신경망 네트워크와 수학적 기반

Norm

- the ***Euclidean norm*** (or just *norm*) of an n-vector x is

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{x^T x}$$

- used to measure the size of a vector
- reduces to absolute value for $n = 1$

Norm and Distance

Properties

for any n-vectors x and y, and any scalar β

- *homogeneity:* $\| \beta x \| = | \beta | \cdot \| x \|$
- *triangle inequality*: $\| x + y \| \leq \| x \| + \| y \|$
- *nonnegativity*: $\| x \| \geq 0$
- *definiteness*: $\| x \| = 0$ only if $x = 0$

Norm and Distance
RMS value

- *mean-square value* of *n*-vector *x* is

$$\frac{x_1^2 + \cdots + x_n^2}{n} = \frac{\|x\|^2}{n}$$

- *root-mean-square value* (RMS value) is

$$\mathbf{rms}(x) = \sqrt{\frac{x_1^2 + \cdots + x_n^2}{n}} = \frac{\|x\|}{\sqrt{n}}$$

- RMS value useful for comparing sizes of vectors of different lengths

Norm of block vectors

- suppose a, b, c are vectors
- $\| (a, b, c) \|^2 = a^Ta + b^Tb + c^Tc = \|a\|^2 + \|b\|^2 + \|c\|^2$
- so we have

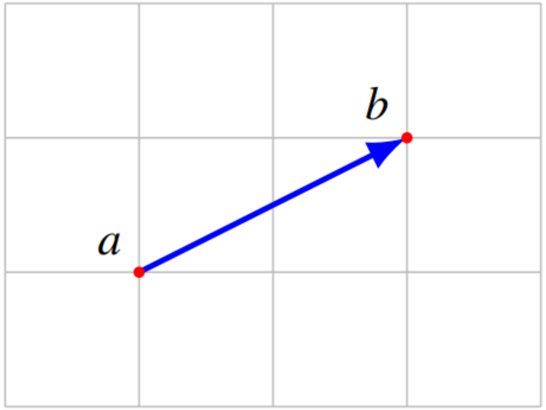
$$\|(a,b,c)\| = \sqrt{\|a\|^2 + \|b\|^2 + \|c\|^2} = \|(\|a\|, \|b\|, \|c\|)\|$$

Distance

- (Euclidean) *distance* between n-vectors a and b is

$$\mathbf{dist}(a,b) = \|a - b\|$$

- agrees with ordinary distance for $n = 1, 2, 3$



- $\mathbf{rms}(a - b)$ is the *RMS deviation* between a and b

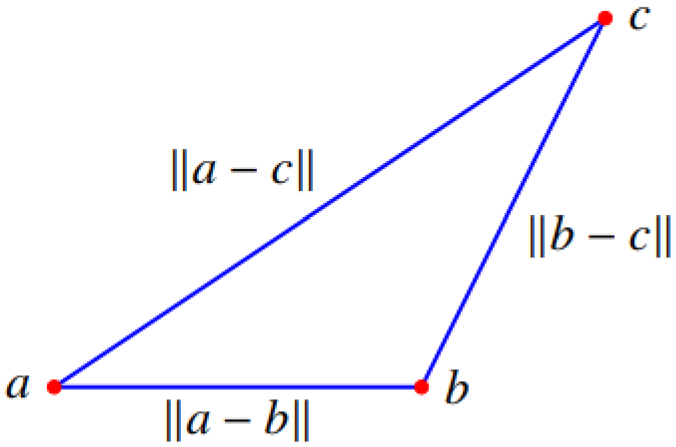
Norm and Distance

Triangle inequality

- triangle with vertices at positions a, b, c
- edge lengths are $\|a-b\|$, $\|b-c\|$, $\|a-c\|$
- by triangle inequality

$$\|a-c\| = \|(a-b)+(b-c)\| \leq \|a-b\| + \|b-c\|$$

i.e., third edge length is no longer than sum of other two

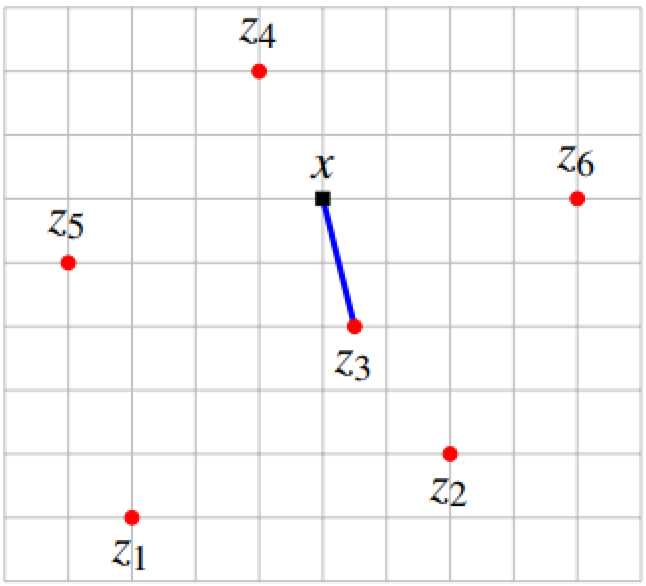


Norm and Distance

Feature distance and nearest neighbors

- if x and y are feature vectors for two entities, $\| x - y \|$ is the feature distance
- if $z_1, \ldots, z_m$ is a list of vectors, z_j is the *nearest neighbor* of x if

$$\|x - z_j\| \leq \|x - z_i\|, \quad i = 1, \ldots, m$$



Norm and Distance

Standard deviation

- for n-vector x, $\mathbf{avg}(x) = \mathbf{1}^T x/n$
- *de-meaned vector* is $\tilde{x} = x - \mathbf{avg}(x)1$ (so $\mathbf{avg}(\tilde{x}) = 0$)
- *standard deviation* of x is

$$\mathbf{std}(x) = \mathbf{rms}(\tilde{x}) = \frac{\|x - (\mathbf{1}^T x/n)\mathbf{1}\|}{\sqrt{n}}$$

- $\mathbf{std}(x)$ gives 'typical' amount x_i vary from $\mathbf{avg}(x)$
- $\mathbf{std}(x) = 0$ only if $x = \alpha\mathbf{1}$ for some α
- greek letters μ and σ commonly used for mean and standard deviation
- a basic formula:

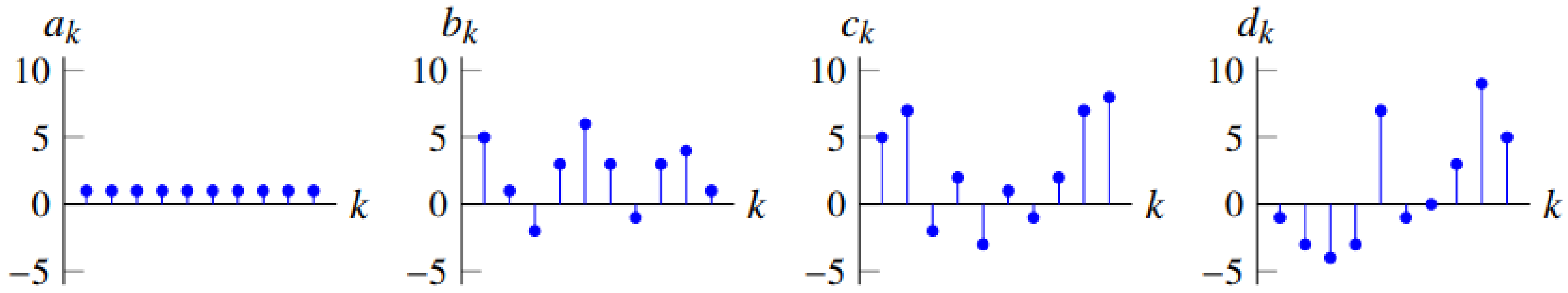
$$\mathbf{rms}(x)^2 = \mathbf{avg}(x)^2 + \mathbf{std}(x)^2$$

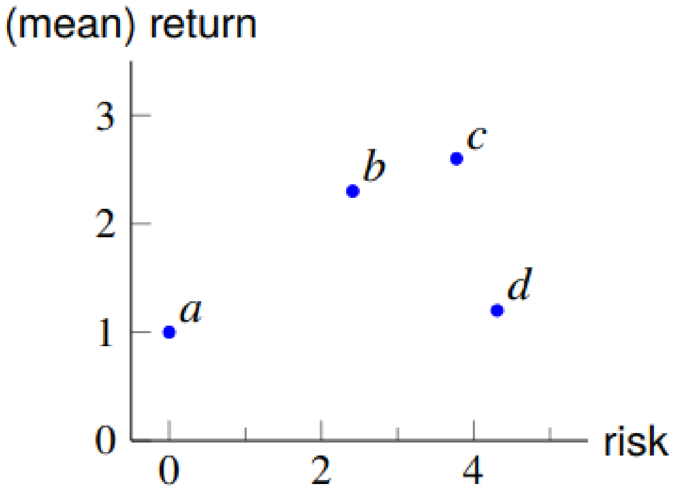
*rms(x^2)로 정정.

Mean return and risk

- x is time series of returns (say, in %) on some investment or asset over some period
- $\mathbf{avg}(x)$ is the mean return over the period, usually just called *return*
- $\mathbf{std}(x)$ measures how variable the return is over the period, and is called the *risk*
- multiple investments (with different return time series) are often compared in terms of return and risk
- often plotted on a *risk-return plot*

Risk-return example





Cauchy–Schwarz inequality

- for two n-vectors a and b, $|a^Tb| \leq \|a\| \, \|b\|$
- written out,

$$|a_1b_1 + \cdots + a_nb_n| \leq \left(a_1^2 + \cdots + a_n^2\right)^{1/2} \left(b_1^2 + \cdots + b_n^2\right)^{1/2}$$

- now we can show triangle inequality:

$$\begin{aligned} \|a+b\|^2 &= \|a\|^2 + 2a^Tb + \|b\|^2 \\ &\leq \|a\|^2 + 2\|a\|\|b\| + \|b\|^2 \\ &= (\|a\| + \|b\|)^2 \end{aligned}$$

Derivation of Cauchy–Schwarz inequality

- it's clearly true if either a or b is 0
- so assume $\alpha = \|a\|$ and $\beta = \|b\|$ are nonzero
- we have

$$\begin{aligned} 0 &\leq \|\beta a - \alpha b\|^2 \\ &= \|\beta a\|^2 - 2(\beta a)^T(\alpha b) + \|\alpha b\|^2 \\ &= \beta^2\|a\|^2 - 2\beta\alpha(a^T b) + \alpha^2\|b\|^2 \\ &= 2\|a\|^2\|b\|^2 - 2\|a\|\,\|b\|(a^T b) \end{aligned}$$

- divide by $2\|a\|\,\|b\|$ to get $a^Tb \leq \|a\|\,\|b\|$
- apply to $-a$, b to get other half of Cauchy–Schwarz inequality

Norm and Distance

Angle

- *angle* between two nonzero vectors a, b defined as

$$\angle(a,b) = \arccos\left(\frac{a^T b}{\|a\|\,\|b\|}\right)$$

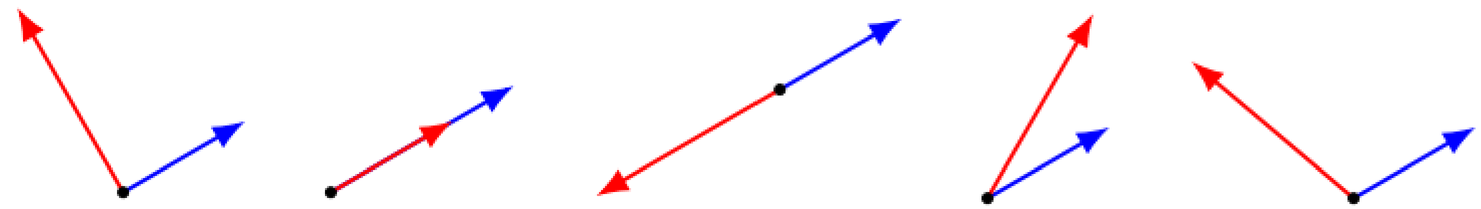
- $\angle(a, b)$ is the number in $[0, \pi]$ that satisfies

$$a^T b = \|a\|\,\|b\| \cos\left(\angle(a,b)\right)$$

Classification of angles

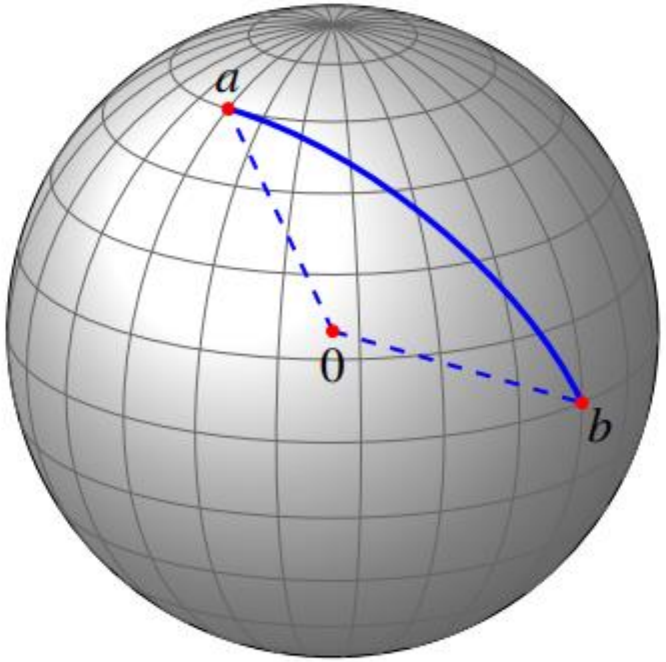
$\theta = \angle(a, b)$

- $\theta = \pi/2 = 90^\circ$: a and b are *orthogonal*, written $a \perp b$ ($a^Tb = 0$)
- $\theta = 0$: a and b are *aligned* ($a^Tb = \|a\|\,\|b\|$)
- $\theta = \pi = 180^\circ$: a and b are *anti-aligned* ($a^Tb = -\|a\|\,\|b\|$)
- $\theta \leq \pi/2 = 90^\circ$: a and b make an *acute angle* ($a^Tb \geq 0$)
- $\theta \geq \pi/2 = 90^\circ$: a and b make an *obtuse angle* ($a^Tb \leq 0$)



Spherical distance

- if a, b are on sphere of radius R, distance *along the sphere* is $R\angle(a, b)$



Correlation coefficient

- vectors a and b, and de-meaned vectors

$$\tilde{a} = a - \mathbf{avg}(a)\mathbf{1}, \qquad \tilde{b} = b - \mathbf{avg}(b)\mathbf{1}$$

- *correlation coefficient* (between a and b, with $\tilde{a} \neq 0$, $\tilde{b} \neq 0$)

$$\rho = \frac{\tilde{a}^T \tilde{b}}{\|\tilde{a}\| \, \|\tilde{b}\|}$$

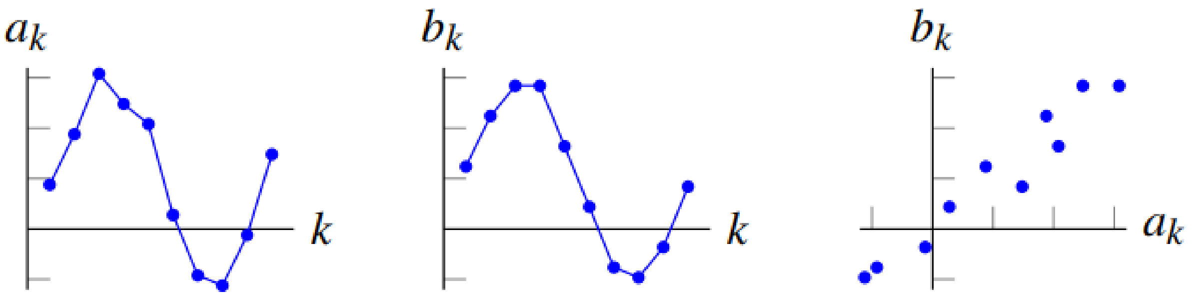
- $\rho = \cos \angle(\tilde{a}, \tilde{b})$

  - $\rho = 0$: a and b are *uncorrelated*

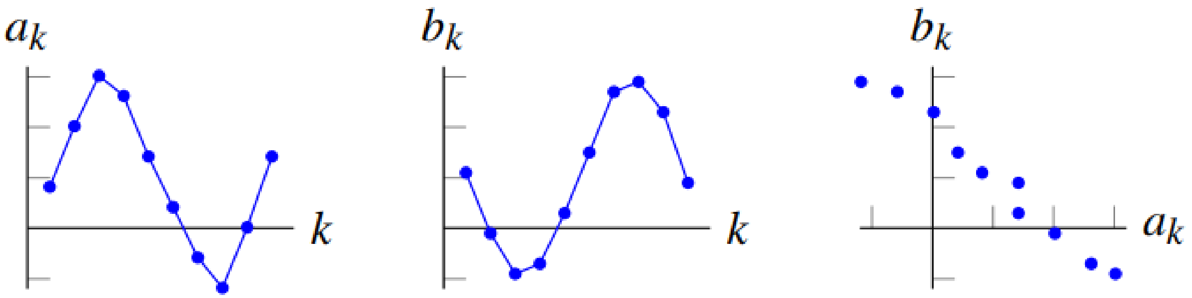
  - $\rho > 0.8$ (or so): a and b are *highly correlated*

  - $\rho < -0.8$ (or so): a and b are *highly anti-correlated*

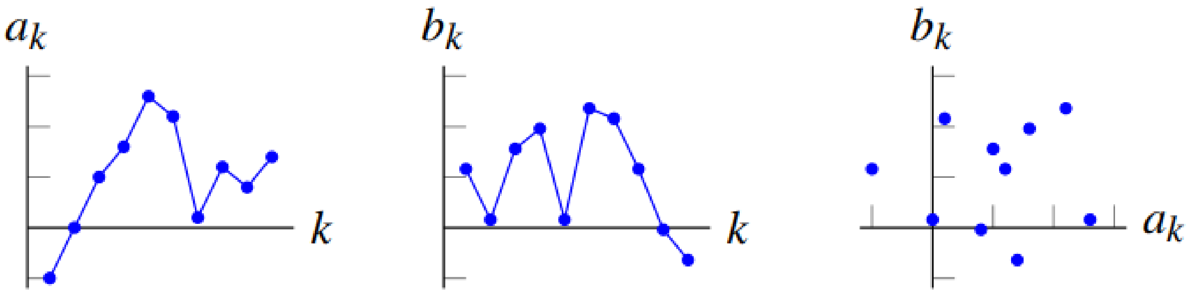
Norm and Distance
Examples



$\rho = 97\%$



$\rho = -99\%$



$\rho = 0.4\%$