

Distance Metrics.*

Greg Kochanski

<http://kochanski.org/gpk>

February 14, 2004

To talk about “close” and “far” in a consistent manner, we may need to have the concept of a distance metric. You don’t need a distance metric if you are just measuring how far you are from one particular point, but you need one any time you want to be able to measure arbitrary points on a line, on a plane, or in space.

A distance metric, $d(A, B)$ is a function or algorithm for calculating a distance between two things, A and B . It has three properties:

1. It is always positive or zero.
2. The distance from a thing¹ to itself is zero.
3. It obeys the triangle inequality: For any three points, A , B , and C , $d(A, B) + d(B, C) \geq d(A, C)$ for any possible choice of B . In other words, the straight line between A and C , which has a length $d(A, C)$, is shorter² than any other path between A and C , such as a path that goes by way of B .

Anything that obeys these three properties is a distance metric. Common examples are Euclidean distance in two dimensions:

$$d_{Euclid}(A, B) = ((A_x - B_x)^2 + (A_y - B_y)^2)^{1/2}, \quad (1)$$

and city block metric [Krowne, 2003]:

$$d_{Euclid}(A, B) = (|A_x - B_x| + |A_y - B_y|). \quad (2)$$

(The Euclidean and city-block metrics generalize to any number of dimensions in a straightforward manner.)

References

[Krowne, 2003] Krowne, A. (2003). *PlanetMath: city-block metric*. PlanetMath.org, <http://planetmath.org/encyclopedia/CityBlockMetric.html>. Canonical name CityBlockMetric,.

*This work is licensed under the Creative Commons Attribution License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA. This work is available under <http://kochanski.org/gpk/teaching/0401Oxford> .

¹ A geometrical point, or a document...

² Strictly speaking this should read “. . . at least as short as. . .,” because if B is on the line between A and C , then $d(A, B) + d(B, C) = d(A, C)$. Also, it’s possible to have a distance measure where certain directions just don’t count: an extreme example is if $d(X, Y) = 0$ for all X and Y . This always-zero (and rather useless) distance metric obeys the triangle inequality because $d(A, B) + d(B, C) = d(A, C)$ for all B .