2

L6: Distances

Jeff M. Phillips

January 28, 2019

distance

$$d : \underbrace{X}_{\text{domain}} \times X \to [0, \infty)$$

$$d(a,b) = 72.3 \qquad a, b \in X$$

metric

(M1) $d(a,b) \geq 0$

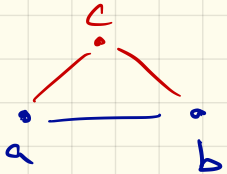
(M2) $d(a,b) = 0$ iff $a = b$ (identity)

(M3) $d(a,b) = d(b,a)$ (symmetry)

(M4) $d(a,b) \leq d(a,c) + d(c,b)$ (triangle inequality

→ pseudometric

→ quasimetric



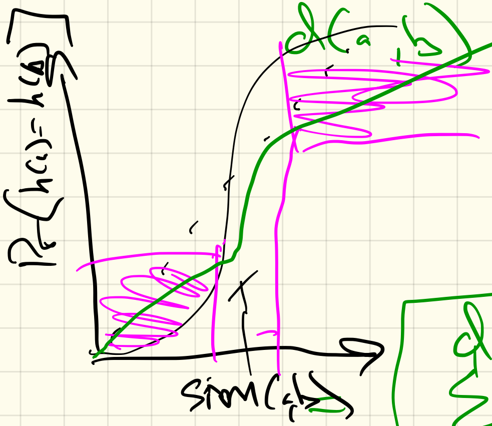
Euclidean Distance (L_2 distance)

Domain $\mathcal{X}$: $\mathbb{R}^d$

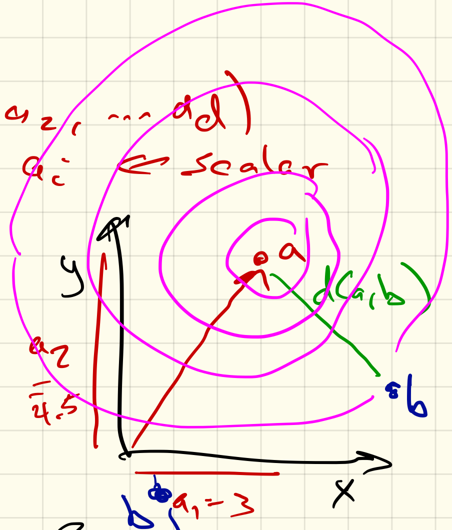
$a \in \mathbb{R}^d$

$a = (a_1, a_2, \ldots, a_d)$

$a_i \Leftarrow$ scalar

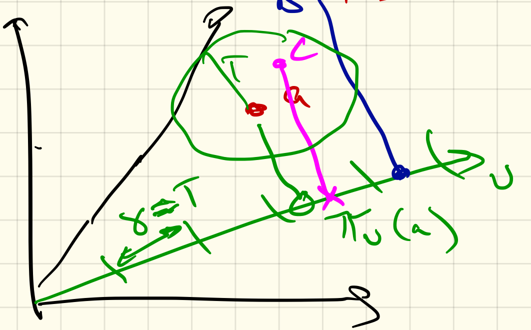


$$d(a,b) = L_2(a,b) = \|a-b\|_2 = \|a-b\| = \sqrt{\sum_{i=1}^{d} (a_i-b_i)^2}$$



LSHable

$u \sim unif(S^{d-1})$



L_p Distances

$d_p: \mathbb{R}^d \times \mathbb{R}^d \to [0,\infty)$

$$d_p(a,b) = \|a-b\|_p = \left(\sum_{i=1}^{d} (|a_i - b_i|)^p\right)^{1/p}$$

valid

$p \in (0,\infty)$

metric

$p \in [1,\infty)$

L-stable

$p \in [1,2] \approx p \in (0,2]$

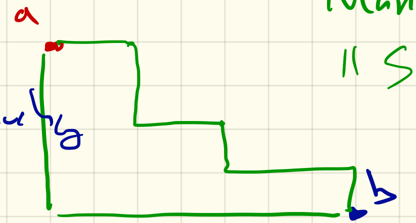
L_1

$$d_1(a,b) = \|a-b\|_1 = \sum_{i=1}^{d} |a_i - b_i|$$

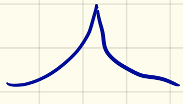
~~Manhattan~~

"SLC"

Eucl	SLC
• $g \sim$ Gauss	$c \sim$ Cauchy
• $u = \frac{g}{\|u\|}$	"
• $x = \langle u, a \rangle$	"
• bin	



Cauchy $\sim \frac{1}{\pi} \frac{1}{1+x^2}$

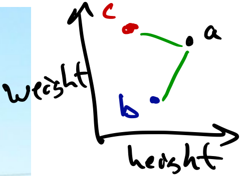


Lp Distances and Units

For $a = (a_1, a_2, \ldots, a_d)$ and $b = (b_1, b_2, \ldots, b_d) \in \mathbb{R}^d$,

$$L_p: \quad \mathtt{d}_p(a, p) = \|a - b\|_p = \left(\sum_{i=1}^{d} (|a_i - b_i|)^p\right)^{1/p}.$$

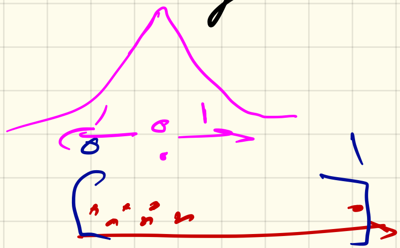
$d(a,b) = d(a,c)$



bad

(a_1, a_2) different units

- don't take distance.
- Distance Metric Learning
- Normalize Data



For each coordinate

$a_1, b_1, c_1, d_1 \ldots \to [0, 1]$, mean 0, stdev 1

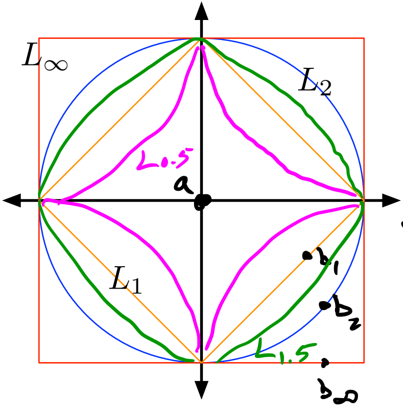
$a_2, b_2, c_2, d_2 \ldots \to [0, 1]$, mean 0, stdev 1

Lp Distances and Unit Balls

For $a = (a_1, a_2, \ldots, a_d)$ and $b = (b_1, b_2, \ldots, b_d) \in \mathbb{R}^d$,

$$L_p : \quad \mathrm{d}_p(a, p) = \|a - b\|_p = \left(\sum_{i=1}^{d} (|a_i - b_i|)^p \right)^{1/p} .$$

Let $a = (0, 0, \ldots, 0)$ and $\|a - b\|_p = 1$.



$$L_0 = \|a-b\|_0 = d - \sum_{i=1}^{d} \mathbb{1}(a_i = b_i)$$

$$L_\infty = \|a-b\|_\infty = \max_i |a_i - b_i|$$

$L_p(x)$ ball $\subset$ $L_{p'}(x)$ ball

$p' \geq p$

Mahalanobis Distance

$M \in \mathbb{R}^{d \times d}$

$$d_M(a, b) = \sqrt{(a-b)^T M (a-b)}$$

if $M = I = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$

$\hookrightarrow d_M = L_2$

M pd $\rightarrow$ metric

Jaccard Distance

$$d_J(A,B) = 1 - JS(A,B)$$

metric

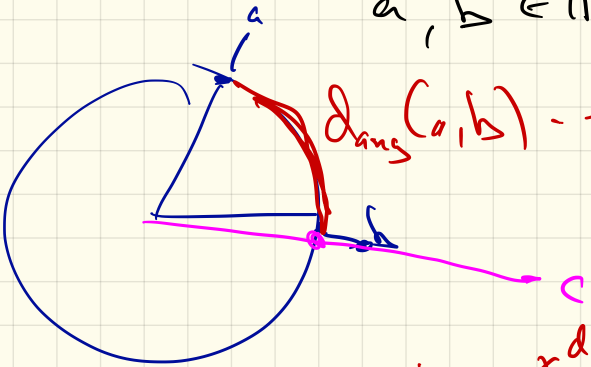
$$1 - \frac{|A \cap B|}{|A \cup B|}$$

$$= \frac{|A \triangle B|}{|A \cup B|}$$

Cosine Distance $\quad 1 - \boxed{\frac{\langle a,b\rangle}{\|a\|\cdot\|b\|}}$

Angular Distance $\quad 1 - \boxed{\arccos\left(\frac{\langle a,b\rangle}{\|a\|\cdot\|b\|}\right)}$

$a, b \in \mathbb{R}^d$



$d_{ang}(a,b) = \pi$

$a, b \in S^{d-1}$

$\hookrightarrow d_{ang}$ metric

$$\frac{\langle a,b\rangle}{\|a\|\cdot\|b\|} = \left\langle \frac{a}{\|a\|}, \frac{b}{\|b\|} \right\rangle$$

$$\langle a,b\rangle = \sum_{i=1}^{d} a_i \cdot b_i$$