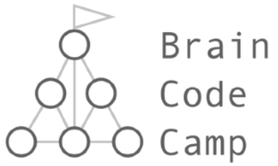
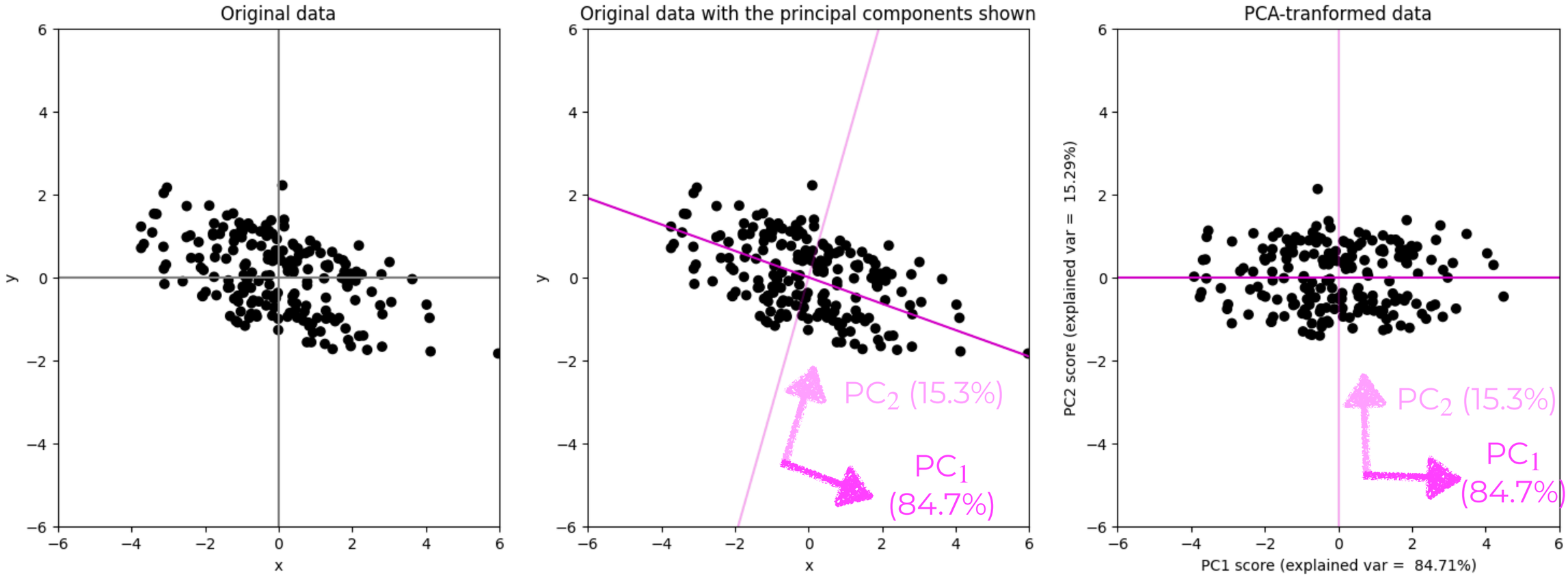


Nonlinear Dimensionality Reduction

Itthi Chatnuntawech

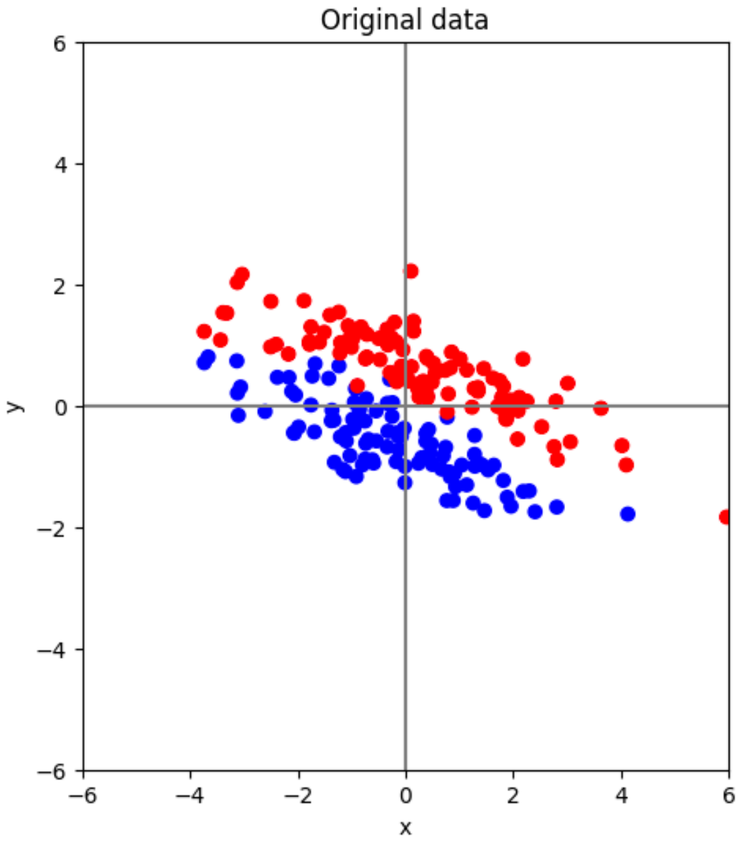
Brain Code Camp

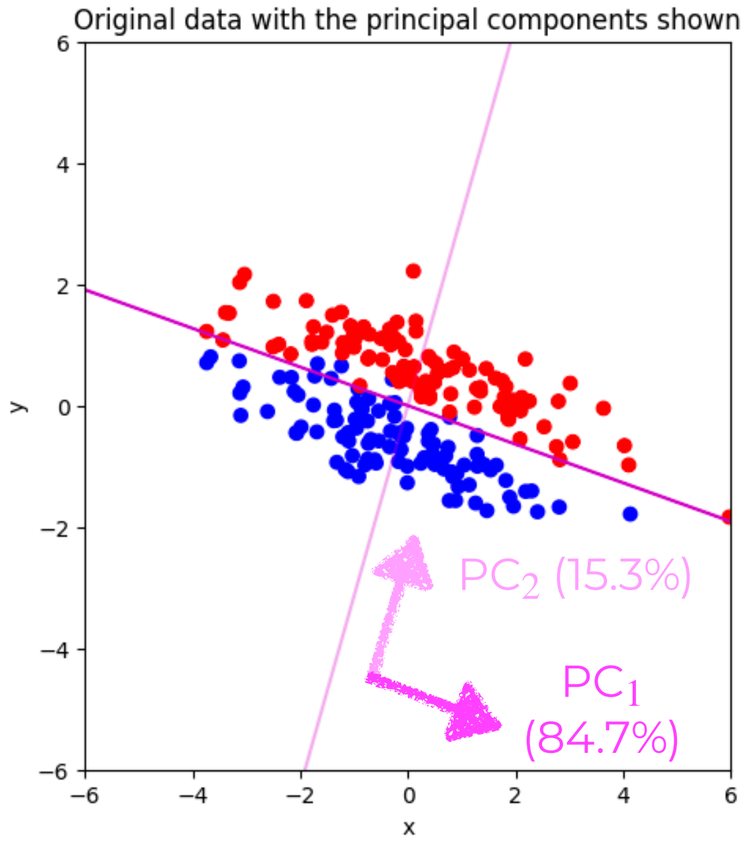
PCA Example

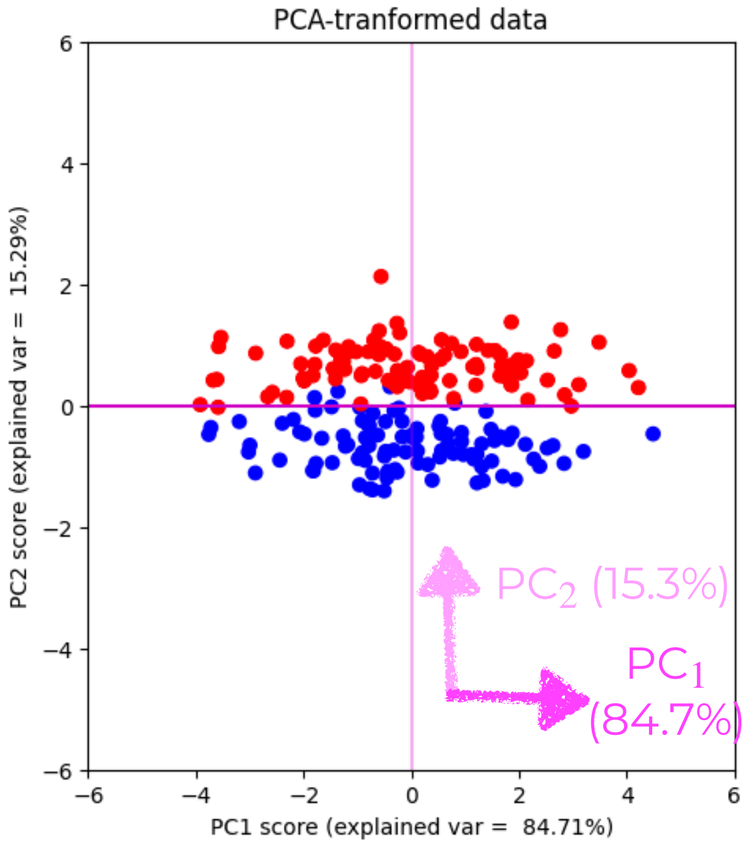


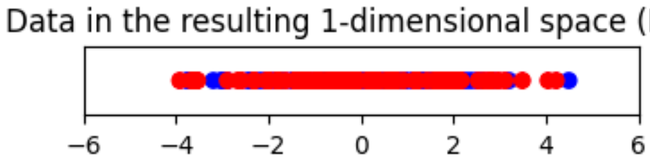
Brain Code Camp

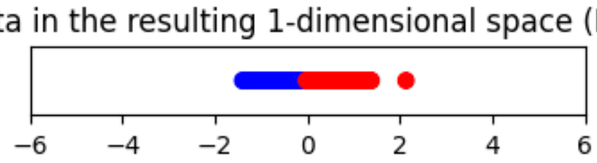
PCA Example





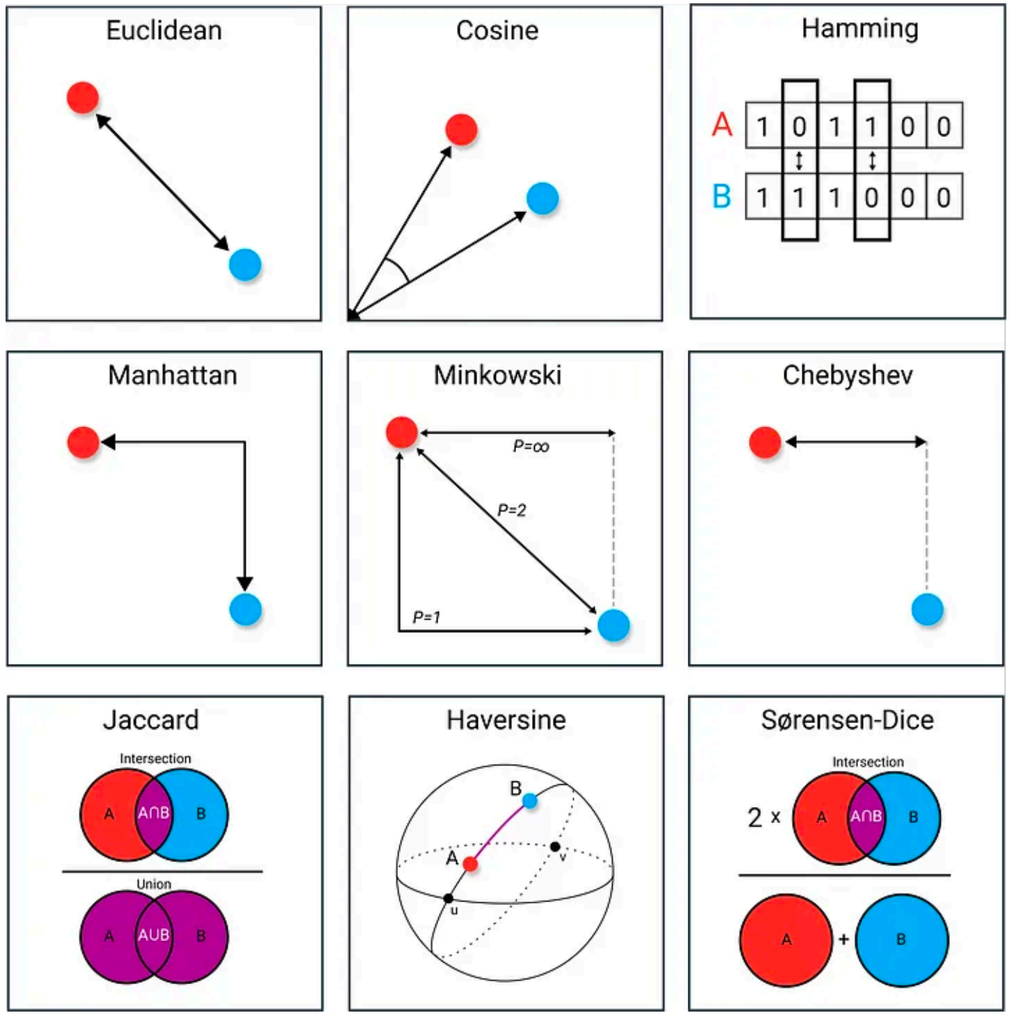






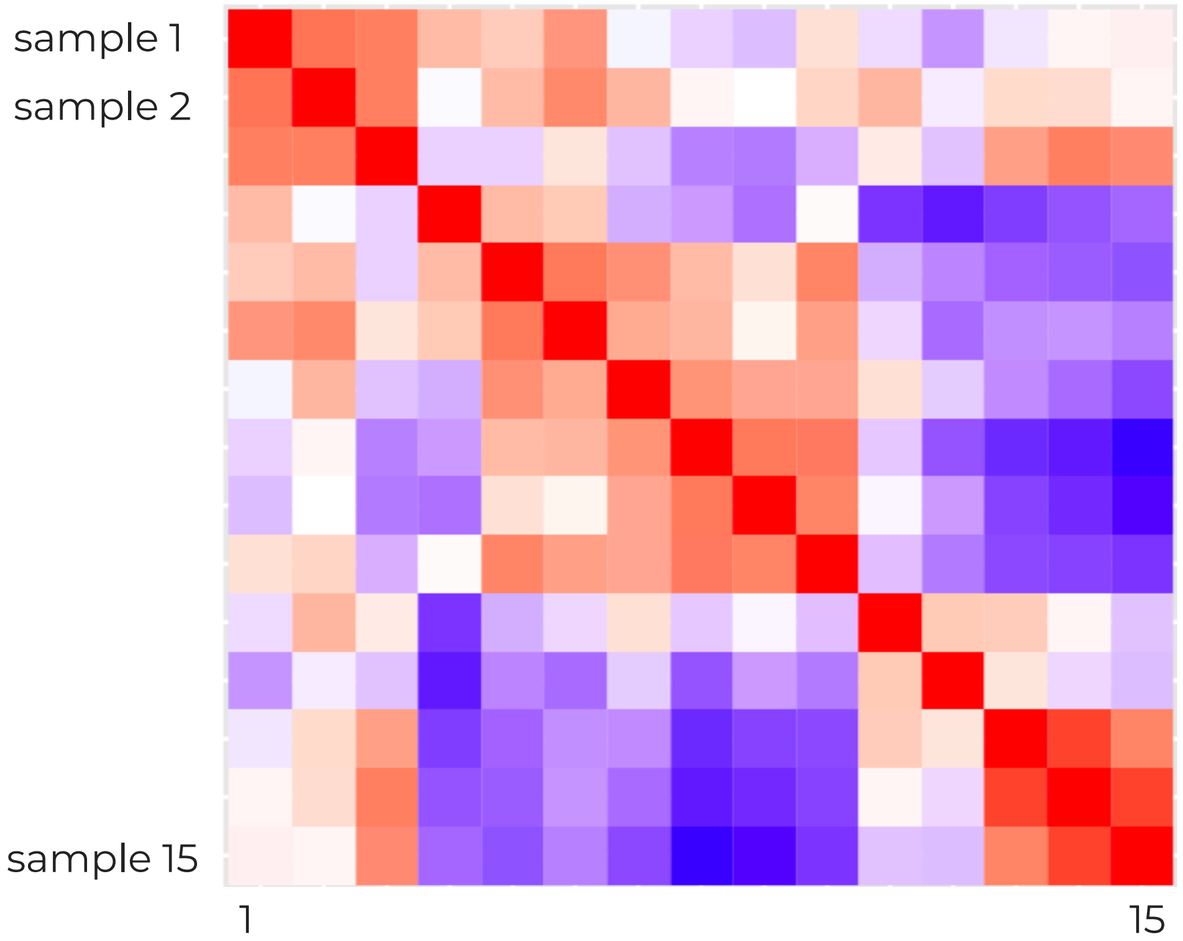
Brain Code Camp

Distance Measures



https://towardsdatascience.com/9-distance-measures-in-data-science-918109d069fa

Pairwise Distance Matrix



The element at row i and column j represents the distance between sample i and sample j

$$D_{ij} = D_{ji}$$

$$D_{ii} = 0$$

Modified from https://www.datanovia.com/en/lessons/clustering-distance-measures/

Brain Code Camp

Generalized Multi-Dimensional Scaling (MDS)

High-dimensional feature space

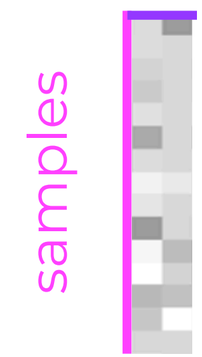
features

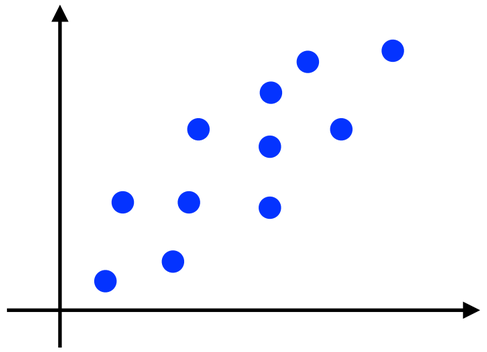
samples

Low-dimensional space

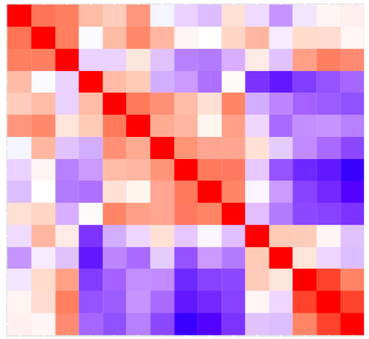
new features

samples

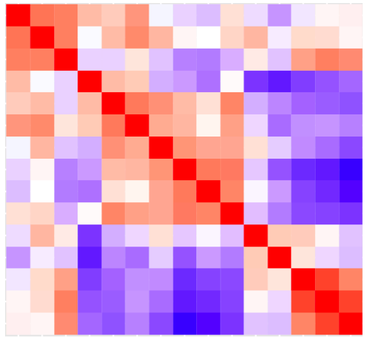




pairwise distance matrix created using any types of distance



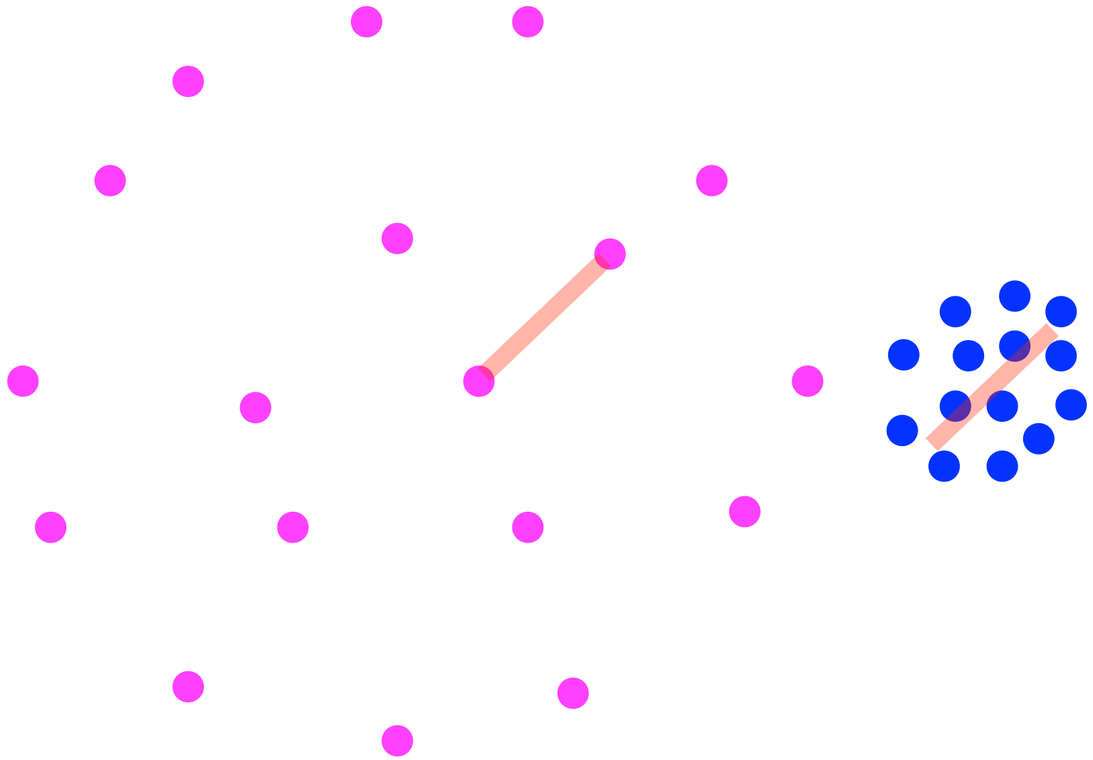
≈



pairwise distance matrix created using any types of distance

Brain Code Camp

Beyond MDS



Should we use distance metrics that take into account data density?

Brain Code Camp

Popular Nonlinear Dimensionality Reduction Methods

t-distributed Stochastic Neighbor Embedding (t-SNE)

sklearn.manifold.TSNE

class `sklearn.manifold.TSNE`(*n_components=2, *, perplexity=30.0, early_exaggeration=12.0, learning_rate='auto', n_iter=1000, n_iter_without_progress=300, min_grad_norm=1e-07, metric='euclidean', metric_params=None, init='pca', verbose=0, random_state=None, method='barnes_hut', angle=0.5, n_jobs=None*) [source]

Uniform Manifold Approximation and Projection (UMAP)

class **umap.umap_.UMAP**(*n_neighbors=15, n_components=2, metric='euclidean', metric_kwds=None, output_metric='euclidean', output_metric_kwds=None, n_epochs=None, learning_rate=1.0, init='spectral', min_dist=0.1, spread=1.0, low_memory=True, n_jobs=-1, set_op_mix_ratio=1.0, local_connectivity=1.0, repulsion_strength=1.0, negative_sample_rate=5, transform_queue_size=4.0, a=None, b=None, random_state=None, angular_rp_forest=False, target_n_neighbors=-1, target_metric='categorical', target_metric_kwds=None, target_weight=0.5, transform_seed=42, transform_mode='embedding', force_approximation_algorithm=False, verbose=False, unique=False, densmap=False, dens_lambda=2.0, dens_frac=0.3, dens_var_shift=0.1, output_dens=False, disconnection_distance=None*) [source]