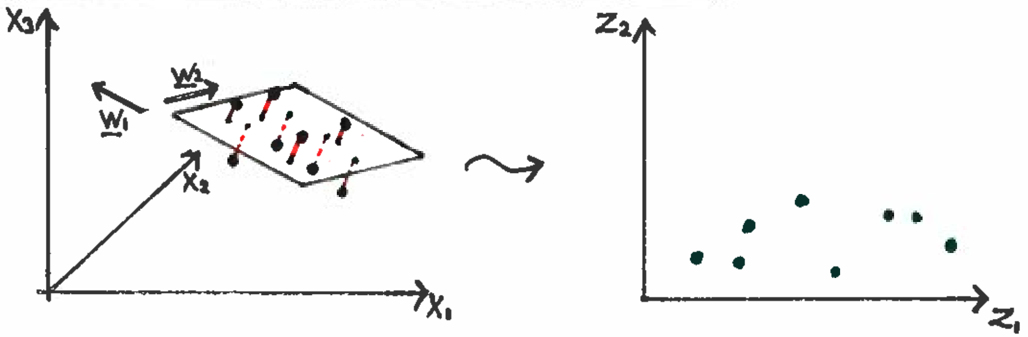
PCA Dimensionality Reduction



Original data space (3D) PC space (2D)

Dimensionality reduction from different representations:

- (Un)centred data:

$$[X = \tilde{X} C_n]; \quad \Sigma = \frac{1}{n} X X^T = U \Lambda U^T; \quad Z = U_k^T X$$

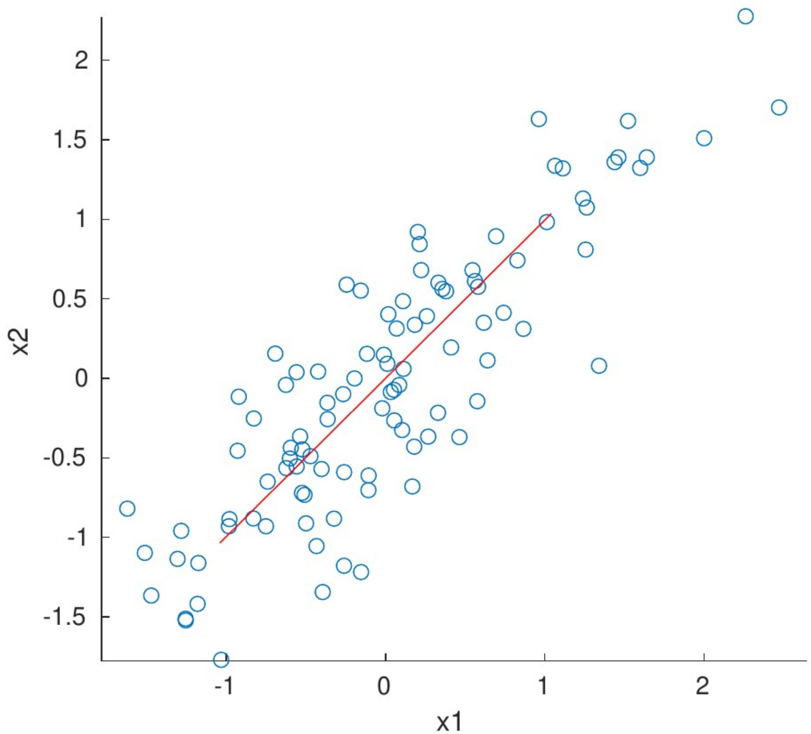
- Inner products between (un)centred data:

$$[G = C_n \tilde{G} C_n]; \quad G = V \tilde{\Lambda} V^T; \quad Z = \sqrt{\tilde{\Lambda}_k}\, V_k^T$$

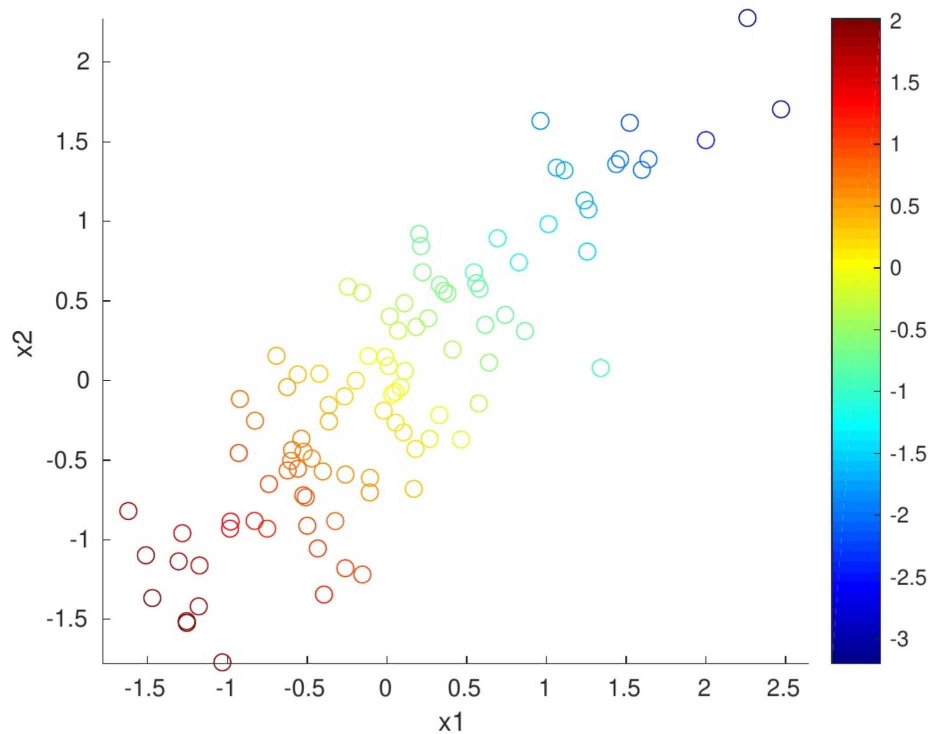
- Squared distances between (un)centred data points:

$$G = -\frac{1}{2} C_n \Delta C_n; \quad G = V \tilde{\Lambda} V^T; \quad Z = \sqrt{\tilde{\Lambda}_k}\, V_k^T$$

Linear Dimensionality Reduction

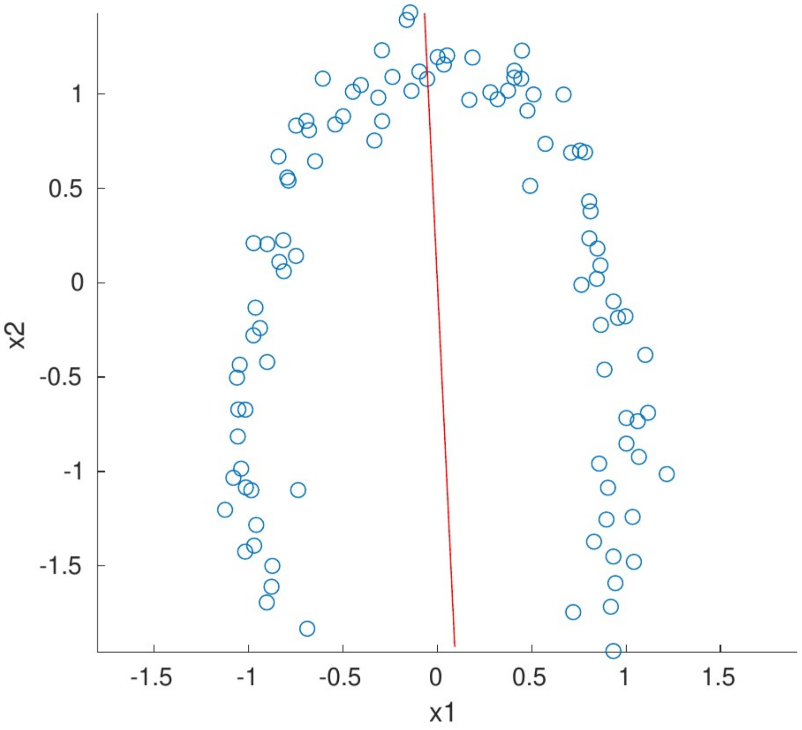


Principal component direction

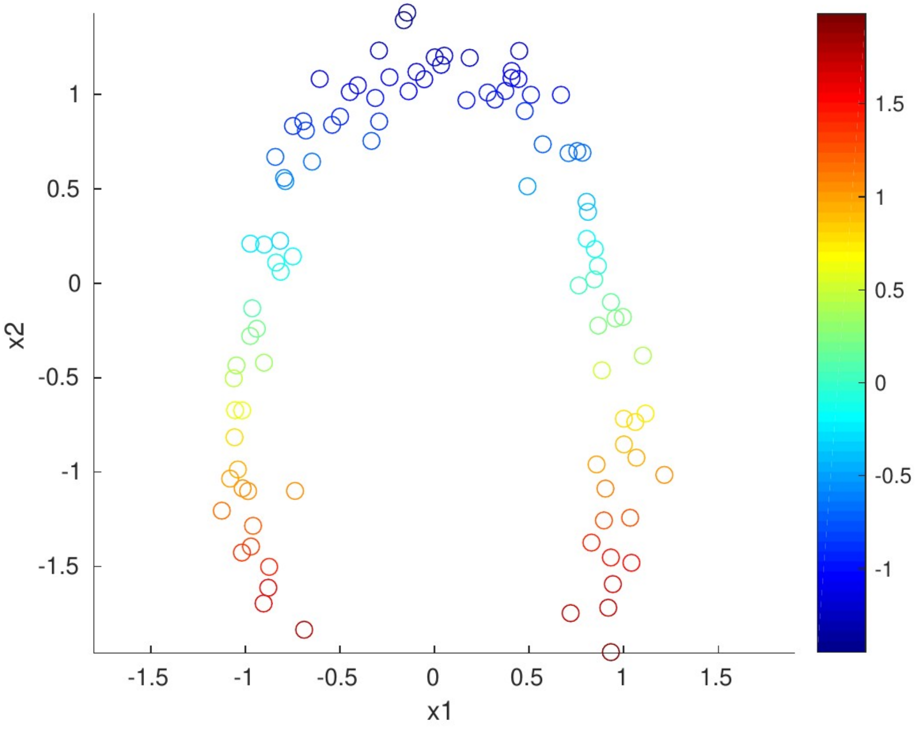


Principal component scores

Linear Dimensionality Reduction (2)



Principal component direction



Principal component scores

Nonlinear PCA

So far: PCA - linear projections to get Z

Now: Transform observations $\underline{x}_i$
to features $\phi(\underline{x}_i)$

Example: $\phi(\underline{x}) = (x_1, \dots, x_d, x_1 x_2, x_1 x_3, \dots x_{d-1} x_d)'$

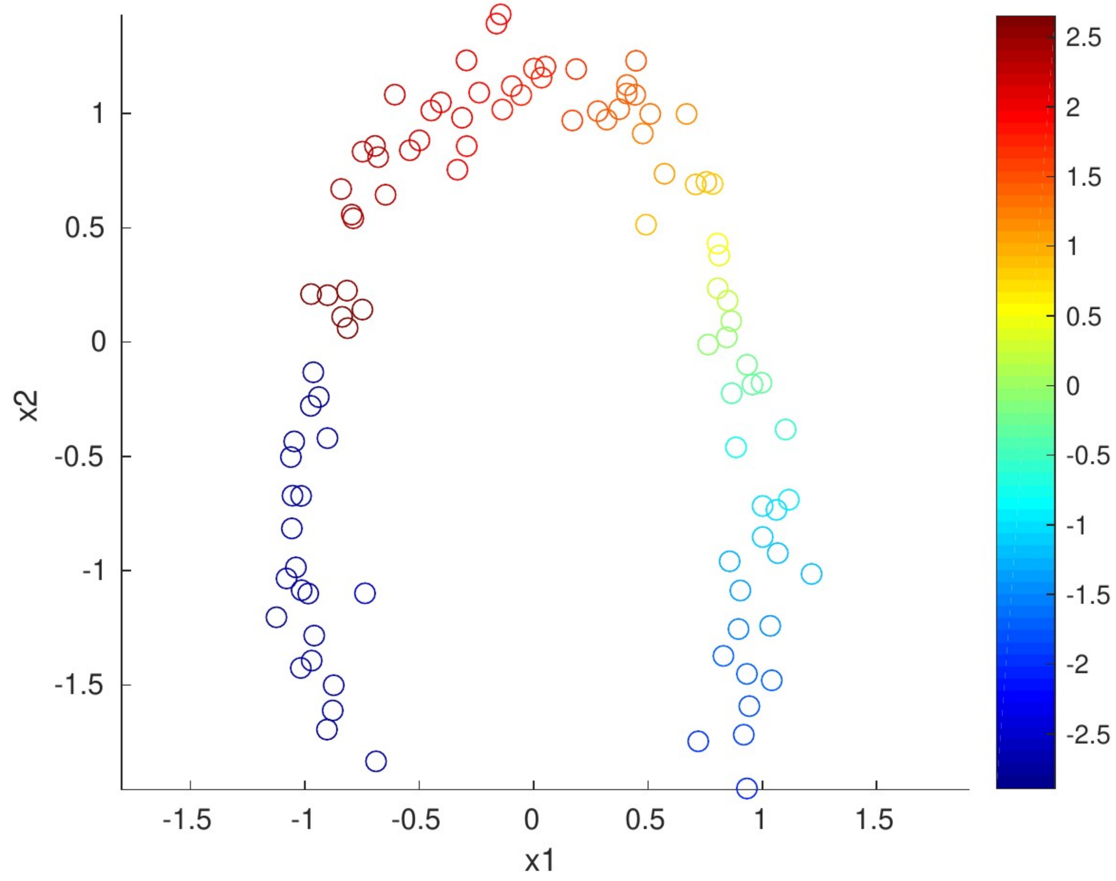
Then: apply PCA to features

$\rightsquigarrow$ k PCs that maximally preserve variance of $\phi(\underline{x}_i)$ rather than $\underline{x}_i$

$\rightsquigarrow$ potentially more information about data <u>preserved</u>

Problem: Choose appropriate nonlinearity ϕ

Non-linear PCA



$$\phi(x_1, x_2) = \left(x_1, \quad x_2, \quad \sqrt{x_1^2 + x_2^2}, \quad \operatorname{atan}(x_2, x_1)\right)^\top$$

Kernel trick:

- For PCA from Gram matrix: sufficient to know inner products
- For some ϕ: $\phi(\underline{x}_i)^T \phi(\underline{x}_j) = k(\underline{x}_i, \underline{x}_j)$
  where $k(\cdot,\cdot)$: "kernel function"
  $\Rightarrow (\tilde{G})_{ij} = \phi(\underline{x}_i)^T \phi(\underline{x}_j) = k(\underline{x}_i, \underline{x}_j)$

Then as before: $G = C_n \tilde{G} C_n$, $Z = \sqrt{\tilde{\Lambda}_k}\, V_k^T$

"kernel PCA"

Kernel function implicitly defines features

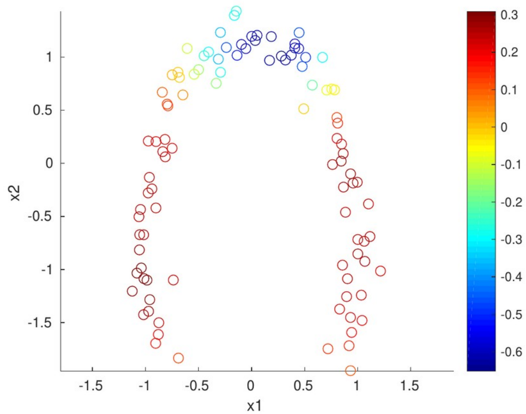
Examples:

$k(\underline{x}, \underline{y}) = (\underline{x}^T \underline{y})^\alpha$ "Polynomial kernel"

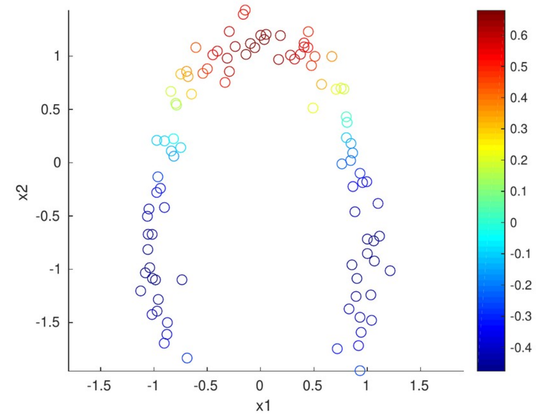
$k(\underline{x}, \underline{y}) = \exp\left(-\frac{\|\underline{x} - \underline{y}\|^2}{2\sigma^2}\right)$ "Gaussian kernel"

Kernel PCA

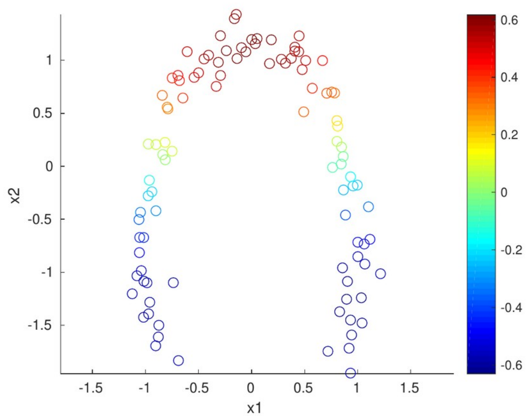
- Gaussian kernel
- σ² determined by quantiles of the distances



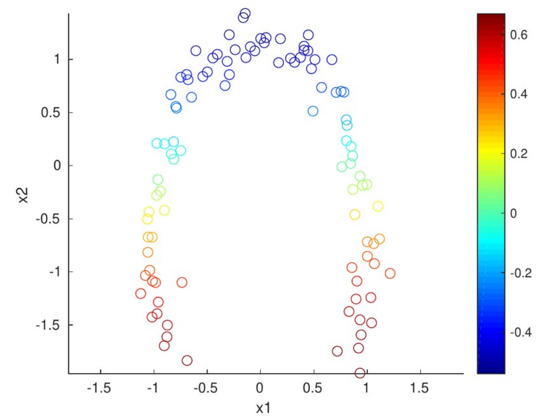
(a) σ^2: 0.01 quantile of all distances



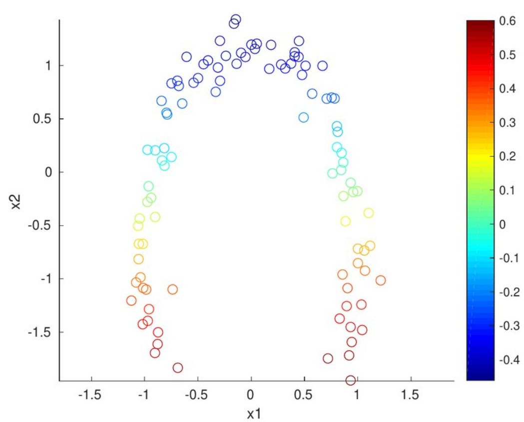
(b) σ^2: 0.05 quantile



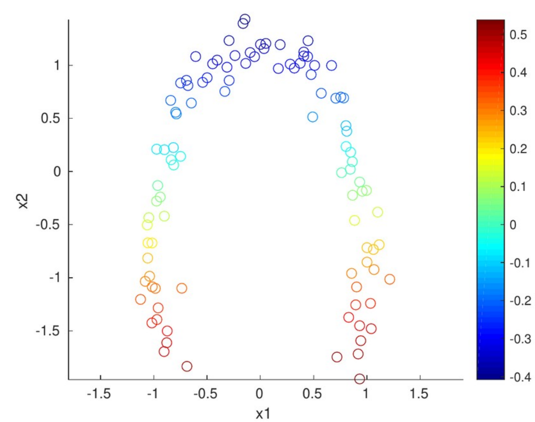
(c) σ^2: 0.1 quantile



(d) σ^2: 0.25 quantile



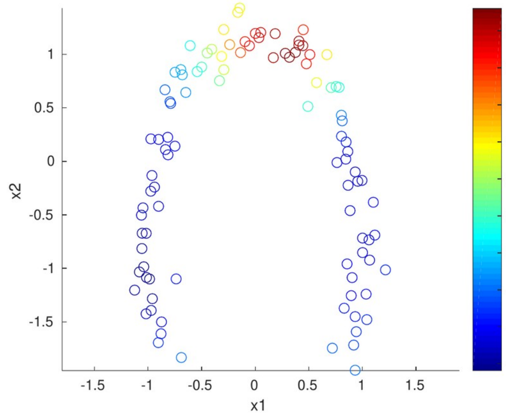
(e) σ^2: 0.5 quantile



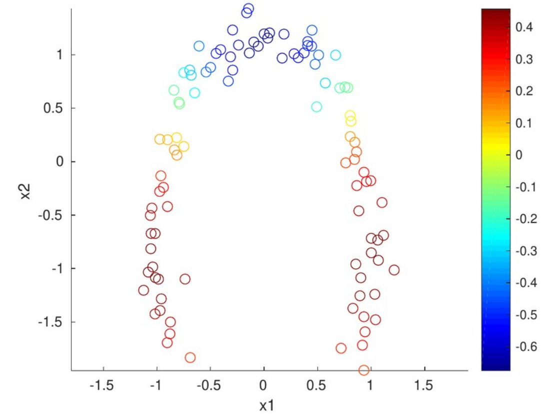
(f) σ^2: 0.75 quantile

Kernel PCA

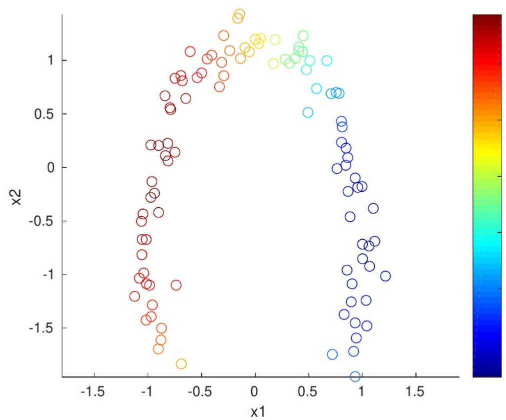
- Gaussian kernel
- σ² determined by quantiles of the distances
- Standardised



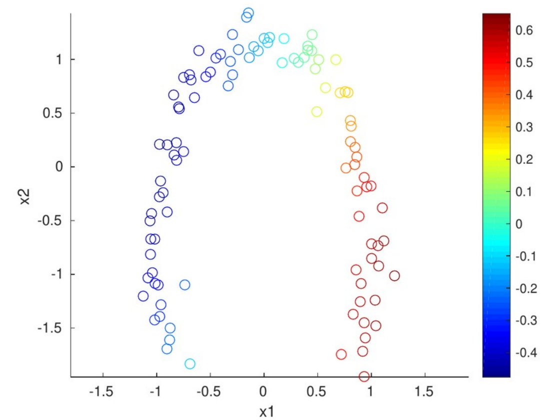
(a) σ^2: 0.01 quantile of all distances



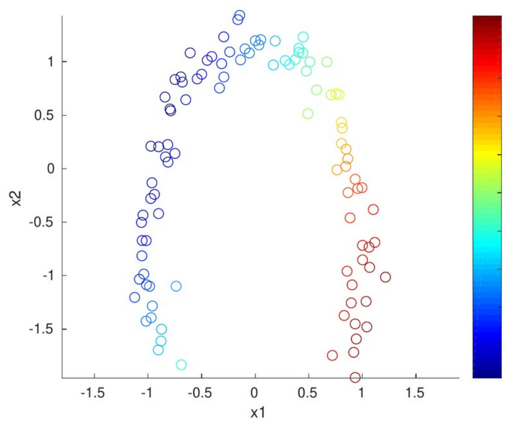
(b) σ^2: 0.05 quantile



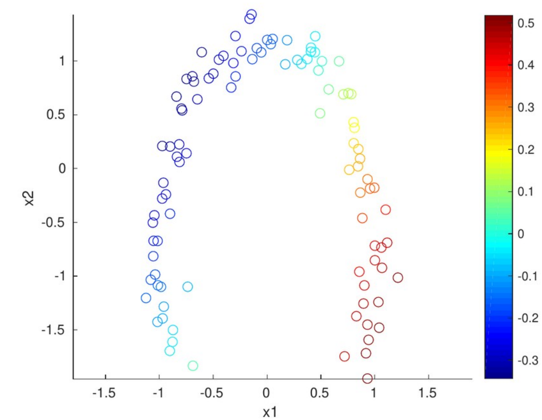
(c) σ^2: 0.1 quantile



(d) σ^2: 0.25 quantile



(e) σ^2: 0.5 quantile



(f) σ^2: 0.75 quantile

Multidimensional scaling (MDS)

Input: dissimilarities Δ

Assumption: there are latent data points with relation to Δ

Find configuration of points such that their distances represent Δ

- Metric MDS

Precise numerical values taken into account

Find n points: $\underline{z}_1, \dots \underline{z}_n \in \mathbb{R}^k$

$$\boxed{\begin{array}{l} \underset{\underline{z}_1, \dots, \underline{z}_n}{\text{minimise}} \sum_{i<j} w_{ij} \left(\| \underline{z}_i - \underline{z}_j \| - \delta_{ij} \right)^2 \\ \text{where } \| \underline{z}_i - \underline{z}_j \| = \sqrt{(\underline{z}_i - \underline{z}_j)^T (\underline{z}_i - \underline{z}_j)} \\ \text{and } w_{ij} \geq 0 \text{ given weights} \end{array}}$$

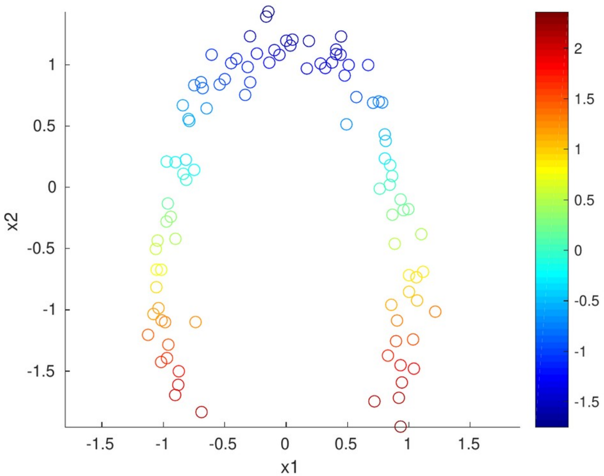
Solve using gradient descent

Special case:

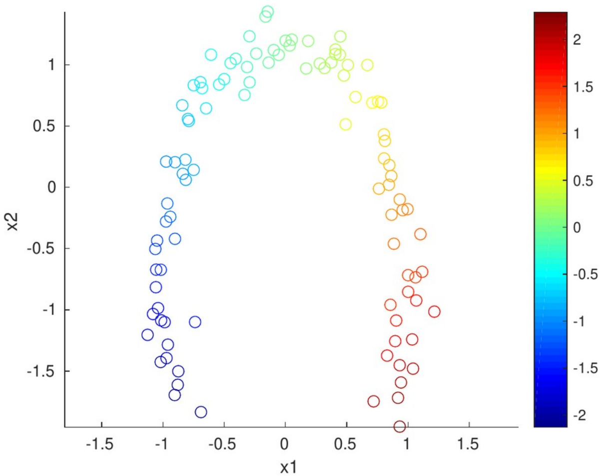
$w_{ij} = 1/\delta_{ij}$ "Sammon nonlinear mapping"

emphasises small dissimilarities

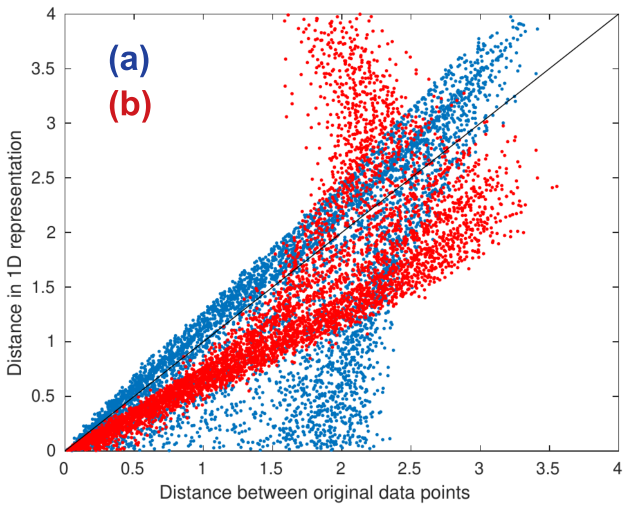
Sammon Nonlinear Mapping



(a) Local optimum



(b) Better local optimum



– Nonmetric MDS

Precise numerical values are disregarded (rank order used)

$$\boxed{\min_{\underline{z}_1,\dots,\underline{z}_n, f} \sum_{i<j} w_{ij}\left(\|\underline{z}_i - \underline{z}_j\| - f(\delta_{ij})\right)^2 \quad \text{where } f: \text{non-decreasing function}}$$

"f converting dissimilarities to distances"

Typically alternating optimisation of $\underline{z}_i$, f

See Izenman 2008,

"Modern multivariate statistical techniques"