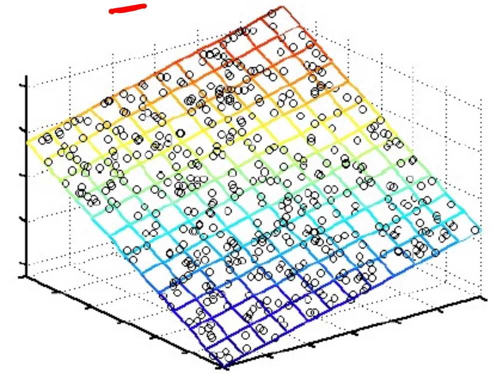
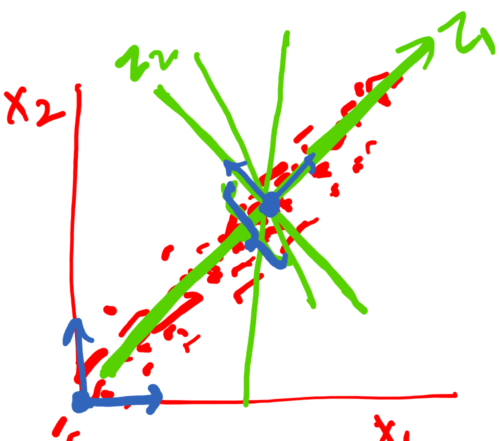
Unsupervised Dimensionality Reduction

$z_i = \sum_{j=1}^{D} w_j x_j$

$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix} \rightarrow \begin{bmatrix} z_1 \\ \vdots \\ z_d \end{bmatrix} \quad d < D$

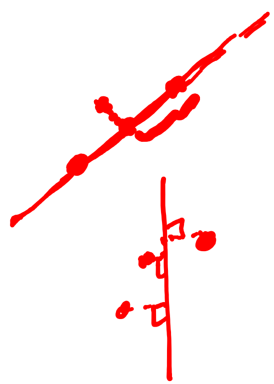
- Linear
  - **Principal Component Analysis (PCA)**
  - Factor Analysis
  - Independent Component Analysis (ICA)

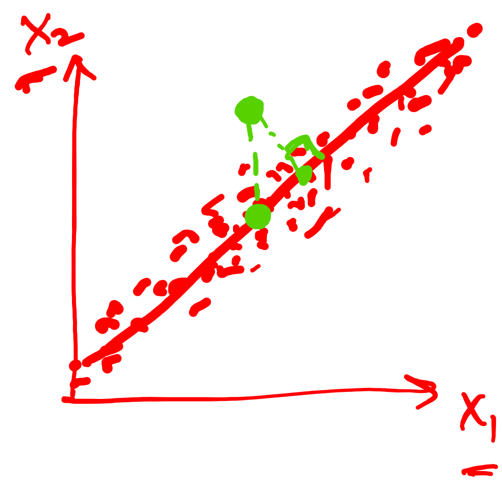




$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rightarrow [z_1]$

max var of projected points

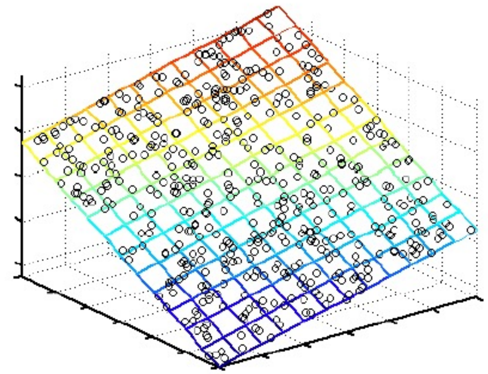


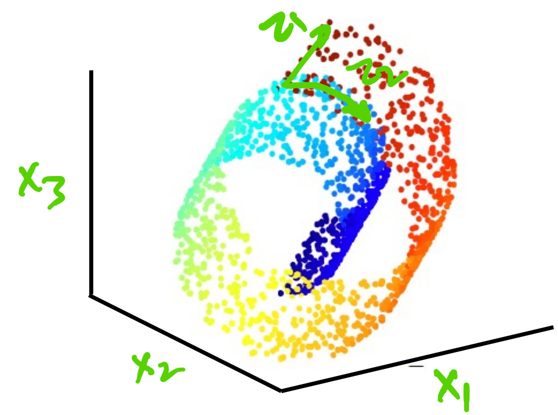


min MSE reconstruction from projection

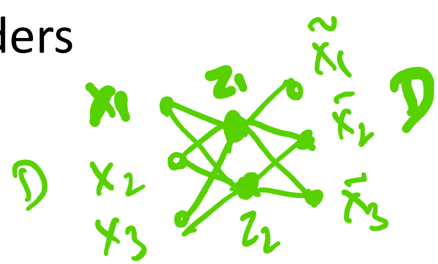
Unsupervised Dimensionality Reduction

- Linear
  - **Principal Component Analysis (PCA)**
  - Factor Analysis
  - Independent Component Analysis (ICA)





- Nonlinear
  - **Kernel PCA**
  - Laplacian Eigenmaps, ISOMAP, LLE
  - Autoencoders



$$\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} \to \begin{bmatrix}z_1\\z_2\end{bmatrix}$$

$z_i = \sum_j w_j \phi(x_j)$

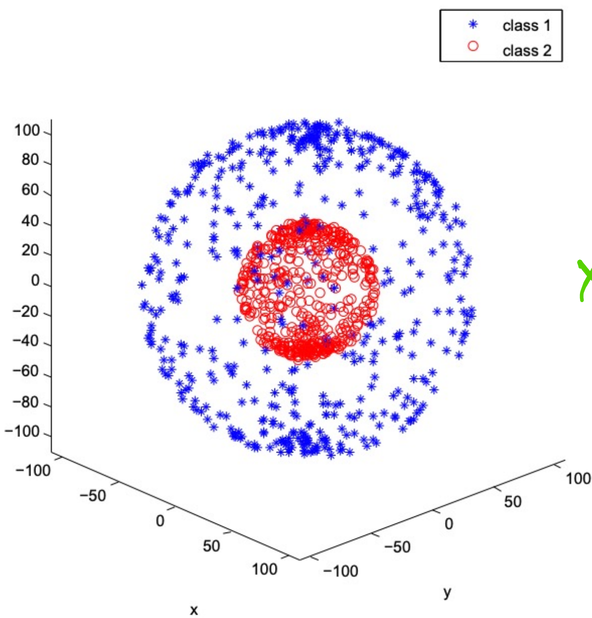
Kernel PCA

$\boxed{X X^T}$ $D \times D$ → $\phi(x)\phi(x)^T$ $\tilde{D} \times \tilde{D}$

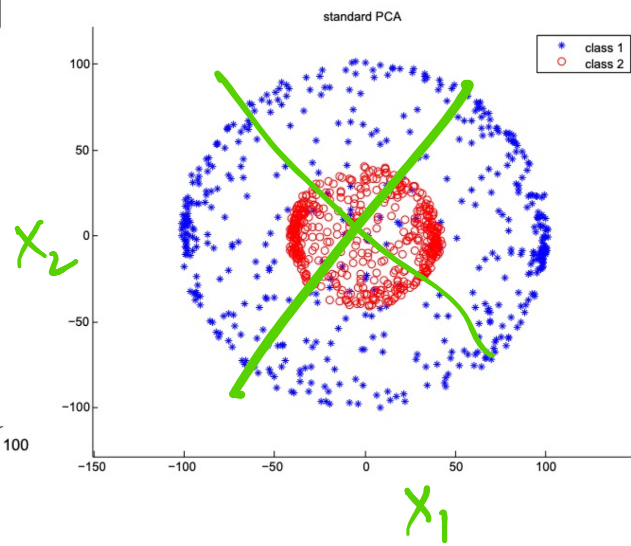
HW4!

Latent features: linear in $\phi(x)$ where $\phi(x).\ \phi(x') = K(x,x')$ that capture maximum variance or minimum reconstruction error

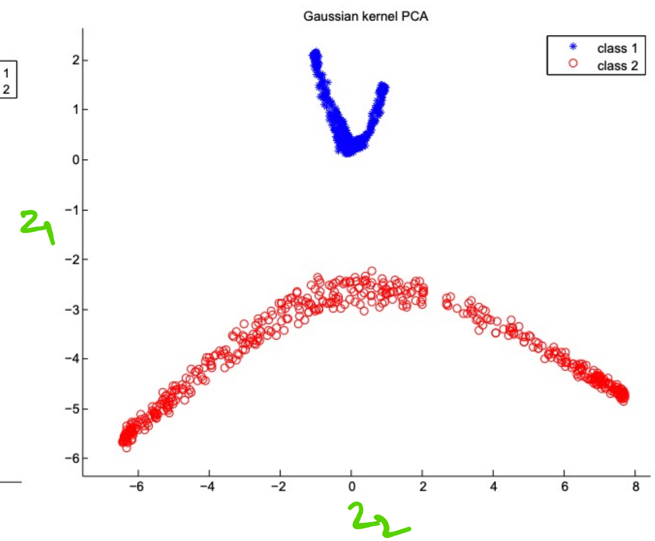
$(K) = \phi(x)\cdot\phi(x')$ $n \times n$



Original data points



PCA



Kernel PCA
Gaussian/RBF kernel

Src: ArXiv 1207.3538

Kernel PCA

HW4!

Latent features: linear in $\phi(x)$ where $\phi(x).\ \phi(x') = K(x,x')$ that capture maximum variance or minimum reconstruction error

PCA:

$d < D$

$v_1 \ldots v_d$

$D \times D$

Top d eigenvectors (each D dimensional) of sample covariance XX^T

$\phi(X)\phi(X)^T$

Low d-dimensional embedding of a point: $[v_1^T x_i,\ v_2^T x_i,\ \ldots\ v_d^T x_i]$

$= z_1 \quad = z_2 \quad = z_d$

Kernel PCA:

$v_1 \ldots v_d$

Top d eigenvectors (each n dimensional) of kernel matrix $K(X,X)$ $n \times n$

Low d-dimensional embedding of a point: $[v_1(i),\ v_2(i),\ \ldots,\ v_d(i)]$

$= z_1 \quad = z_2 \quad = z_d$

Eigenvectors are not PCs but projections of data points

PCA Summary

- PCA finds latent features linear in original features x that capture
  - Maximum variance amongst all linear features
  - Minimum reconstruction error when recovering points from PC projections

- Non-convex problem with simple solution:

  PCs = eigenvectors of sample covariance matrix

  Lower (d < D) dimensional embedding of data point = projection of data point onto d PCs

- Kernel PCA: latent features linear in $\phi(x)$ where $\phi(x).\ \phi(x') = K(x,x')$ that capture maximum variance or minimum reconstruction error
  - Directly get projections of data points