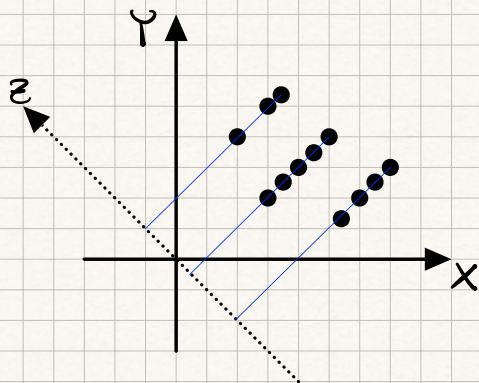


DIMENSIONALITY REDUCTION

In alcune circostanze è meglio costruire un nuovo sottospazio piuttosto che usare sottospazi del DB originale.

Esempio



Nella nuova dimensione Z i cluster non si sovrappongono come fanno nelle dimensioni X e Y .

SPECTRAL CLUSTERING

THE NG-JORDAN-WEISS ALGORITHM (NJW)

Dati gli oggetti $o_1 \dots o_n$ in D , e la loro distanza $\text{dist}(o_i, o_j)$, trovare i K cluster desiderati.

Calcolare poi una matrice detta **matrice di affinità** W nel seguente modo:

$$w_{ij} = e^{\left(-\frac{\text{dist}(o_i, o_j)}{\sigma} \right)}$$

Dove con il parametro σ si cerca di controllare la velocità con cui l'affinità di due oggetti diminuisce all'aumentare della loro distanza. In NJW $w_{ii} = 0$.

Definiamo poi una seconda matrice D , matrice diagonale dove:

$$D_{ii} = \sum_{j=1}^n w_{ij}$$

Poi abbiamo una terza matrice $A = f(W)$ che si costruisce come:

$$A = D^{-\frac{1}{2}} \cdot W \cdot D^{-\frac{1}{2}}$$

Trovare poi gli **autovettori** di A tale per cui, da definizione di autovettori v_i e autovalori λ_i vale che $A\vec{v} = \lambda\vec{v}$.

Usare i Top-K autovettori v per definire un nuovo sottospazio su cui proiettare i dati di D .

Poi eseguire un algoritmo classico di clustering nel nuovo sottospazio.