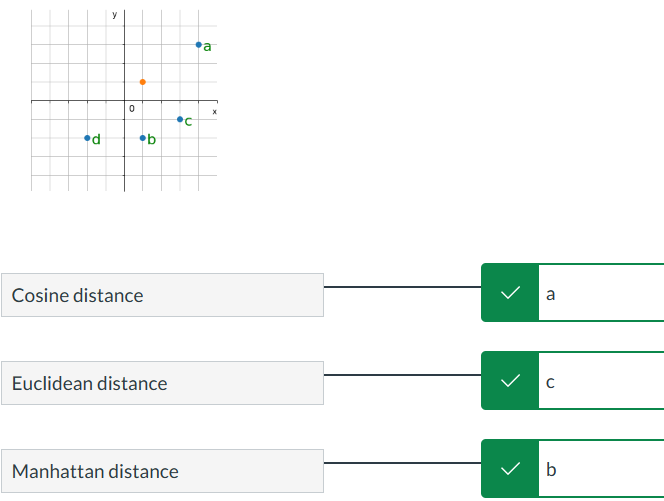
5.1 & 5.2

1. Which blue point is closest to the orange one, according to the following distance measures?



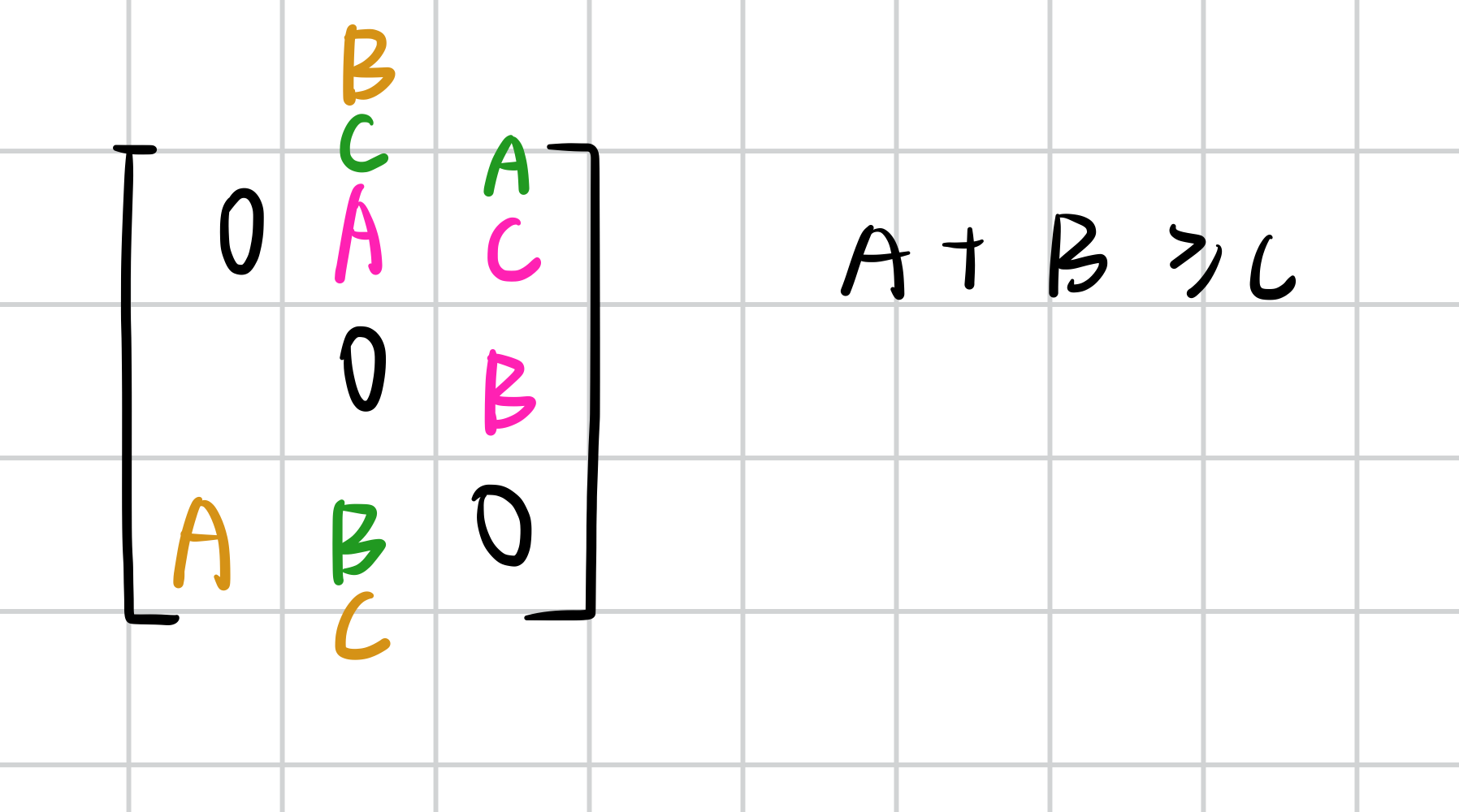
1. Which of the following metric properties **do not apply** as-is for a **pseudometric**?

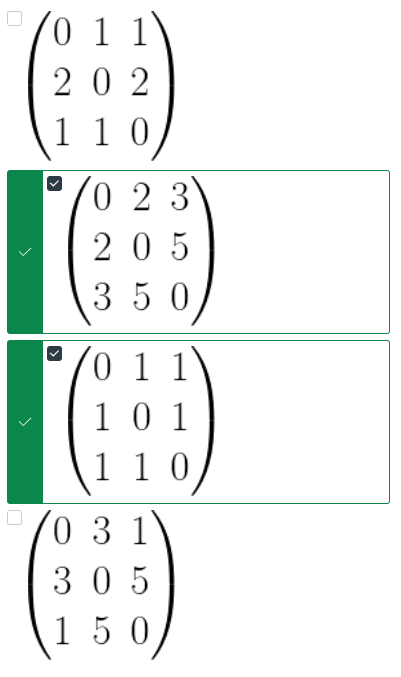
* Triangle inequality
* Identity of indiscernibles
* Symmetry

1. Yes - Assume you have a metric and you want to push distinct points further apart, so you define as follows:

for some fixed positive t. Is f a metric?

1. Which of the following are valid distance matrices for a metric?

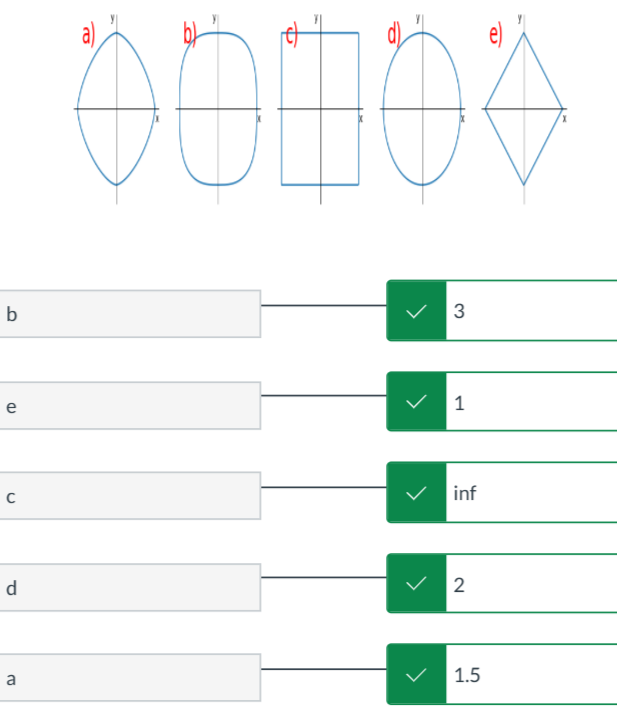
****



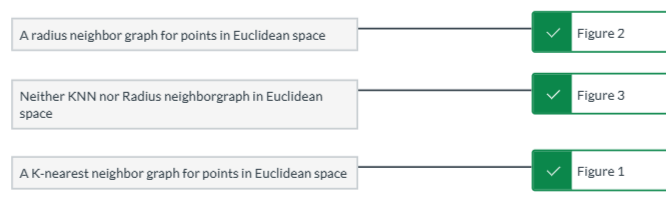
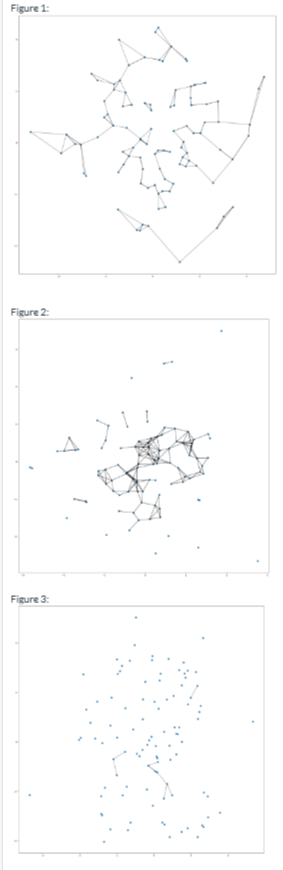
1. True - Is the following statement true or false? "All metrics are equivarivant to point scaling, meaning the following holds:
2. Suppose we are given a dataset of distinct points where the smallest distance d between any two points is m>0 and the largest distance between any two of the points is M>0. Which of the following statements are true?

* As we increase k in KNN(X, d, k) the geodesic distance in the graph between two vertices increases also
* In the K nearest neighbor graph KNN(X, d, k) every vertex has exactly k outgoing directed edges emanating from as long as there are at least k elements in X
* The construction of KNN(X, d, k) requires the data X to lie in a vector space.
* As we increase r in RN(X, d, r) the geodesic distance in the graph between two vertices may either increase or shrink
* The bigger k in KNN(X, d, k), the better the approximation of geodesic distance of an underlying surface/manifold on which the data was generated.
* For r>M the radius neighborhood graph has n vertices and n choose 2 edges connecting any possible combination of two distinct vertices.
* For any dataset X of n points in Euclidean space, all the edges of KNN(X, d, k) are also edges in RN(X, d, r) for k<r
* For 0<r<m the radius neighborhood graph has n vertices and not a single edge connecting distinct vertices.
* In the K nearest neighbor graph for X, it is possible that we have a directed edge from to but not from to

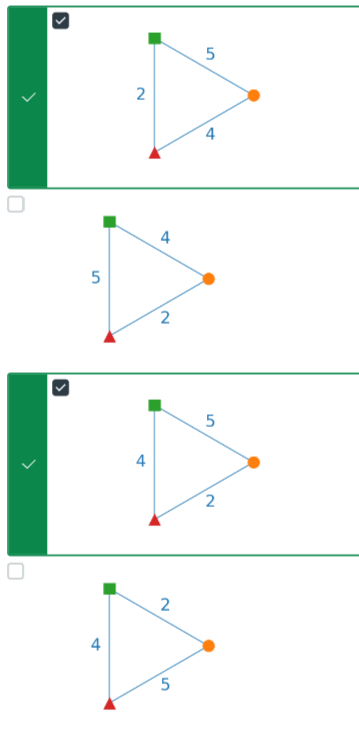
1. Displayed below are different levelsets in the plane under different metrics, p∈{1,1.5,2,3,∞}. Match the level sets with the correct p parameter:



1. An norm of a vector z∈ is defined as ∥z∥p=(z,0).As differentiability is a key property for gradient-based optimization methods, which of the following functions are differentiable over z **everywhere on** ?
2. What do we see in these graphs?



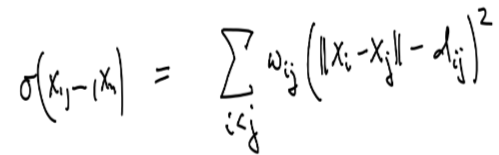
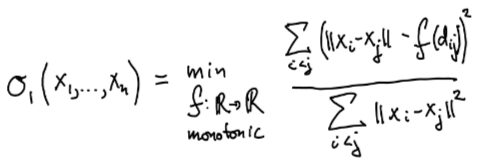
1. False - Cosine similarity is well-defined on whole
2. The following graphs graphically show the metric distances between three points denoted with a square, circle and a triangle.Select **exactly two** of the depicted metrics and such that minimum between them metrics is in fact not a metric. This gives a proof that the minimum between two metrics is not always a metric.



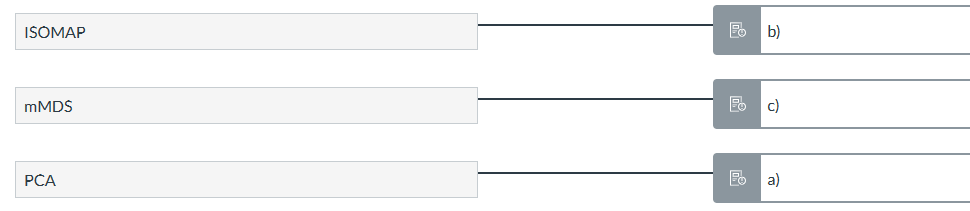
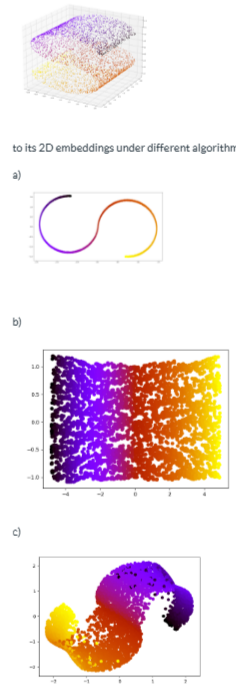
1. False - "The KL-divergence is a metric between probability density functions.

5.3

1. Which of these statements are true?

* Suppose the mMDS strain functionhas value p at . Then also for any orthogonal matrix R and fixed translation by a vector c. In particular, more than one optimal solution exists to recovering an embedding with mMDS.
* Choosing larger values for for a particular choice of i and j increases the relative importance of matching the in the mMDS function compared to matching the distances for (s, t) not equal to (i, j) in the objective function
* The nMDS loss does not change value at all if we replace the input distances by

1. Match the the following 3D dataset



1. **True - Suppose we are given the pairwise Euclidean distances between n datapoints in some Euclidean space. cMDS is able to determine n vectors in ∈ such that the pairwise Euclidean distances between these vectors is exactly equal to as long as q is sufficiently high dimensional.**
2. Which of the following are concerns when running ISOMAP on a very large dataset of n vectors in d-dimensional Euclidean space?

* We need to construct a neighborhood graph for a large number of inputs - this may have a very high computational cost and not be realistically doable in reasonable time, in particular when choosing parameters k (for KNN) and r (for RN) too large.
* We need to store a dxd distance matrix of pairwise distances that serves as input to MDS
* We do not need to be worried at all. ISOMAP has linear complexity in n.
* We need to store a nxn distance matrix of pairwise distances that serves as input to MDS
* Computing geodesic distance approximations requires running a shortest path algorithm such as Dijkstra's algorithm that may also contribute to computational complexity challenge of applying ISOMAP.

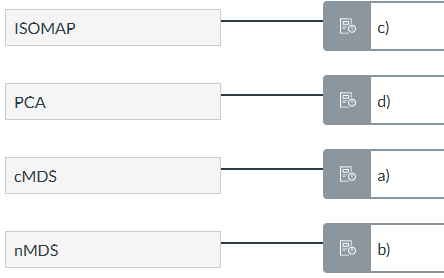
1. Match the following algorithms to the most appropriate input data:

a) Euclidean data for which we are given only a distance matrix, but not the original coordinates in Euclidean space

b) Data, where care mostly about the relative ranking of similarities between items

c) Data for which we can approximate geodesic distances using Neighborhood graphs

d) Euclidean data that can be projected orthogonally to lower dimensions while retaining the majority of its variance



1. Suppose S denotes the empirical covariance matrix for a dataset of n vectors in d dimensions. Let v be any d dimensional unit norm vector - how can we interpret the following quantity:

* In general it does not have an interpretation - only when v is an Eigenvector of S can we say anything with certainty.
* is the mean of the dataset
* It is equal to the empirical variance of the data under the projection

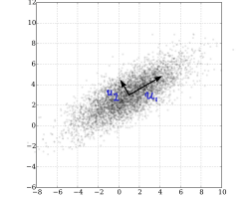
1. The 1st principal component direction in PCA is chosen such that:

* The empirical variance of the data when projected onto the the line along direction is maximized.
* corresponds to the smallest Eigenvalue of the empirical covariance matrix.
* The emprical variance of the data when projected onto the the line along direction is minimized

1. Which of the following are true

* We can use Lagrange multipliers to derive that the first principal direction corresponds to an eigenvector of the empirical covariance matrix with largest Eigenvalue.
* Suppose data is sampled on a 1d subspace (a line) in 2 dimensions and we have at least two distinct samples.  
  In that case the empirical covariance matrix of the data has only 1 non-zero Eigenvalue with a corresponding 1 dimensional Eigenspace in the direction of the line and the zero eigenvalue has a corresponding complementary Eigenspace of dimension 1.
* Each Eigenspace corresponding to an eigenvalue of the empirical covariance matrix S is always exactly one dimensional and therefore spanned by a single unit norm vector that is unique up to plus/minus sign.
* The empirical covariance matrix may have negative eigenvalues.
* PCA can be motivated alternatively as an optimization problem finding a linear projection to k dimensions with the lowest mean squared reconstruction error, but we have not covered this in detail in our lecture notes.

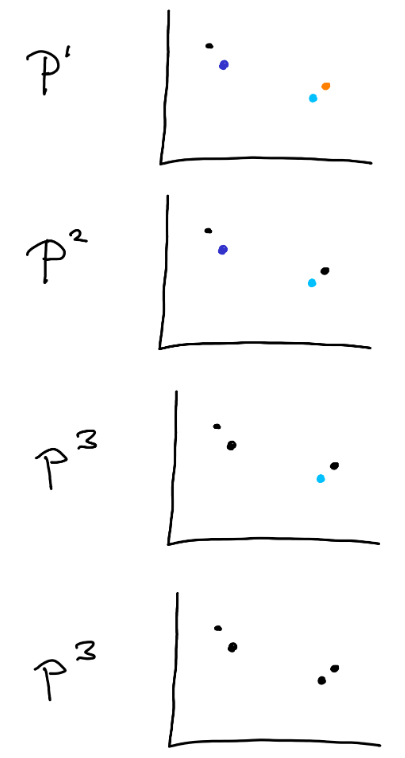
1. What can we conclude about this visualization of a 2d dataset:



Adaptaion of Image: CC BY 4.0 Wikimedia Commons user [NicoguaroLinks.](https://commons.wikimedia.org/wiki/File:GaussianScatterPCA.svg)

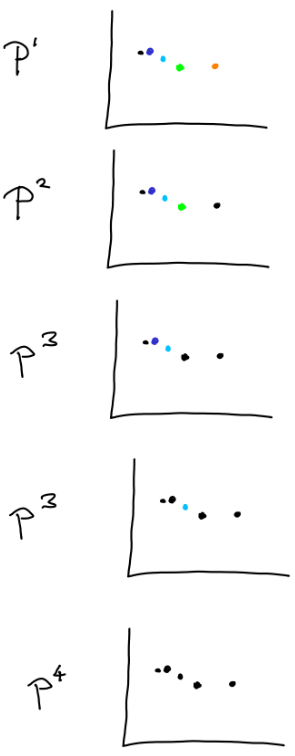
* This dataset looks like it has been generated as samples from a Multivariate Gaussian distribution
* This dataset looks quite suitable for applying a PCA projection onto the first principal component.
* The direction vector looks like an Eigenvector corresponding to the largest Eigenvalue of the empirical covariance matrix for this data.
* The dataset has an interesting covariance structure: One eigenvalue of the empirical covariance matrix is definitely larger than the other.

5.4



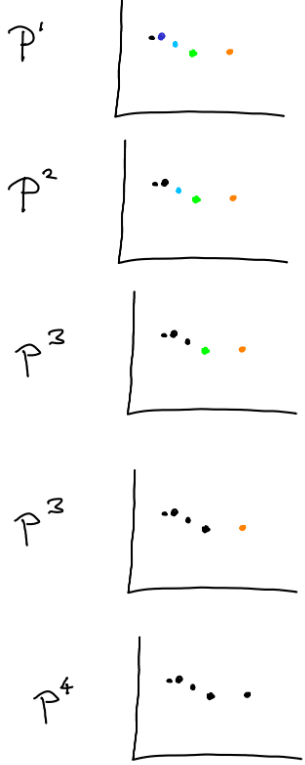
The pictures above show an agglomerative clustering of 4 points in 2 dimensions, where we use the Euclidean metric in 2d to measure distances between individual points, which statements hold?

* This is some other clustering method we have not discussed
* This is the result we would expect from single linkage clustering
* This is the result we would expect from complete linkage clustering



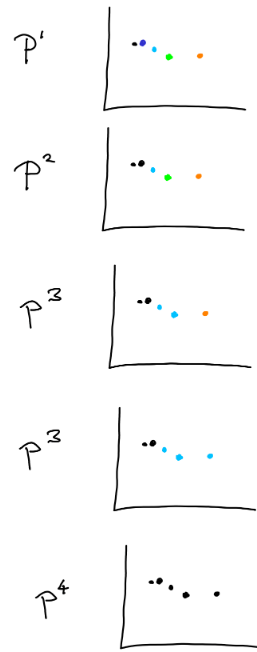
The pictures above show an agglomerative clustering of 5 points in 2 dimensions, where we use the Euclidean metric in 2d to measure distances between individual points, which statements hold?

* This is the result we would expect from complete linkage clustering
* This is the result we would expect from single linkage clustering
* This is some other clustering method we have not discussed. It does not look intuitive



The pictures above show an agglomerative clustering of 5 points in 2 dimensions, where we use the Euclidean metric in 2d to measure distances between individual points, which statements hold?

* This is the result we would expect from single linkage clustering
* This is some other clustering method we have not discussed
* This is the result we would expect from complete linkage clusterin



The pictures above show an agglomerative clustering of 5 points in 2 dimensions, where we use the Euclidean metric in 2d to measure distances between individual points, which statements hold?

* This is the result we would expect from complete linkage clustering
* This is the result we would expect from single linkage clustering
* This is some other clustering method we have not discussed

1. Kleinberg's results that we discuss in the notes imply that:

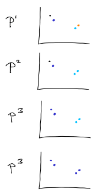
* K Medoids can achieve Richness and Consistency
* Our intuition about what clustering can achieve is wrong if we think that Richness, Consistency and Scale Invariance should all be satisfied simultaneously by any clustering method.
* Single linkage clustering can achieve any two of Richness, Consistency and Scale Invariance at the same time, but it depends on the stopping critereon we choose.
* Complete linkage clustering cannot achieve Richness, Consistency and Scale invariance at the same time

1. The single linkage distance

* Returns the distance between the two closest points a in A and b in B.
* Satisfies d(A, B)=d(B, A)
* Satisfies triangle inequality
* Defines a metric on clusters

1. The complete linkage distance

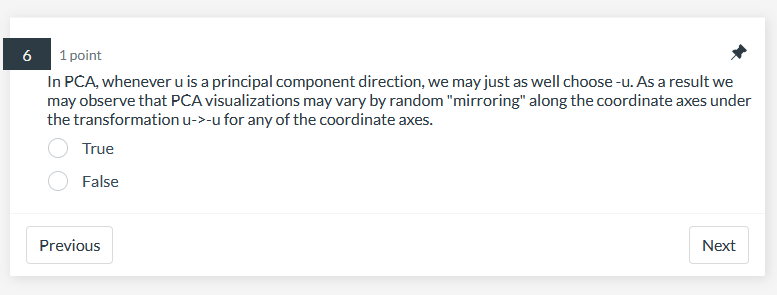
* Returns the distance between the two furthest apart points a in A and b in B.
* Defines a metric on clusters
* Satisfies d(A, B)=d(B, A

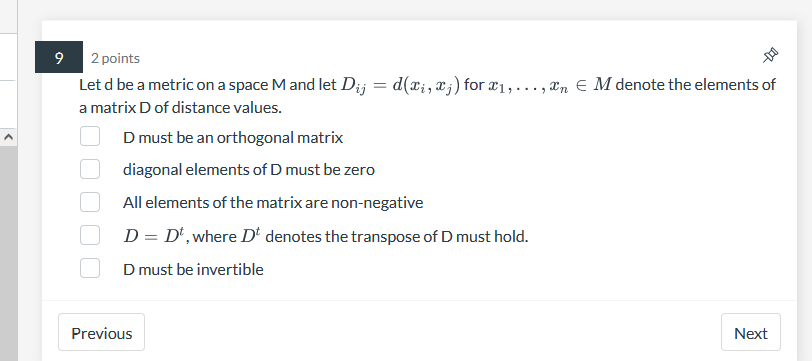


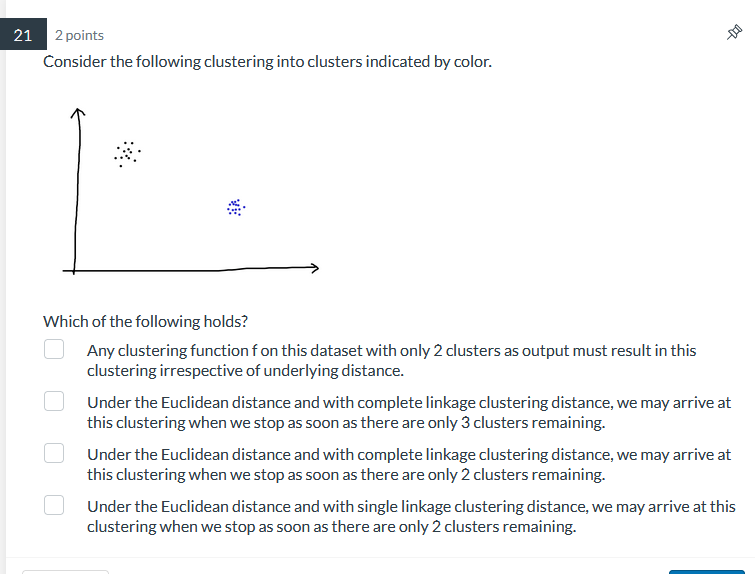
The pictures above show an agglomerative clustering of 4 points in 2 dimensions, where we use the Euclidean metric in 2d to measure distances between individual points, which statements hold?

* This is definitely some other clustering method we have not discussed
* This is the result we would expect from single linkage clustering.
* If we use the k-cluster stopping critereon for k=2, we would return
* This is the result we would expect from average linkage clustering
* This is the result we would expect from complete linkage clustering
* If we use the k-cluster stopping critereon for k=2, we would return

1. True - None of the distances: average linkage, complete linkage or single linkage satisfy all three properties of a metric.







Any clustering function f on this dataset with only 2 clusters as output must result in this clustering irrespective of underlying distance.

Under the Euclidean distance and with complete linkage clustering distance, we may arrive at this clustering when we stop as soon as there are only 3 clusters remaining.

Under the Euclidean distance and with complete linkage clustering distance, we may arrive at this clustering when we stop as soon as there are only 2 clusters remaining.

Under the Euclidean distance and with single linkage clustering distance, we may arrive at this clustering when we stop as soon as there are only 2 clusters remaining