

Outline

Norm

Distance

Standard deviation

Angle

Cauchy–Schwarz inequality

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

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

- ▶ now we can show triangle inequality:

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

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 $a^T b \leq \|a\|\|b\|$
- ▶ apply to $-a, b$ to get other half of Cauchy–Schwarz inequality

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

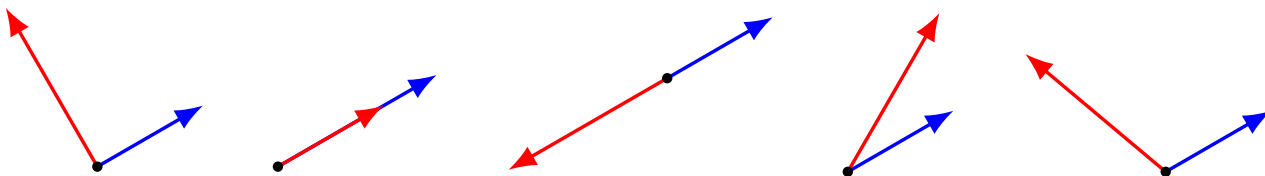
$$a^T b = \|a\| \|b\| \cos(\angle(a, b))$$

- ▶ coincides with ordinary angle between vectors in 2-D and 3-D

Classification of angles

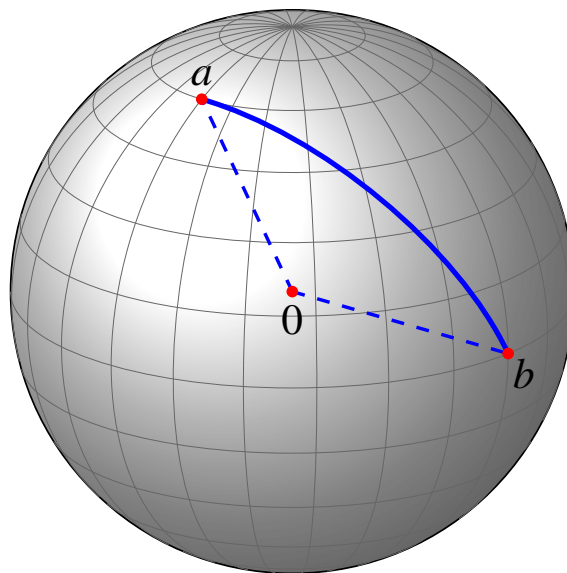
$$\theta = \angle(a, b)$$

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



Spherical distance

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



Document dissimilarity by angles

- ▶ measure dissimilarity by angle of word count histogram vectors
- ▶ pairwise angles (in degrees) for 5 Wikipedia pages shown below

	Veterans Day	Memorial Day	Academy Awards	Golden Globe Awards	Super Bowl
Veterans Day	0	60.6	85.7	87.0	87.7
Memorial Day	60.6	0	85.6	87.5	87.5
Academy A.	85.7	85.6	0	58.7	85.7
Golden Globe A.	87.0	87.5	58.7	0	86.0
Super Bowl	87.7	87.5	86.1	86.0	0

Correlation coefficient

- ▶ vectors a and b , and de-meaned vectors

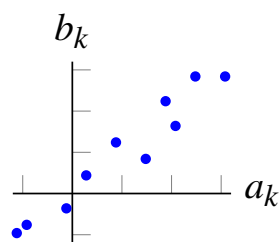
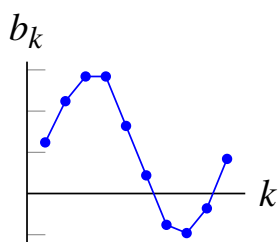
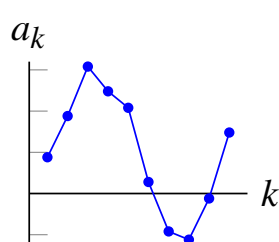
$$\tilde{a} = a - \mathbf{avg}(a)\mathbf{1}, \quad \tilde{b} = b - \mathbf{avg}(b)\mathbf{1}$$

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

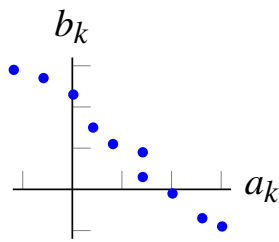
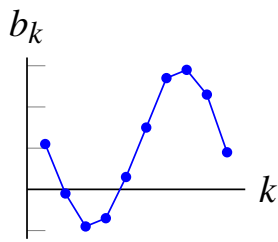
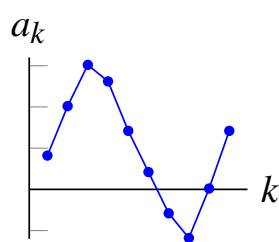
$$\rho = \frac{\tilde{a}^T \tilde{b}}{\|\tilde{a}\| \|\tilde{b}\|}$$

- ▶ $\rho = \cos \angle(\tilde{a}, \tilde{b})$
 - $\rho = 0$: 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*
- ▶ very roughly: highly correlated means a_i and b_i are typically both above (below) their means together

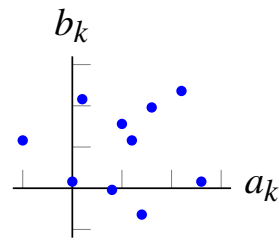
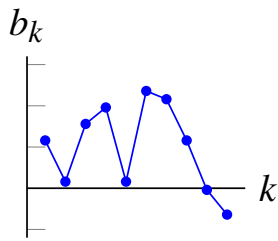
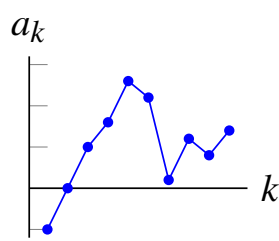
Examples



$$\rho = 97\%$$



$$\rho = -99\%$$



$$\rho = 0.4\%$$

Examples

- ▶ highly correlated vectors:
 - rainfall time series at nearby locations
 - daily returns of similar companies in same industry
 - word count vectors of closely related documents (e.g., same author, topic, ...)
 - sales of shoes and socks (at different locations or periods)
- ▶ approximately uncorrelated vectors
 - unrelated vectors
 - audio signals (even different tracks in multi-track recording)
- ▶ (somewhat) negatively correlated vectors
 - daily temperatures in Palo Alto and Melbourne

5. Linear independence

Outline

Linear independence

Basis

Orthonormal vectors

Gram–Schmidt algorithm

Linear dependence

- ▶ set of n -vectors $\{a_1, \dots, a_k\}$ (with $k \geq 1$) is *linearly dependent* if

$$\beta_1 a_1 + \dots + \beta_k a_k = 0$$

holds for some $\beta_1, \dots, \beta_k$, that are not all zero

- ▶ equivalent to: at least one a_i is a linear combination of the others
- ▶ we say ' $a_1, \dots, a_k$ are linearly dependent'
- ▶ $\{a_1\}$ is linearly dependent only if $a_1 = 0$
- ▶ $\{a_1, a_2\}$ is linearly dependent only if one a_i is a multiple of the other
- ▶ for more than two vectors, there is no simple to state condition

Example

- ▶ the vectors

$$a_1 = \begin{bmatrix} 0.2 \\ -7 \\ 8.6 \end{bmatrix}, \quad a_2 = \begin{bmatrix} -0.1 \\ 2 \\ -1 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 0 \\ -1 \\ 2.2 \end{bmatrix}$$

are linearly dependent, since $a_1 + 2a_2 - 3a_3 = 0$

- ▶ can express any of them as linear combination of the other two, e.g.,

$$a_2 = (-1/2)a_1 + (3/2)a_3$$