Prev topic Next topic Prev page
Next page

The Lecture Contains:

- General idea
- Distortion
- Reduction to one dimension

Prev topic Next topic Prev page
Next page

General idea

- Feature vector v in k dimensions
- Transformation f on to yield another feature vector $f(v)$ of dimensionality $k' < k$
- Distance measure may change from d to d'

Prev topic Next topic Prev page
Next page

General idea

- Feature vector v in k dimensions
- Transformation f on to yield another feature vector $f(v)$ of dimensionality $k' < k$
- Distance measure may change from d to d'
- Aim is to maintain, as far as possible

$$d(x, y) \approx d'(f(x), f(y))$$

Prev topic Next topic Prev page
Next page

General idea

- Feature vector v in k dimensions
- Transformation f on to yield another feature vector $f(v)$ of dimensionality $k' < k$
- Distance measure may change from d to d'
- Aim is to maintain, as far as possible

$$d(x,y) \approx d'(f(x), f(y))$$

- Pruning property

$$\forall x, y,\ d'(f(x), f(y)) \leq d(x,y)$$

- Mappings that obey pruning property are contractive
- Useful for range queries

Module 6: Dimensionality Reduction

Lecture 28: Dimensionality Reduction

General idea

- Feature vector v in k dimensions
- Transformation f on to yield another feature vector $f(v)$ of dimensionality $k' < k$
- Distance measure may change from d to d'
- Aim is to maintain, as far as possible

$$d(x, y) \approx d'(f(x), f(y))$$

- Pruning property

$$\forall x, y,\ d'(f(x), f(y)) \leq d(x, y)$$

- Mappings that obey pruning property are contractive
- Useful for range queries
- Proximity-preserving property

$$\forall x, y, z,\ d(x, y) \leq d(x, z) \implies d'(f(x), f(y)) \leq d'(f(x), f(z))$$

- Useful for nearest-neighbor queries

Prev topic Next topic Prev page
Next page

Distortion

$$\frac{1}{c_1}.d(x,y) \leq d'(f(x),f(y)) \leq c_2.d(x,y)$$

- $c_1, c_2 \geq 1$ provides bounds on distance after dimensionality reduction

- c_1 is contraction and c_2 is expansion

Distortion

$$\frac{1}{c_1}.d(x,y) \leq d'(f(x),f(y)) \leq c_2.d(x,y)$$

- $c_1, c_2 \geq 1$ provides bounds on distance after dimensionality reduction
- c_1 is contraction and c_2 is expansion

$$\text{Distortion } = c_1.c_2$$

- For arbitrary dimensionality reductions,

$$\text{Distortion} = \max_{\forall x,y}\left\{\frac{d(x,y)}{d'(f(x),f(y))}\right\} \times \max_{\forall x,y}\left\{\frac{d'(f(x),f(y))}{d(x,y)}\right\}$$

- Contractive embeddings: $c_2 = 1$
- Isometric embeddings: $c_1 = c_2 = 1$

Prev topic Next topic Prev page
Next page

Reduction to one dimension

- Hashing
- Space-filling curves
- Choosing single dimension
    - Choose dimension with largest spread or largest variance
- Pruning and proximity-preserving properties are generally violated
- Loss of information and distortion are too high
- In general, not very useful