Measures and metrics

Scientific Perspectives on GMT 2019/2020

Marc van Kreveld

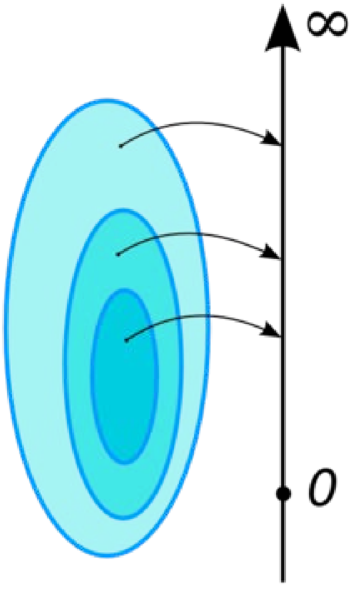
A number of related concepts

- Measure
  - Math
  - Other
- Metric
  - Math
  - Other
- Indicator: same as measure/metric, other
- Measurement

Desperate times call for desperate measures
- English proverb (Hippocrates?)

Measures in mathematics

- Functions from “subsets” to the reals
- A ***measure*** obeys the properties:
    1. Non-negativeness: for any subset X, f(X) ≥ 0
    2. Null empty set: For the empty set, f(∅) = 0
    3. Additivity: for two disjoint subsets X and Y, f(X ∪ Y) = f(X) + f(Y)



Measures in mathematics

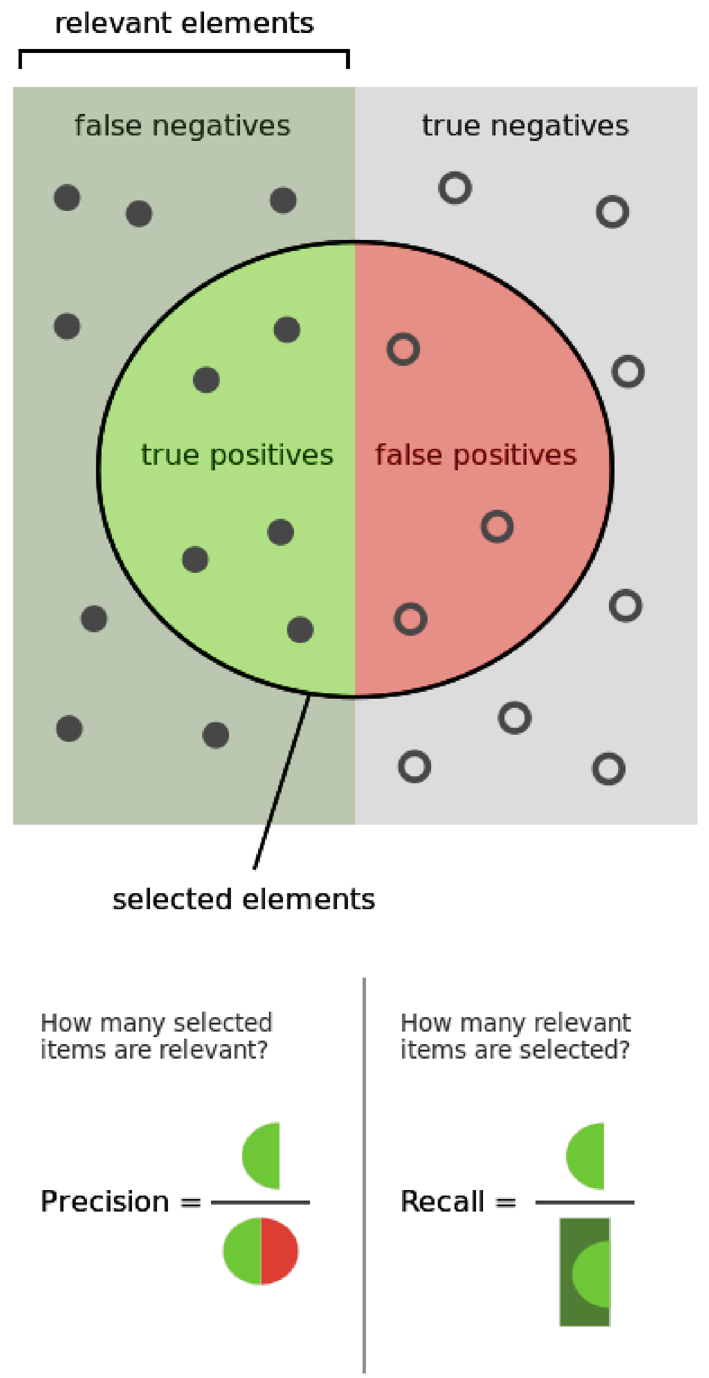
- Example 1: Space is the real line, subsets are disjoint unions of intervals, measure is (total) length
- Example 2: Space is all integers, subsets are finite subsets of integers, measure is number of integers in a subset
- Example 3: Space is outcomes of an experiment (die rolling), measure is probability of the outcome(s)

Measures in the rest of science

- Functions from "something" to the nonnegative reals
- Capture an intuitive aspect: size, quality, difficulty, distance, similarity, usefulness, robustness, … into something well-defined

- Precision and recall in information retrieval
- Support and confidence in association rule mining

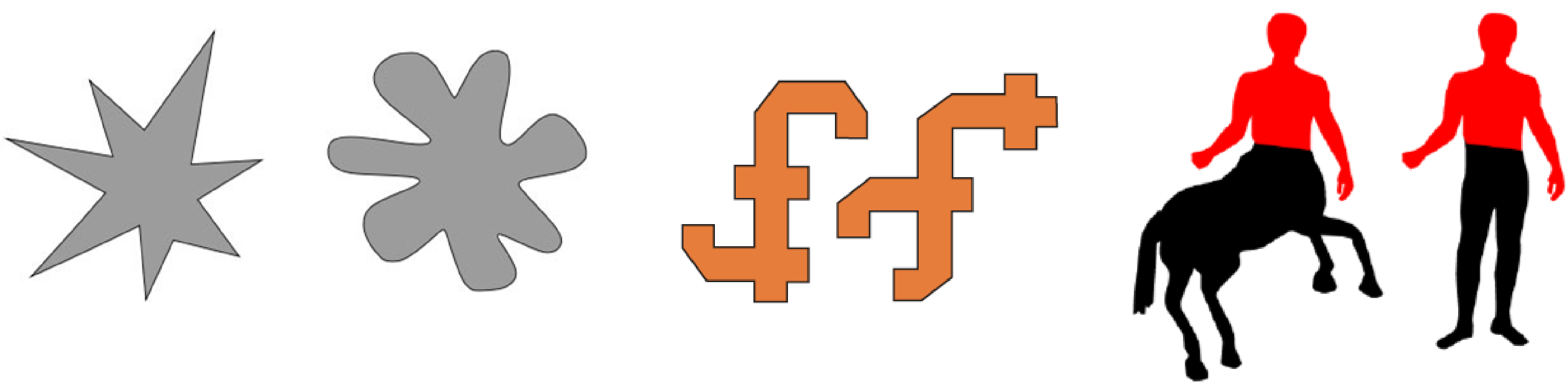
- In the world at large: body mass index, ecological footprint, …


relevant elements
false negatives
true negatives
true positives
false positives
selected elements
How many selected items are relevant?
How many relevant items are selected?
Precision =
Recall =

Measures in the rest of science

- Albeit well-defined, the real connection of the (abstract) function to the intuitive concept is not guaranteed ➔ Needs to be justified or tested

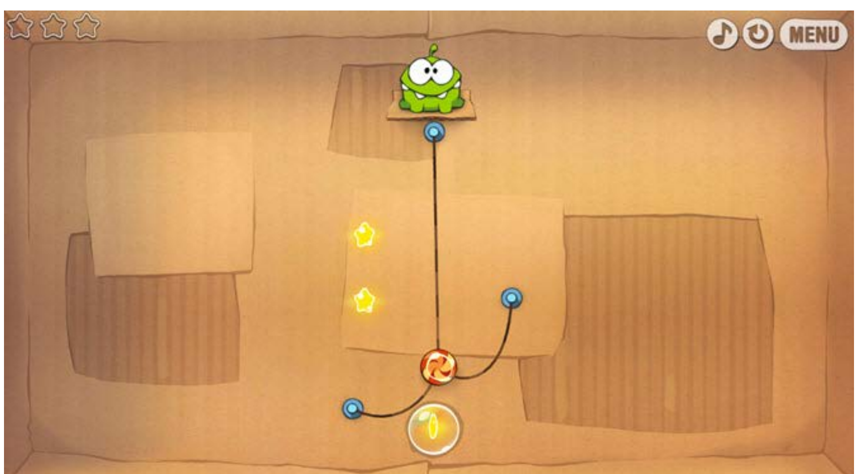
- Example 1: similarity measure for two shapes



Measures in the rest of science

- Albeit well-defined, the real connection of the (abstract) function to the intuitive concept is not guaranteed ➔ Needs to be justified or tested

- Example 2: difficulty rating of a level in a puzzle game

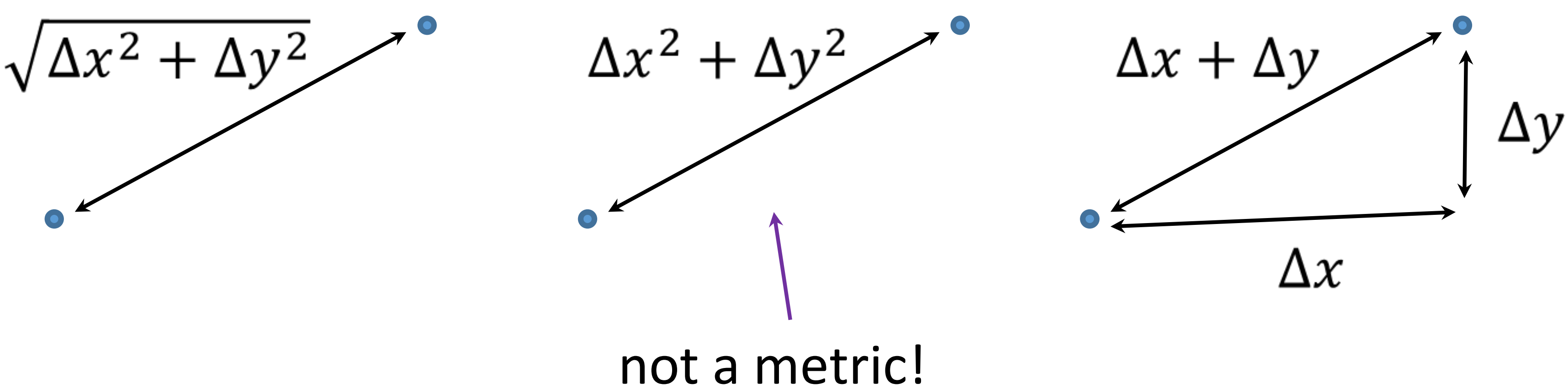


Distance functions, or metrics

- Distance: how far things are apart
- A metric or distance function takes ***two*** arguments and returns a nonnegative real
- Distances on a set X; for any x, y, z in X, a ***metric*** is a function $d(x,y) \rightarrow R$ (the reals) where:
  1. $d(x,y) \geq 0$ non-negative
  2. $d(x,y) = 0$ if and only if $x = y$ coincidence
  3. $d(x,y) = d(y,x)$ symmetry
  4. $d(x,z) \leq d(x,y) + d(y,z)$ triangle inequality

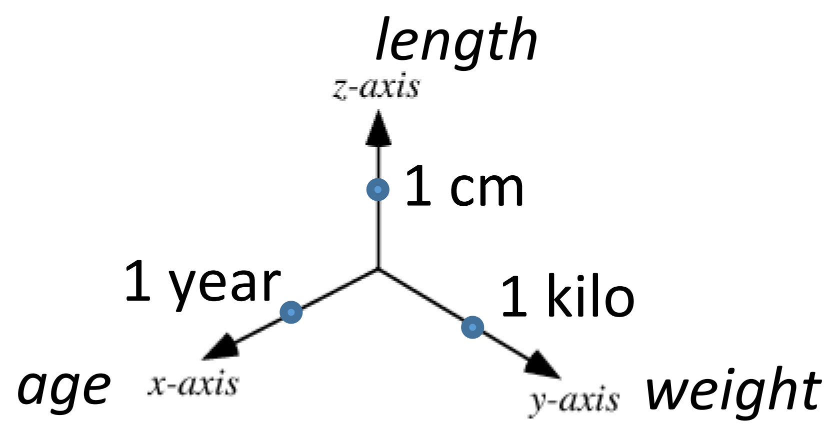
Examples of metrics on points

- Euclidean distance on the line, in the plane or in a higher-dimensional space, L_2 distance
  Note: Squared Euclidean distance is not a metric
- City block, Manhattan, or L_1 distance (are the same)
- L_∞ distance (max of differences in the coordinates)



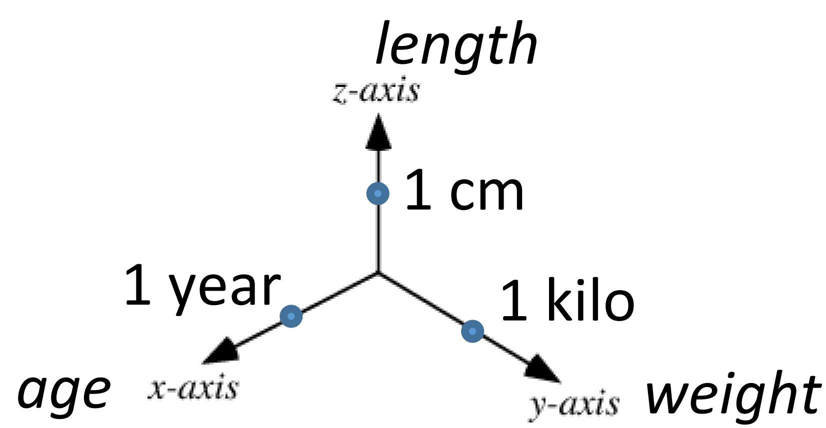
Distances between points in an attribute space?

- Suppose points in 3D represent people with their age, weight, and length
- Any metric that uses these components is influenced by normalization or scaling of an axis
- Any metric makes a choice on how many years correspond to one kilo or one centimeter, and therefore weighs the relevance of the components



Distances between points in an attribute space?

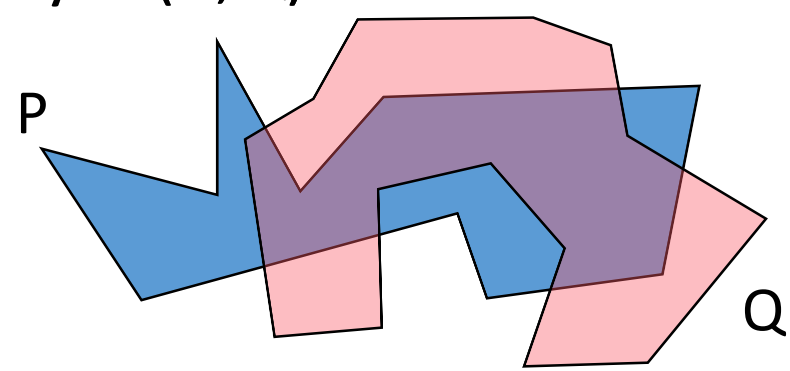
- For a specific point set in an attribute space, one can normalize its axes by making the unit the standard deviation of its values
- … but then, two different point sets in spaces with the same attributes use different distances



Example of metric on polygons

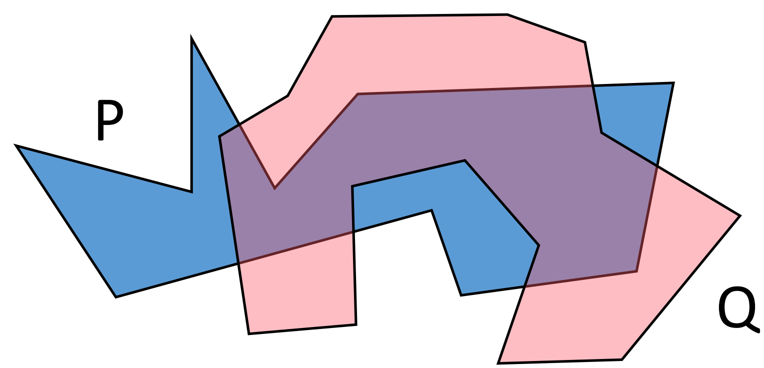
- Area of symmetric difference Asym, is it a metric?
- Three properties (nonnegative, coincidence, symmetry) clear, to be verified: triangle inequality. It reads:

  Given three polygons P, Q, R, we always have
  Asym(P,Q) ≤ Asym(P,R) + Asym (R,Q)



Example of metric on polygons

- Given three polygons P, Q, R, we always have Asym(P,Q) ≤ Asym(P,R) + Asym (R,Q)
- For any polygon R, we consider the parts counted in Asym(P,Q):
  - In P, not in Q, and in R: also counted in Asym(R,Q)
  - In P, not in Q, and not in R: also counted in Asym(P,R)
  - Not in P, in Q, and in R: also counted in Asym(P,R)
  - Not in P, in Q, and not in R: also counted in Asym(R,Q)
  - In both or neither P, Q: not counted in Asym(P,Q)



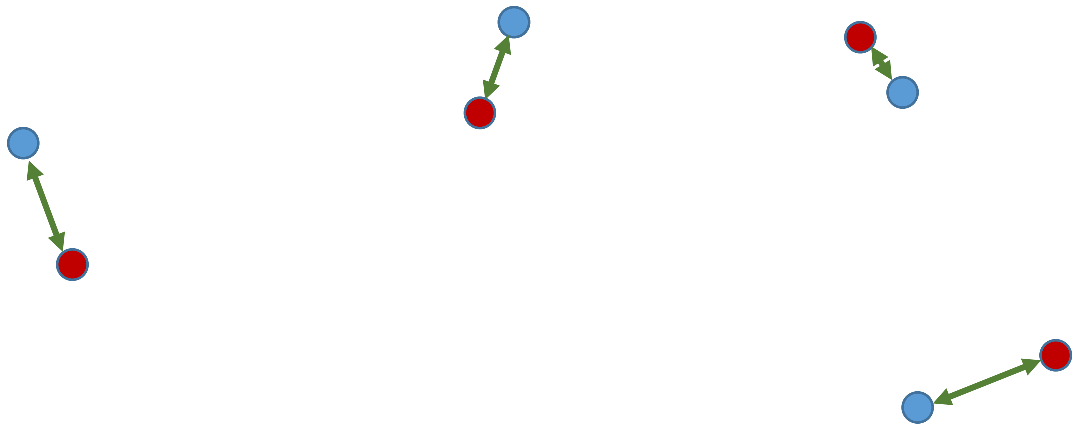
Interesting aspects for measures in geometric situations

“Measures” in the loose sense:

- Size (descriptive measure for many things)
- Elongatedness (descriptive measure for a polygon)
- Spread (descriptive measure for a point set)
- Goodness of fit (for e.g. a shape and a point set)
- Similarity / distance (for two things of the same type)
- …

Aggregation in measures

- When defining the distance between two *point sets*, we may want to combine several point-to-point distances into one distance measure
- This can be called **aggregation** of distances



Aggregation in measures

- **Bottleneck**: aggregation is done by taking a minimum or maximum over values
  Examples: Hausdorff, Fréchet

- **Sum**: aggregation is done by taking the sum over values
  Examples: DTW, EMD, area of symmetric difference

- **Sum-of-squares**: aggregation is done by taking the sum-of-squares over values
  Example: Error of regression line model

Aggregation in measures

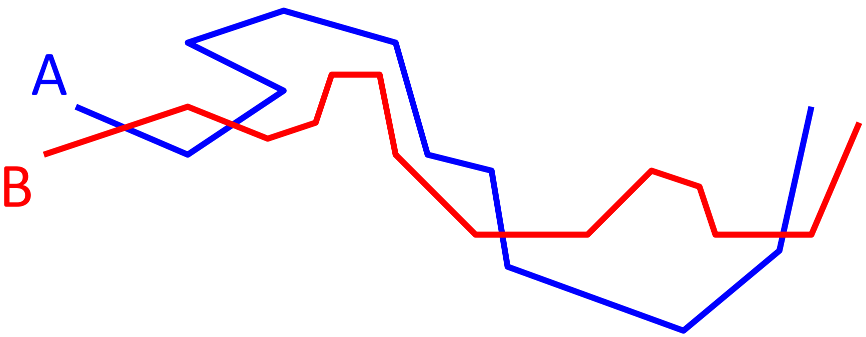
- **Bottleneck**: very sensitive to outliers
- **Sum**: mildly sensitive to outliers
- **Sum-of-squares**: moderately sensitive to outliers

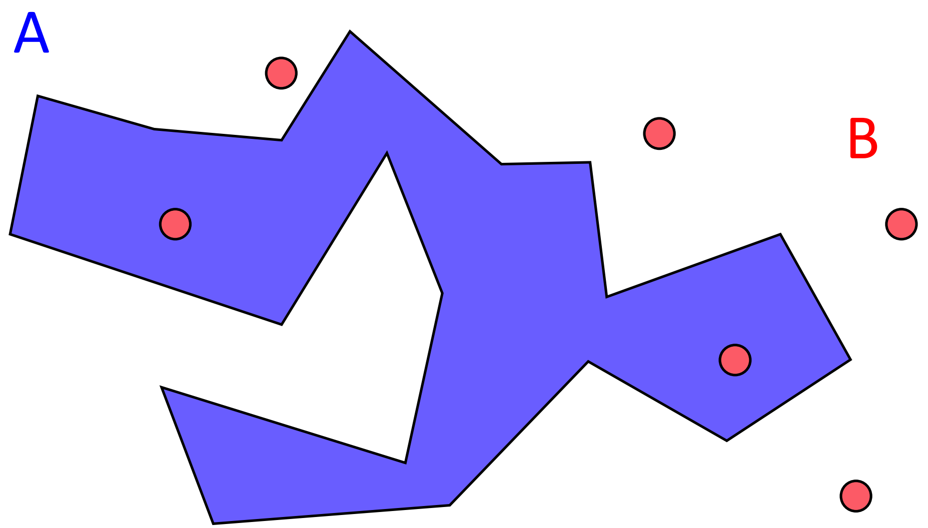
Well-known geometric metrics/measures

- Hausdorff distance (any set; asymmetric, symmetric)
- Area of symmetric difference (for polygons)
- Fréchet distance (for curves)
- Dynamic Time Warping (for time series, or for curves)
- Earth Mover's Distance

Hausdorff distance

- Defined for any two subsets of the plane (two point sets, two curves, two polygons, a curve and a polygon, …)





Hausdorff distance

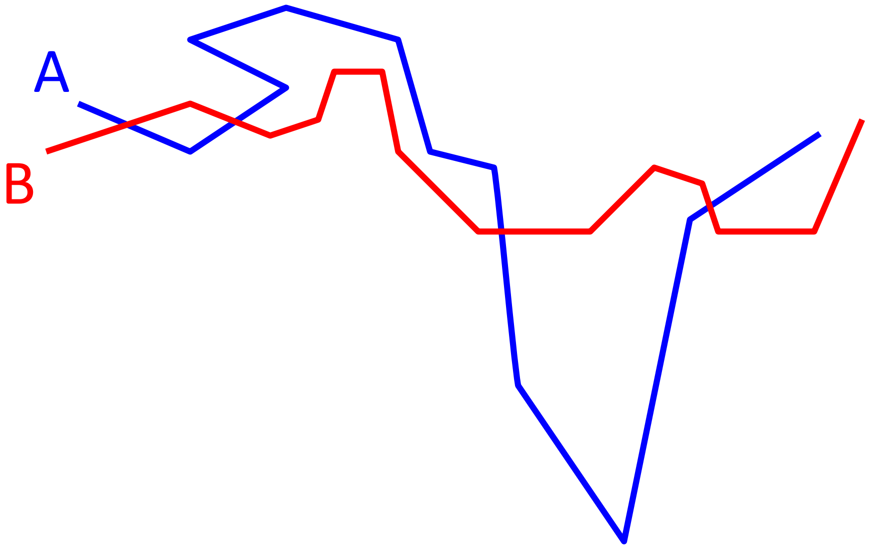
- Defined for any two subsets of the plane (two point sets, two curves, two polygons, a curve and a polygon, …)
- Bottleneck metric
- Asymmetric version: A → B (or B → A); not a metric
- Symmetric version: Max of the asymmetric versions:

$$\text{Max}\left(\max_{a \in A} \min_{b \in B} \text{dist}(a,b)\ ,\ \max_{b \in B} \min_{a \in A} \text{dist}(b,a) \right)$$

A → B B → A

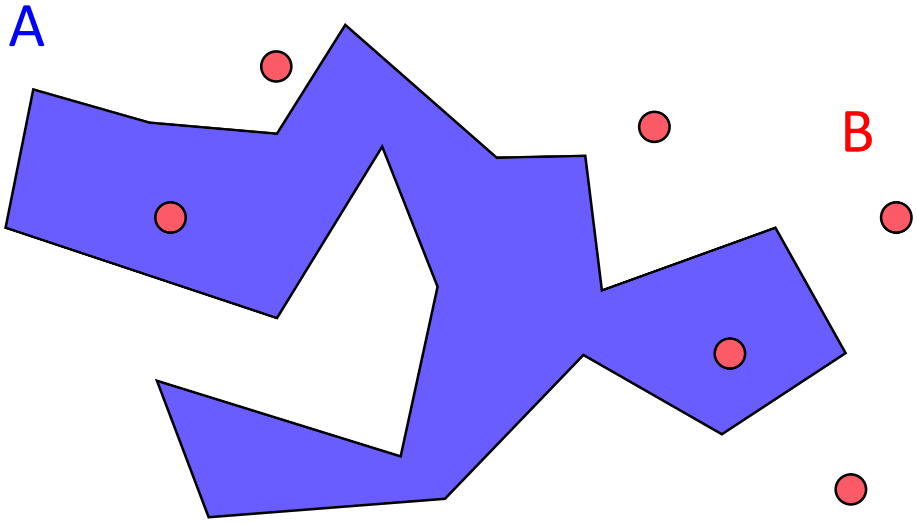
Hausdorff distance

- Which is larger: the Hausdorff distance A → B or B → A ?



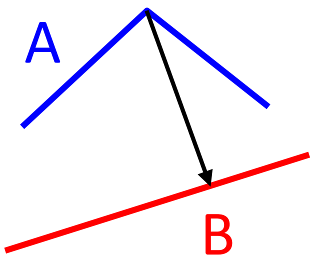
Hausdorff distance

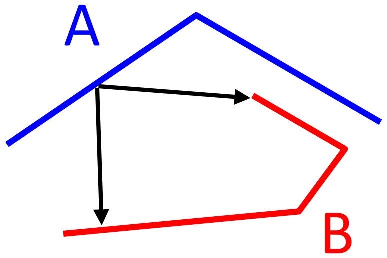
- Which is larger: the Hausdorff distance A → B or B → A ?



Properties Hausdorff distance

- Where can largest distance from A to B occur?





Vertex of A

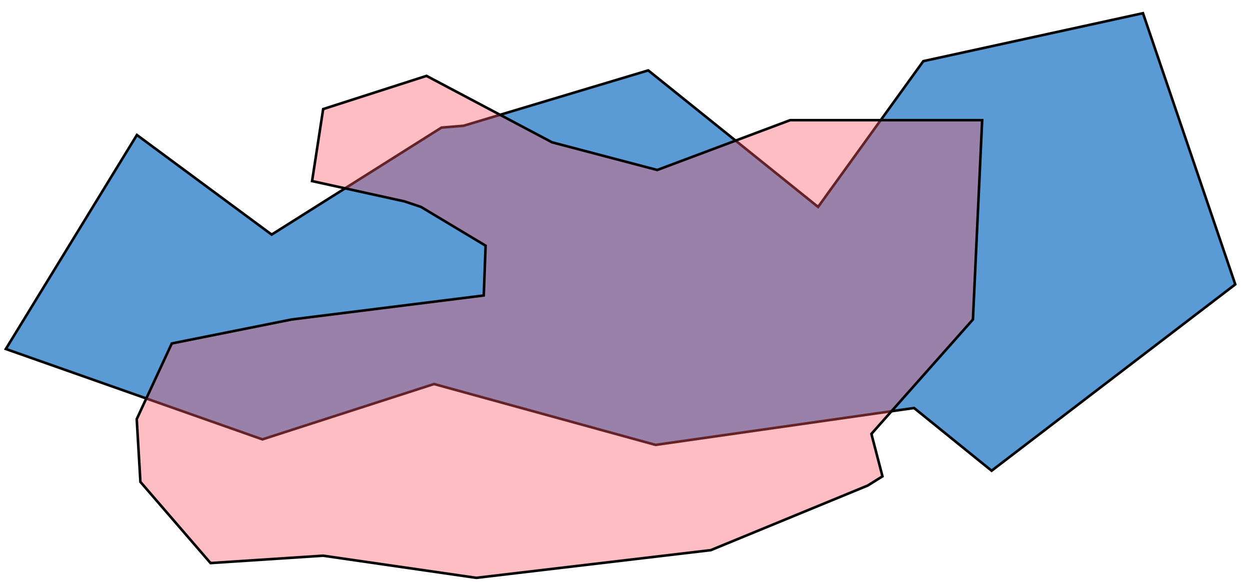
Point internal to edge of A

⇓

In this case, the minimum distance must be attained from that point on A to *two places* on B

Computation area of symmetric difference

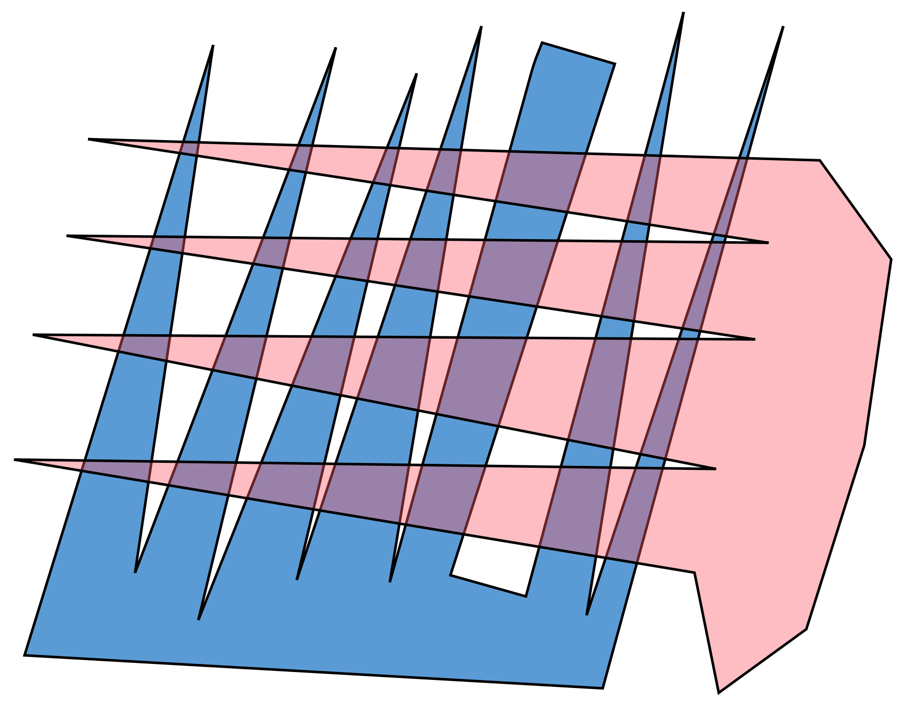
- Perform map overlay (Boolean operation) on the two polygons
- Compute area of symmetric difference of the polygons and add up



Computation area of symmetric difference

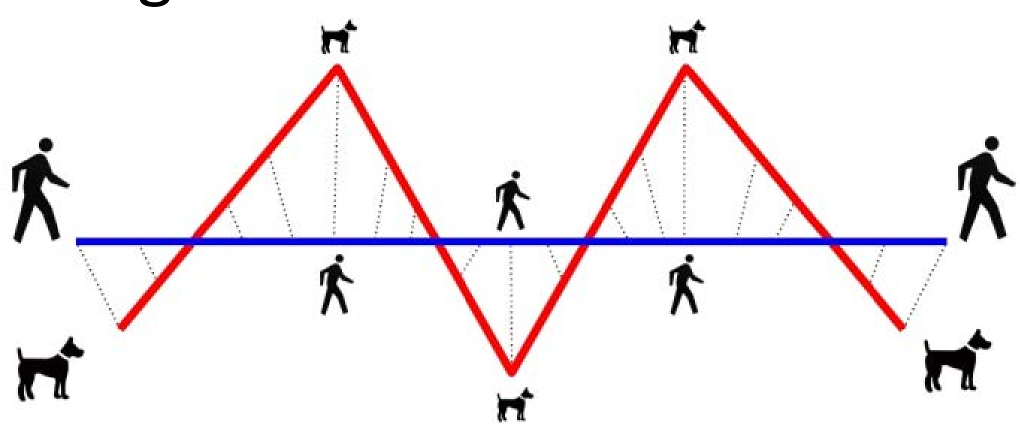
- Perform map overlay (Boolean operation) on the two polygons
- Compute area of symmetric difference of the polygons and add up

- Worst case:
  $O(nm \log (nm))$ time
- Typical case:
  $O((n+m) \log (n+m))$ time

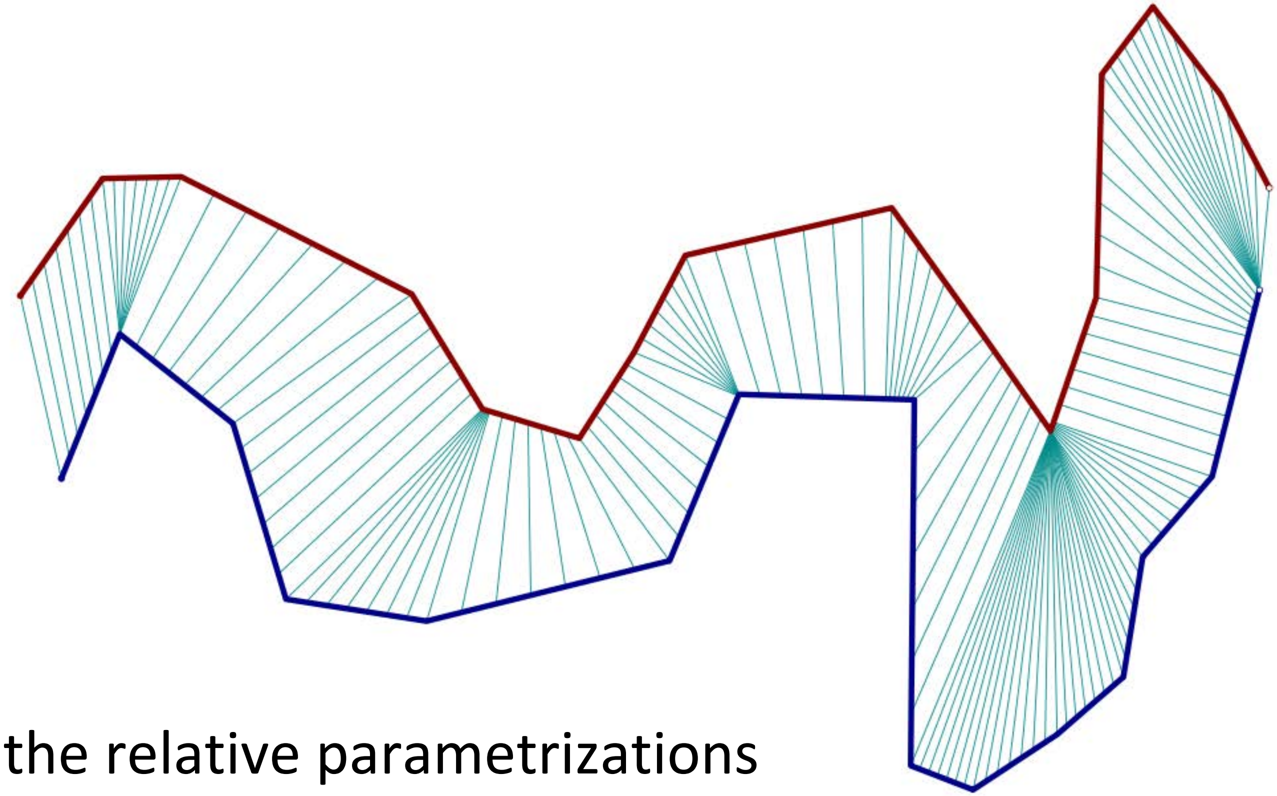


Fréchet distance

- For two oriented curves in 2D or 3D
- Extensions to surfaces exist (but are not treated)
- Intuitively: a man walks on one curve with possibly varying speed, but only forward, and a dog does the same on the other curve. The Fréchet distance is the minimum leash length needed to allow this (man-dog distance, leash distance)



Fréchet distance



Needed: the relative parametrizations over the two curves

Fréchet distance

- Definition:
  let α : [0,1] be a parametrization of curve A and
  let β : [0,1] be a parametrization of curve B where
  - $\alpha(0)$ = the start of A
  - $\alpha(1)$ = the end of A
  - $\beta(0)$ = the start of B
  - $\beta(1)$ = the end of B
- α is a continuous bijection between [0,1] and A, and
  β is a continuous bijection between [0,1] and B

Fréchet distance

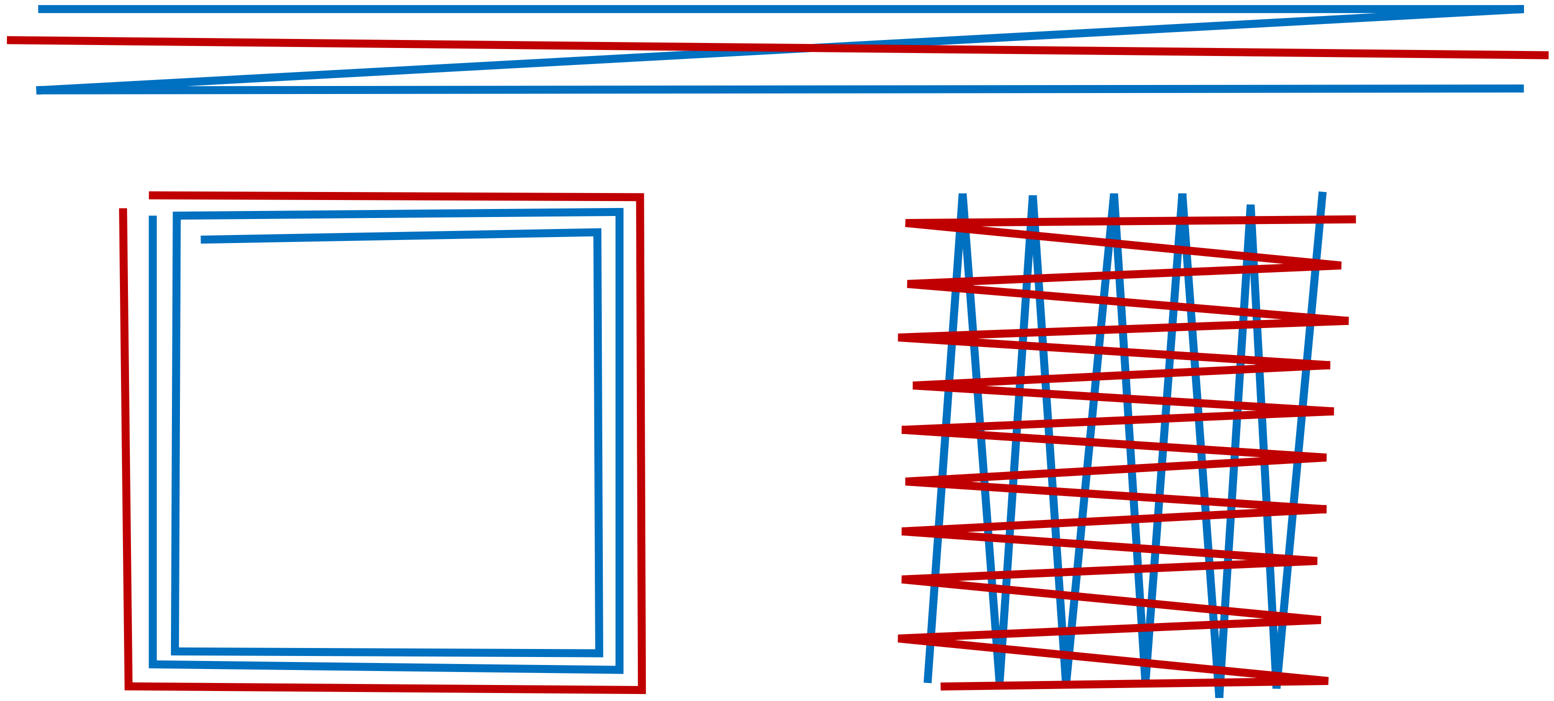
- Definition:

$$\inf_{\alpha, \beta} \left(\max_{t \in [0,1]} \text{dist}(\alpha(t), \beta(t)) \right)$$

- Choosing α, β is choosing the relative "speeds"
- Bottleneck distance due to the max over t
- The Fréchet distance is never smaller than the Hausdorff distance; often they are the same

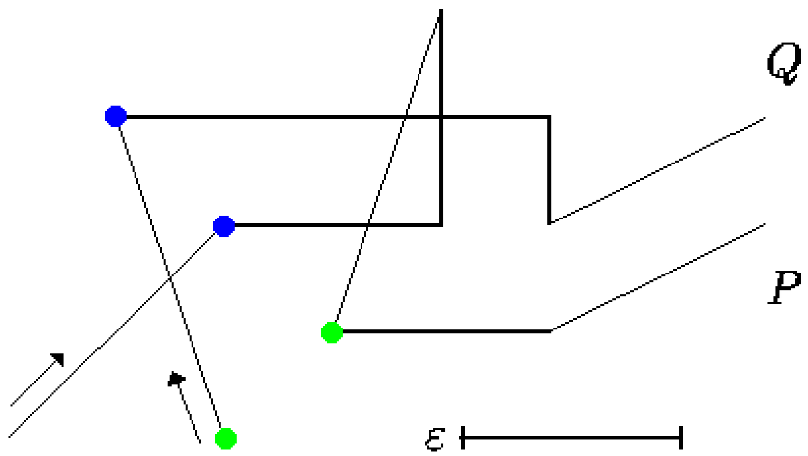
Fréchet distance

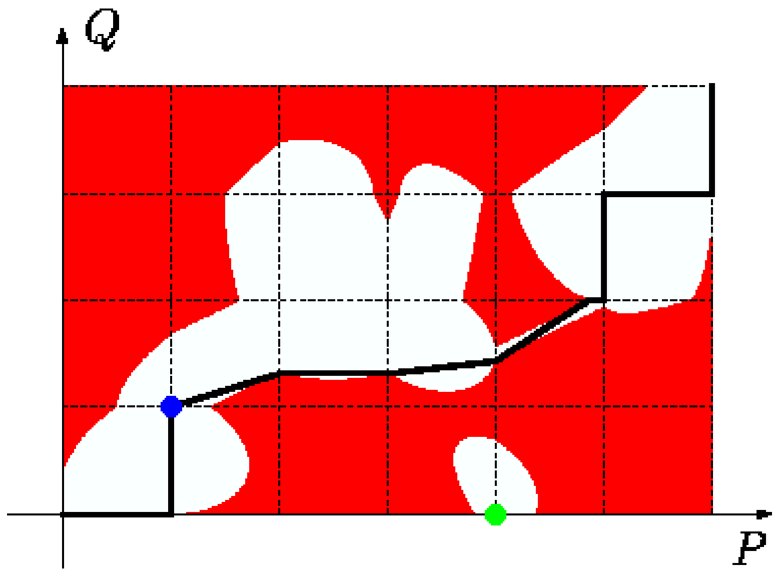
- When are the Fréchet distance and Hausdorff distance clearly different?



Fréchet distance

- Computation using the free-space diagram





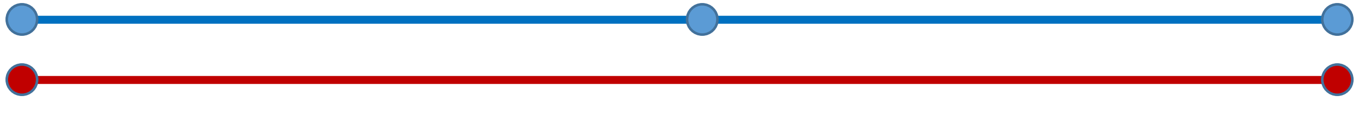
free-space diagram to decide whether the Fréchet distance is at most ε

In-class exercise

- Suppose we want to compare shapes like the ones shown. Which measure appears better:
  - Hausdorff distance
  - Frechet distance (on the boundaries)
  - Area of symmetric difference
- When does which one not work well?
- Think and discuss with your neighbors

Discrete Fréchet distance

- The *discrete* Fréchet distance is like the Fréchet distance, but only measured between vertices
- Vertices must be visited in the right order, but a vertex can be used more than once
- The discrete Fréchet distance can be larger or smaller than the normal Fréchet distance
- The discrete Fréchet distance can be computed in O(*nm*) time by standard dynamic programming

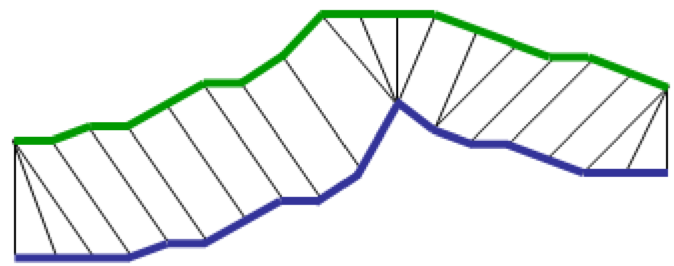


Dynamic Time Warping

- Popular distance measure in time series analysis
- Uses summed distances, not a bottleneck distance
- Uses only distances between vertices

A[1],…, A[n]

B[1],…, B[m]



DTW(A[i..n], B[j..n]) = min

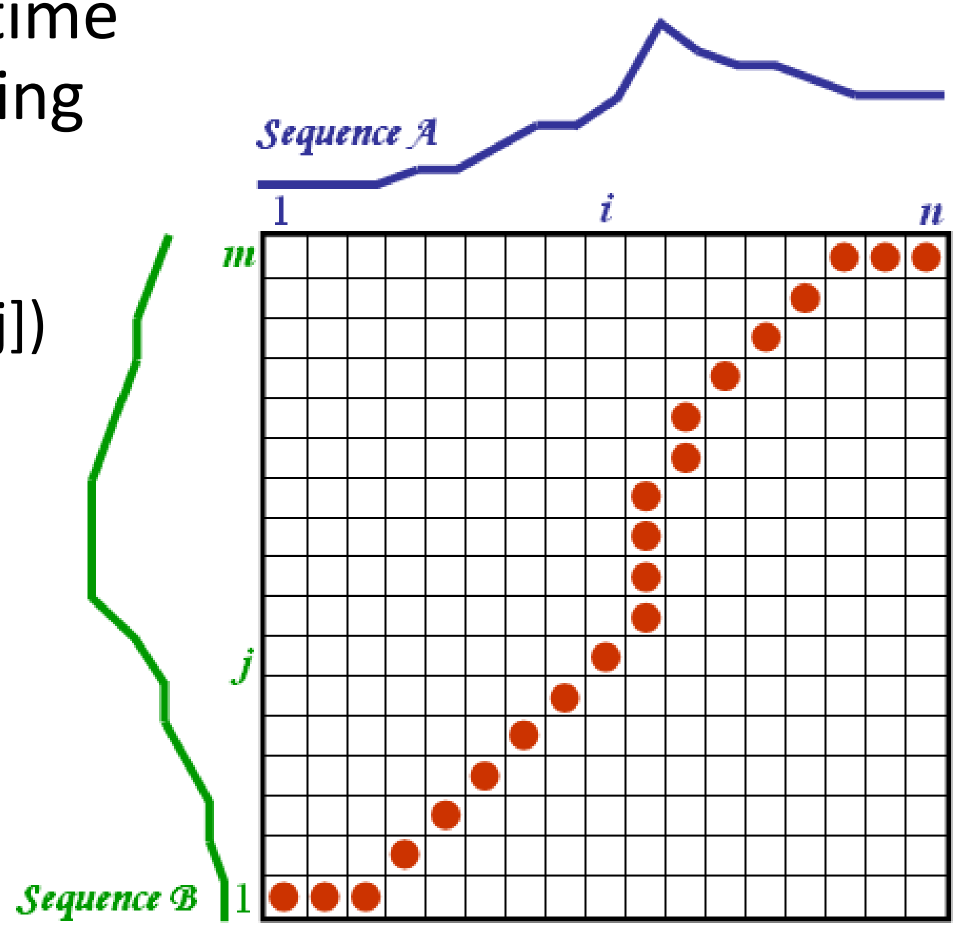
- dist(A[i],B[j]) + DTW(A[i+1..n], B[j+1..m])
- dist(A[i],B[j]) + DTW(A[i..n], B[j+1..m])
- dist(A[i],B[j]) + DTW(A[i+1..n], B[j..m])

Dynamic Time Warping

- Computable in O(*nm*) time by dynamic programming

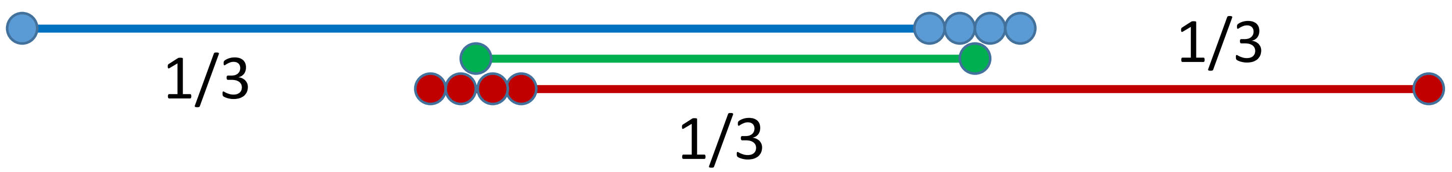
Matrix M with dist(A[i], B[j]) in entry M[i,j]

The DTW distance is the cost of the cheapest path



Dynamic Time Warping

- DTW distance is not a metric: it does not satisfy the triangle inequality



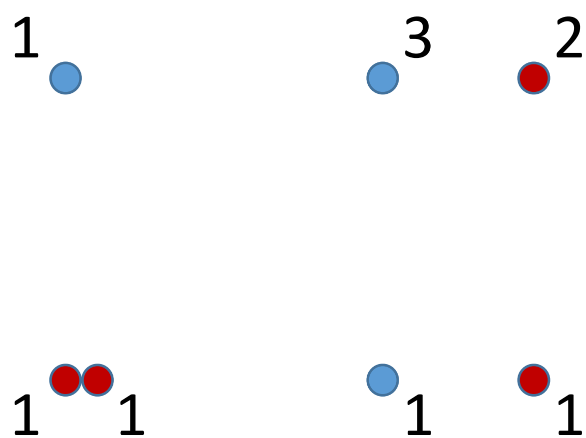
DTW(blue, red) ≈ 5/3

DTW(blue, green) ≈ 1/3
DTW(green, red) ≈ 1/3

DTW(blue, red) > DTW(blue, green) + DTW(green, red)

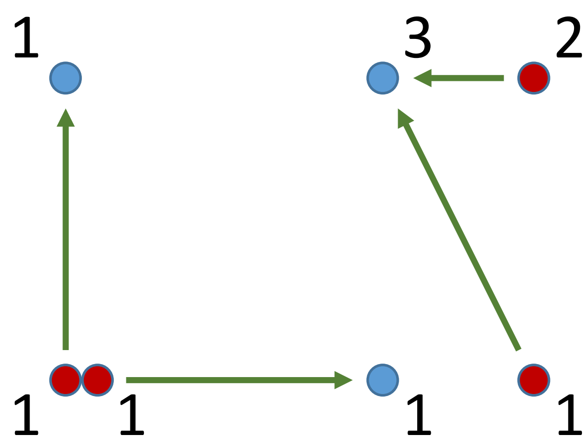
Earth Mover's Distance

- Metric for distance between two weighted point sets with the same total weight
- Captures the minimum amount of energy needed to transport the weight from the one set to the other, where: energy = weight x distance



Earth Mover's Distance

- Metric for distance between two weighted point sets with the same total weight
- Captures the minimum amount of energy needed to transport the weight from the one set to the other, where: energy = weight x distance

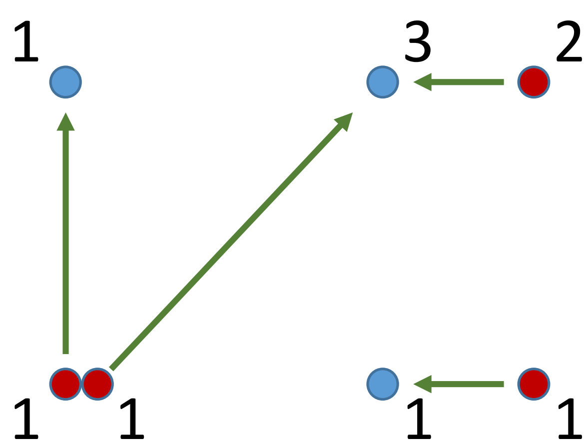


$1 + 1 + 2 \times \frac{1}{2} + \sqrt{1\frac{1}{4}}$

$= 3 + \sqrt{1\frac{1}{4}} = 4.12$

Earth Mover’s Distance

- Metric for distance between two weighted point sets with the same total weight
- Captures the minimum amount of energy needed to transport the weight from the one set to the other, where: energy = weight x distance



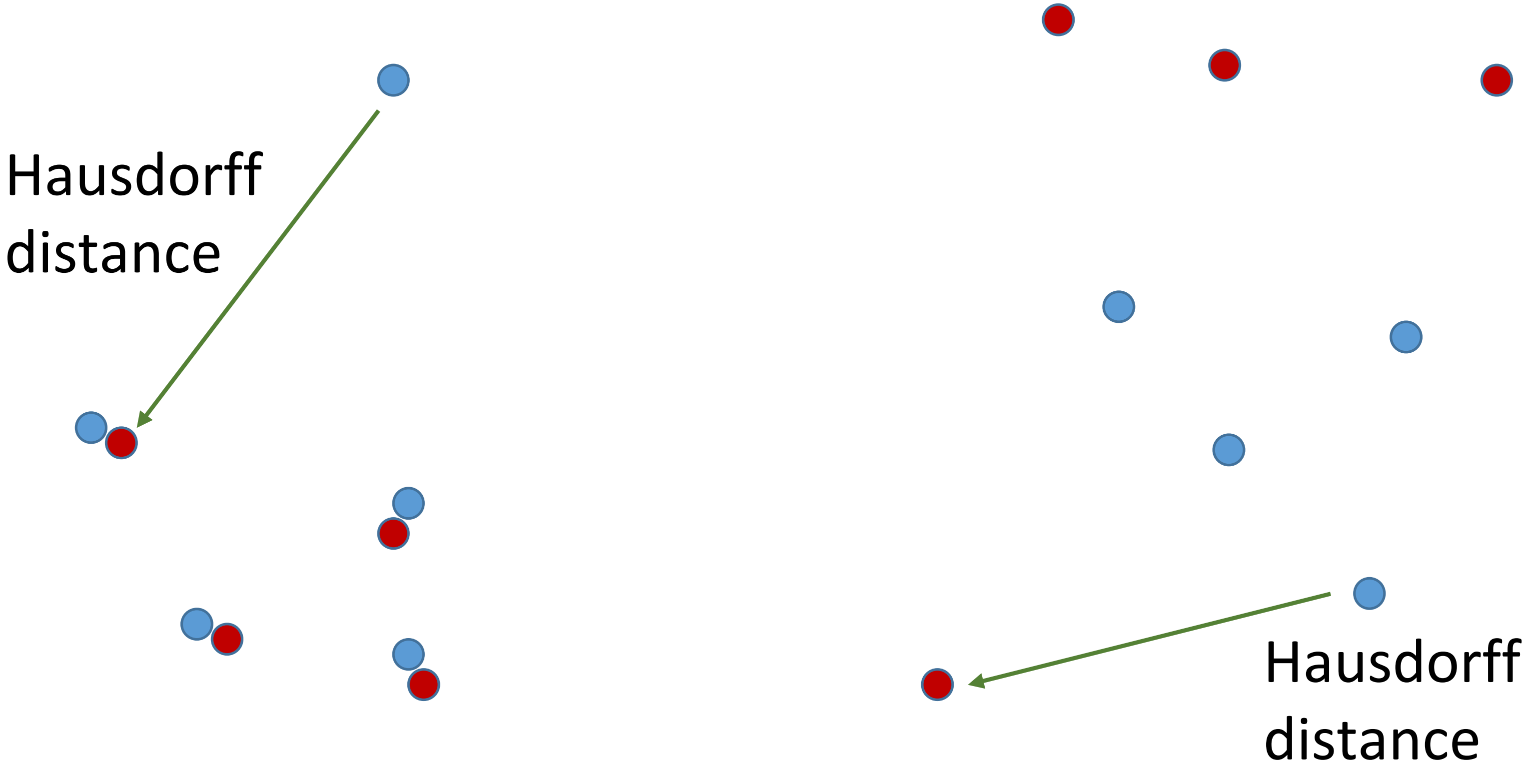
$$1 + \tfrac{1}{2} + 2 \times \tfrac{1}{2} + \sqrt{2}$$
$$= 2\tfrac{1}{2} + \sqrt{2} = 3.91$$

Earth Mover's Distance

- Also known as the Wasserstein distance/metric
- Computable in $O(n^3)$ time when there are *n* points, using a solution to the assignment problem (Hungarian algorithm)

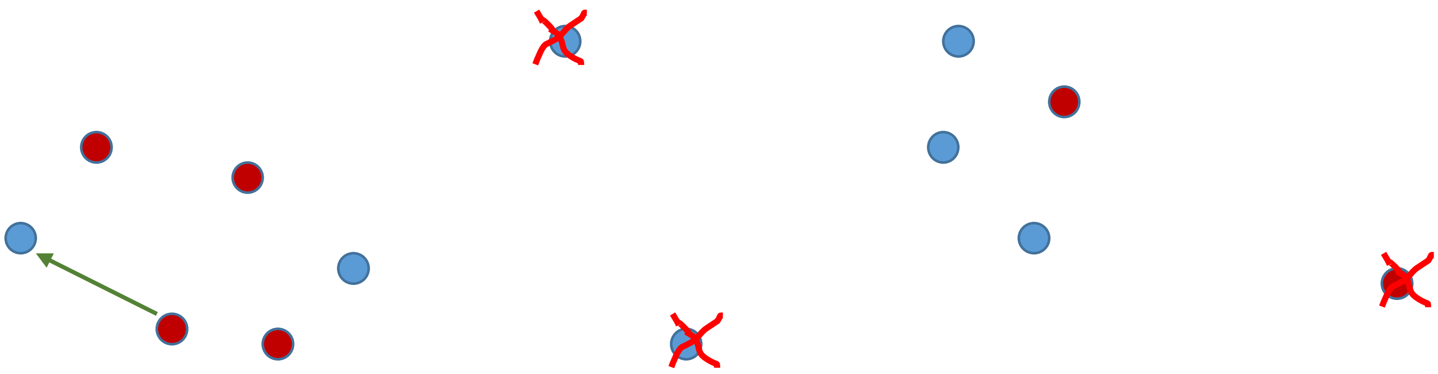
Outliers and measures

- Outliers can influence bottleneck measures significantly, but also sum-of-squares measures and (less so) sum measures



Outliers and measures

- Solutions include:
  - Removing outliers in preprocessing
  - Redefining the measure to not include a small subset of the data
  - Using a different aggregation, like sum-of-square-roots, when sum is considered too sensitive to outliers

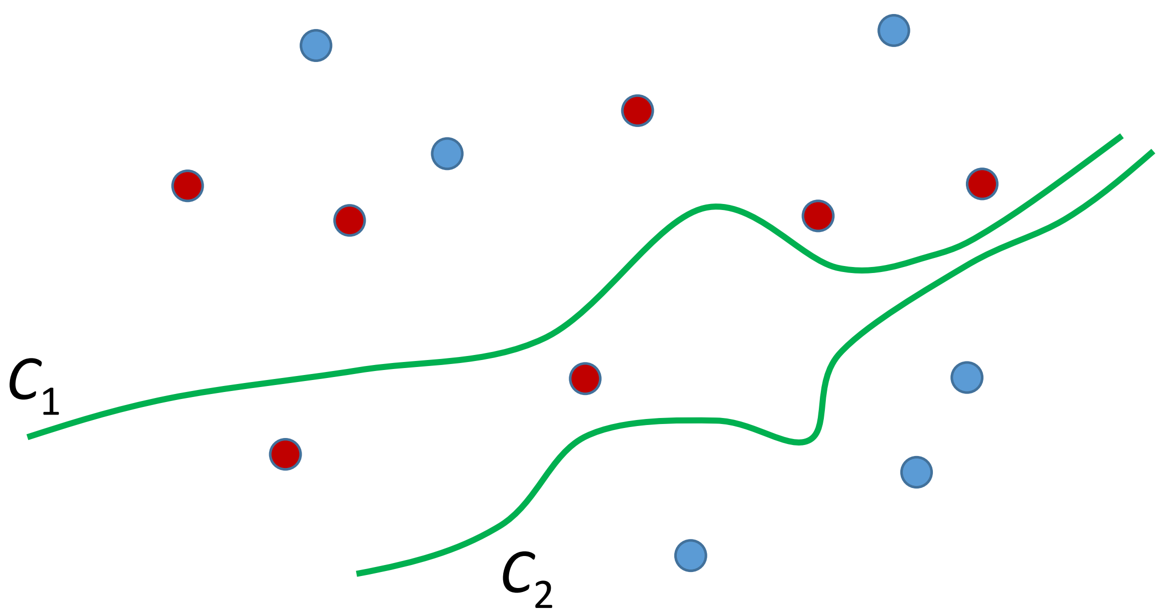


Designing measures

- A balance between simplicity and relevance
- Simplicity:
  - Intuitively easy to understand
  - Easy to define properly, mathematically
  - Easy and efficient to compute
- Relevance:
  - Captures the intuitive notion well – no more and no less
  - Can discriminate differences well

Designing measures

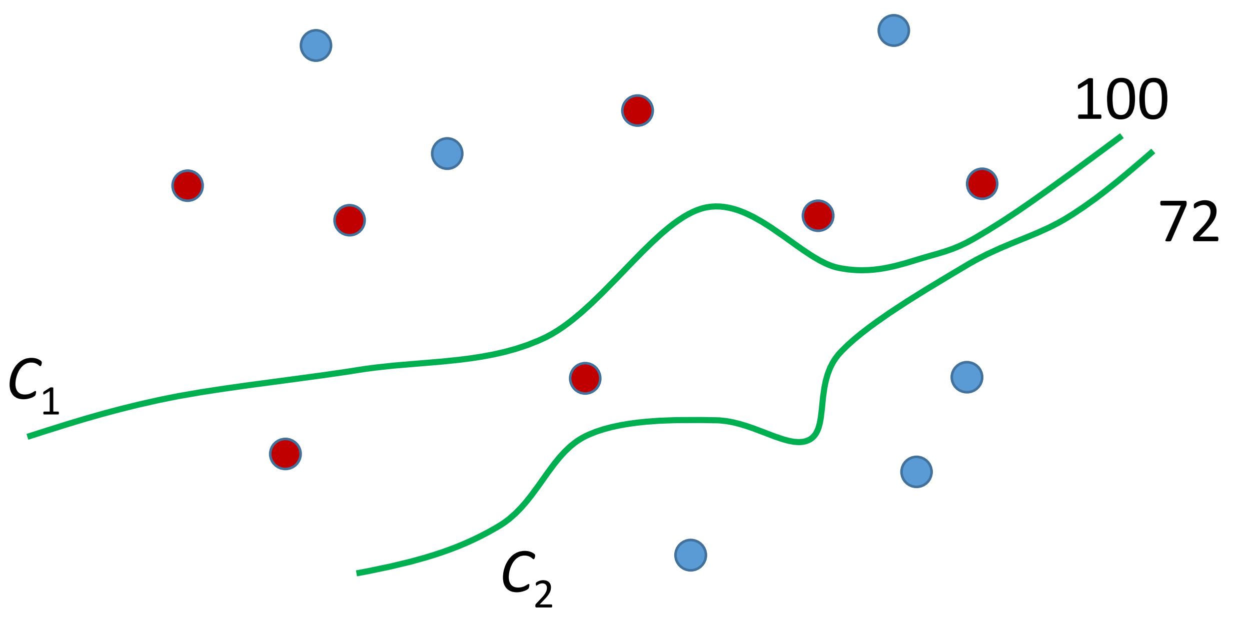
- Example 1: Given a set of red points *R* and a set of blue points *B*, design a measure (score) that captures for any curve *C*, that *C* is close to *R* and not close to *B*



C_1 should get a higher value in the measure than C_2

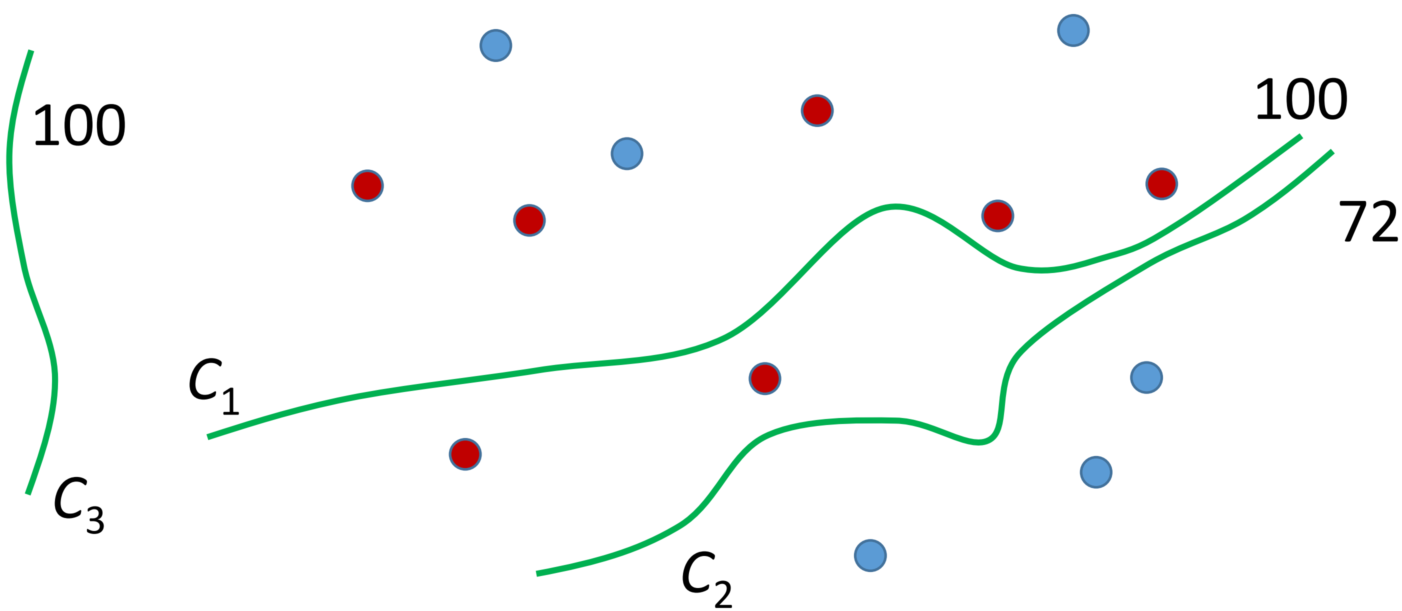
Designing measures

- Possibility 1: Percentage of the length of *C* that is closer to *R* than to *B*, based on closest point



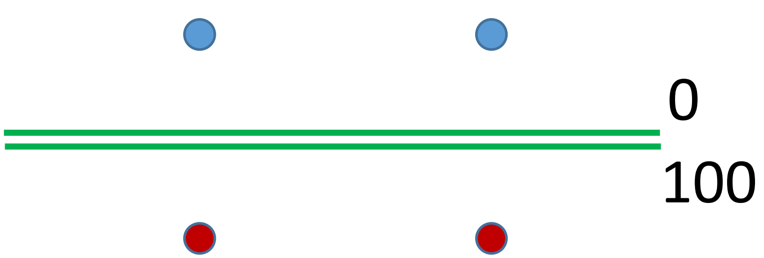
Designing measures

- Possibility 1: Percentage of the length of *C* that is closer to *R* than to *B*, based on closest point
  - Does not capture closeness itself; a curve twice as far may still get a score of 100



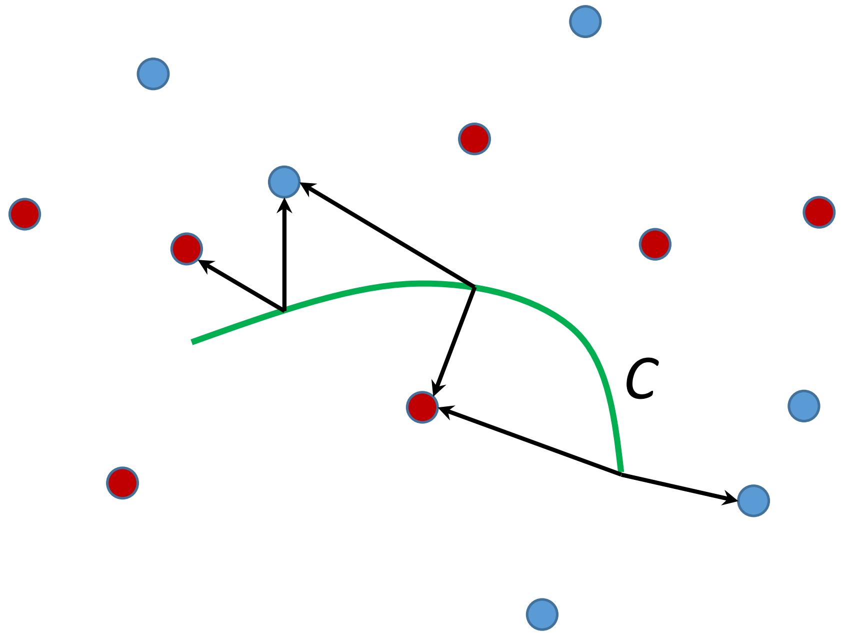
Designing measures

- Possibility 1: Percentage of the length of *C* that is closer to *R* than to *B*, based on closest point
  - Does not capture closeness itself; a curve twice as far may still get a score of 100
  - Not “robust”: a small movement of the curve can change its score from 0 to 100
  - When there are no blue points, any curve gets score 100 (so it does not capture that *C* is close to *R*)



Designing measures

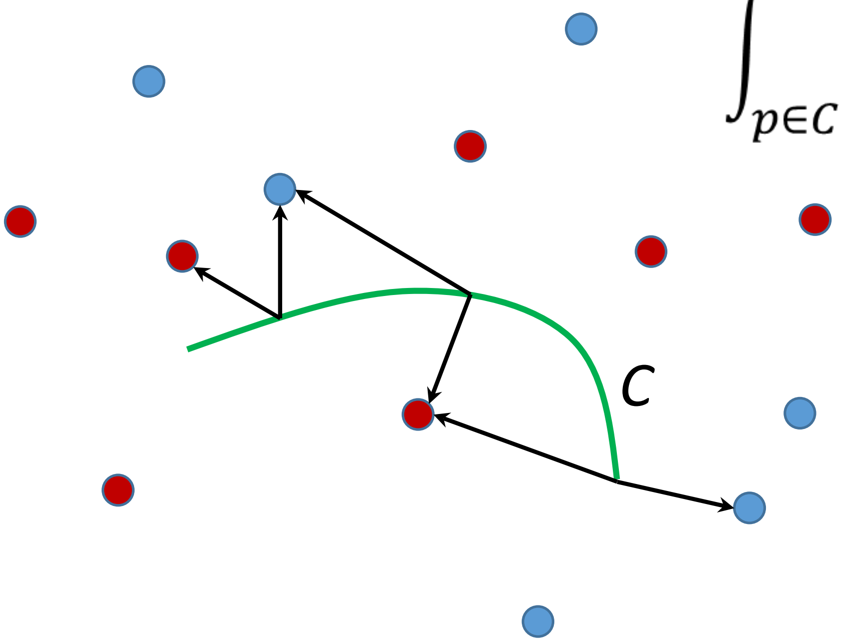
- Possibility 2: Average (over the curve length) of the distance to the nearest blue point – distance to the nearest red point



Designing measures

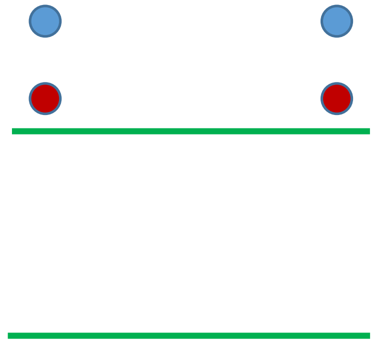
- Possibility 2: Average (over the curve length) of the distance to the nearest blue point – distance to the nearest red point

$$\int_{p \in C} \frac{(dist(p,b) - dist(p,r))dp}{\|C\|}$$



Designing measures

- Possibility 2: Average (over the curve length) of the distance to the nearest blue point – distance to the nearest red point
  - Robust
  - Not scale-invariant
  - Does not capture closeness to *R*, only relative to *B*
  - Does not work when there are no blue points

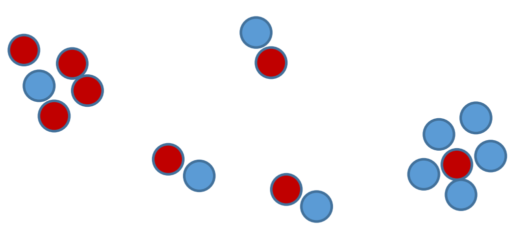


nearly same score

Designing measures

- Example 2: Given a set of red points *R* and a set of blue points *B*, design a distance measure for them

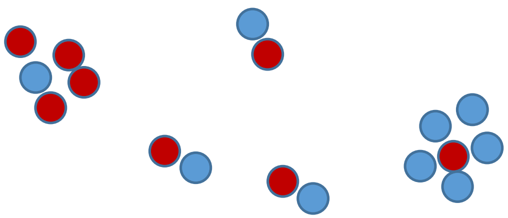
- Immediate question: Are *R* and *B* samples from a region, and we are really interested in how much these regions are alike, or are *R* and *B* really point data (e.g. locations of burglaries and car break-ins)?



In the first case these point sets are very similar, in the second case they are not

Designing measures

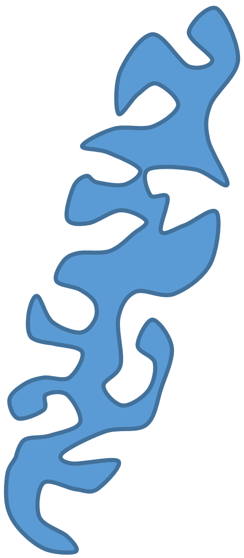
- In the first case: reconstruct the regions (e.g. by alpha-shapes) and use area of symmetric difference
- Alternatively, use the Hausdorff distance
- In the second case: equalize the total weights in the two sets by making the points in the smaller set heavier than 1, and use the Earth Mover's Distance



Combining measures

Suppose we have a measure in [0,1] for elongatedness of a shape and another one for frilliness, called *E* and *F*

How can we combine these into a score for both elongatedness and frilliness?



Combining measures

Suppose we have a measure in [0,1] for elongatedness of a shape and another one for frilliness, called E and F

How can we combine these into a score for both elongatedness and frilliness?

- Weighted linear combination: $\alpha E + (1-\alpha) F$ with $\alpha \in [0,1]$

Combining measures

Suppose we have a measure in [0,1] for elongatedness of a shape and another one for frilliness, called E and F

How can we combine these into a score for both elongatedness and frilliness?

- Weighted linear combination: $\alpha E + (1-\alpha) F$ with $\alpha \in [0,1]$
- Multiplication: $E F$

Combining measures

Suppose we have a measure in [0,1] for elongatedness of a shape and another one for frilliness, called *E* and *F*

How can we combine these into a score for both elongatedness and frilliness?

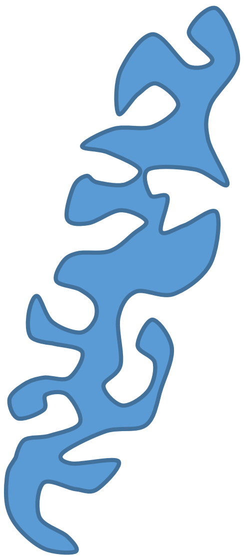
- Weighted linear combination: $\alpha E + (1-\alpha) F$ with $\alpha \in [0,1]$
- Multiplication: $E F$
- Weighted version: $E^{\alpha} F^{1-\alpha}$ with $\alpha \in [0,1]$

Combining measures

Suppose we have a measure in [0,1] for elongatedness of a shape and another one for frilliness, called *E* and *F*

How can we combine these into a score for both elongatedness and frilliness?

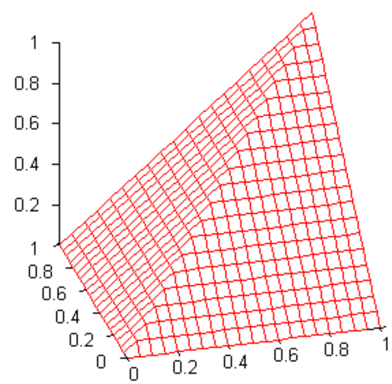
elongated	frilly	combined WLC α = 0.5	combined Mult. α = 0.5
0	0	0	0
1	1	1	1
0	1	0.5	0
0.5	0.5	0.5	0.5
0.5	1	0.75	0.707
0.75	0.75	0.75	0.75

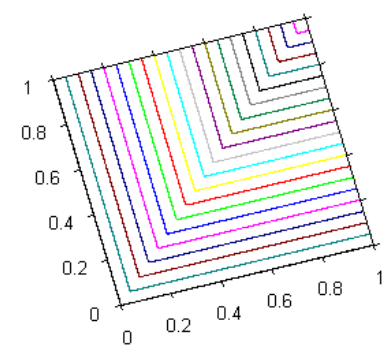


t-norms

- A t-norm is a function T: [0, 1] × [0, 1] → [0, 1] which satisfies the following properties:
  - Commutativity: T(a, b) = T(b, a)
  - Monotonicity: T(a, b) ≤ T(c, d) if a ≤ c and b ≤ d
  - Associativity: T(a, T(b, c)) = T(T(a, b), c)
  - The number 1 acts as identity element: T(a, 1) = a

Minimum t-norm
T(a, b) = min(a, b)

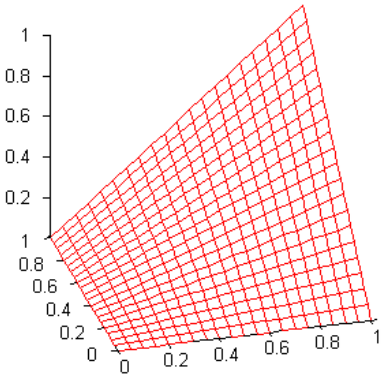


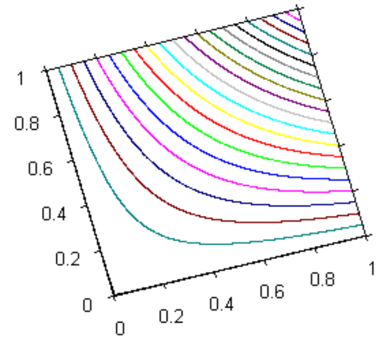


t-norms

- A t-norm is a function T: [0, 1] × [0, 1] → [0, 1] which satisfies the following properties:
  - Commutativity: T(a, b) = T(b, a)
  - Monotonicity: T(a, b) ≤ T(c, d) if a ≤ c and b ≤ d
  - Associativity: T(a, T(b, c)) = T(T(a, b), c)
  - The number 1 acts as identity element: T(a, 1) = a

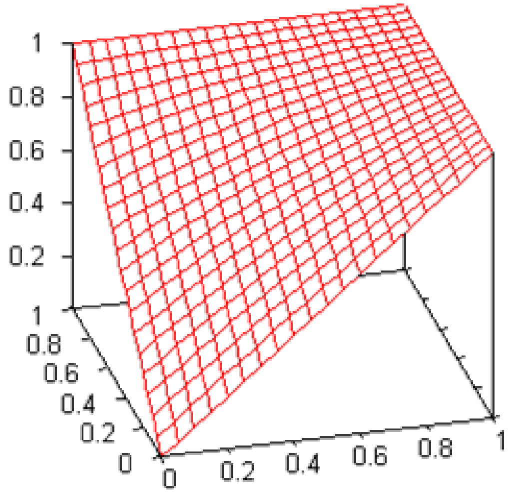
Product t-norm
T(a, b) = ab

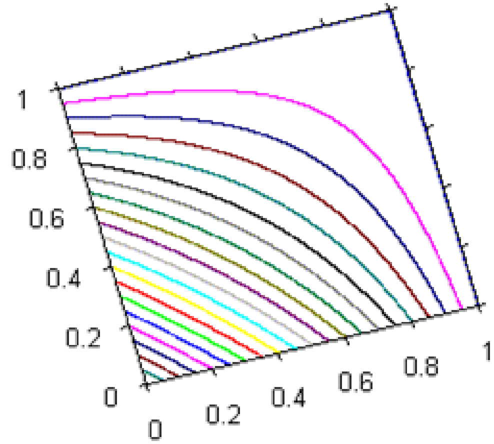




t-conorms

- Similar to t-norms but 0 is the identity: T(a, 0) = a
- Example: Einstein sum T(a, b) = (a + b) / (1 + ab)





Summary

- Measures and metrics are useful to have things to optimize and things to compare quantitatively
- There are many established measures and metrics
- Sometimes one has to define one's own measure or metric for specific situations
- Computation of measures requires geometric algorithms
- Combining measures can be done using the concept of t-norms and t-conorms