|_2}.$$

Then

$$\|u\|_2 = \|v\|_2 = 1,$$

and the angle between u and v is still θ.

Although u and v may lie in a high-dimensional space $\mathbb{R}^n$, they both belong to the at-most two-dimensional subspace

$$\operatorname{span}\{u, v\}.$$

Therefore, ordinary two-dimensional geometry applies.

Consider the triangle whose sides are u, v, and $u - v$. By the law of cosines,

$$\|u - v\|_2^2 = \|u\|_2^2 + \|v\|_2^2 - 2\|u\|_2\|v\|_2 \cos\theta.$$

Since $\|u\|_2 = \|v\|_2 = 1$,

$$\|u - v\|_2^2 = 2 - 2\cos\theta. \tag{1}$$

On the other hand, using the inner product,

$$\begin{aligned} \|u - v\|_2^2 &= \langle u - v, u - v\rangle \\ &= \langle u, u\rangle - 2\langle u, v\rangle + \langle v, v\rangle \\ &= 2 - 2\langle u, v\rangle. \end{aligned} \tag{2}$$

Comparing (1) and (2), we obtain

$$2 - 2\cos\theta = 2 - 2\langle u, v\rangle,$$

so

$$\cos\theta = \langle u, v\rangle.$$

Finally,

$$\begin{aligned} \cos\theta &= \left\langle \frac{x}{\|x\|_2}, \frac{y}{\|y\|_2} \right\rangle \\ &= \frac{\langle x, y\rangle}{\|x\|_2\|y\|_2}. \end{aligned}$$

Hence,

$$\boxed{\cos\theta = \frac{\langle x, y\rangle}{\|x\|_2\|y\|_2}.}$$

In particular,

$$\langle x, y\rangle = 0 \iff \cos\theta = 0 \iff \theta = \frac{\pi}{2}.$$

Therefore,

$$\boxed{x \perp y \iff \langle x, y\rangle = 0.}$$

Therefore,

$$\boxed{\langle x, y\rangle = 0 \iff x \perp y.}$$

Cauchy–Schwarz

$$\boxed{|\langle x, y\rangle| \le \|x\|_2\|y\|_2.}$$

For every $t \in \mathbb{R}$,

$$0 \le \|tx + y\|_2^2.$$

Expanding,

$$\begin{aligned} 0 &\le \langle tx + y, tx + y\rangle \\ &= t^2\|x\|_2^2 + 2t\langle x, y\rangle + \|y\|_2^2. \end{aligned}$$

This quadratic polynomial must be nonnegative for every t.

Therefore its discriminant satisfies

$$(2\langle x, y\rangle)^2 - 4\|x\|_2^2\|y\|_2^2 \le 0.$$

Hence

$$\langle x, y\rangle^2 \le \|x\|_2^2\|y\|_2^2,$$

so

$$\boxed{|\langle x, y\rangle| \le \|x\|_2\|y\|_2.}$$

Euclidean Norm

The inner product immediately gives a notion of length:

$$\boxed{\|x\|_2 = \sqrt{\langle x, x\rangle} = \sqrt{x_0^2 + \cdots + x_{n-1}^2}.}$$

Then

$$\cos\theta = \frac{\langle x, y\rangle}{\|x\|_2\|y\|_2}.$$

Now return to compression for a second:

If two vectors x and $\tilde{x}$ are supposed to be close, a natural error is

$$\|x - \tilde{x}\|_2.$$

Similarly, for matrices we will want quantities such as

$$\|W - B\|.$$

So norms are exactly what we need to make "close" mathematically precise.

General Norm

Now abstract away from ℓ_2.

A norm is a function

$$\|\cdot\| : V \to [0, \infty)$$

satisfying:

1. Positivity
$$\|x\| \geq 0, \qquad \|x\| = 0 \iff x = 0.$$
2. Homogeneity
$$\|\alpha x\| = |\alpha|\|x\|.$$
3. Triangle inequality
$$\boxed{\|x + y\| \leq \|x\| + \|y\|.}$$

Then show why ℓ_2 satisfies triangle inequality:

$$\|x + y\|_2^2 = \|x\|_2^2 + 2\langle x, y\rangle + \|y\|_2^2.$$

By Cauchy–Schwarz,

$$\begin{aligned} &\leq \|x\|_2^2 + 2\|x\|_2\|y\|_2 + \|y\|_2^2 \\ &= (\|x\|_2 + \|y\|_2)^2. \end{aligned}$$

Therefore,

$$\boxed{\|x + y\|_2 \leq \|x\|_2 + \|y\|_2.}$$

Other Norms We Will Need

For a vector,

$$\|x\|_1 = \sum_j |x_j|,$$

$$\|x\|_2 = \left(\sum_j x_j^2\right)^{1/2},$$

$$\|x\|_\infty = \max_j |x_j|.$$

For a matrix,

$$\boxed{\|A\|_F = \left(\sum_{j,k} A_{jk}^2 \right)^{1/2}}$$

is the Frobenius norm.