Feature…

- …a property of interest that can help us index an object
- For a "student record"
  - student_ID

  can be a feature
- What are the features for an image?

Image features

- There are many possible features
  - Color histogram
  - Texture
  - Edges
  - Shapes
  - Objects
  - Object or scene semantics
- Feature selection: which one to use for indexing?

Good feature..

- A good feature is significant and enables us to differentiate objects from others as much as possible

- A good feature corresponds to users' perception as much as possible
  - Relevance feedback!!!!

What does “significant” mean

- Information theoric sense:
  - An event is more significant if it carries more information

What does “significant” mean

- Information theoric sense:
  - An event is more significant if it carries more information
  - An event that has high occurrence rate carries less information
    - Solar eclipse is more interesting then sunset

High frequency ----- less information
Low frequency ----- high information

Entropy

- Total information content (uncertainty)

$$H(X) \equiv \sum_{x \in \mathcal{A}_X} P(x) \log \frac{1}{P(x)}$$

Entropy (example)

- Total information content (uncertainty)

$$H(X) \equiv \sum_{x \in \mathcal{A}_X} P(x) \log \frac{1}{P(x)}$$

P(a) = 0.5, P(b) = 0.5 → H > 0 — more uncertain

P(a) = 1.0, P(b) = 0.0 → H = 0 — less uncertain

Maria Luisa Sapino (BDM 2018)

Entropy (example)

- Total information content (uncertainty)

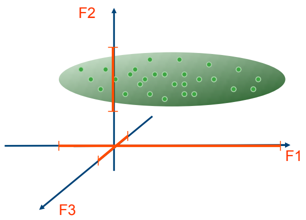
$$H(X) \equiv \sum_{x \in \mathcal{A}_X} P(x) \log \frac{1}{P(x)}$$

P(a) = 0.5, P(b) = 0.5 → H > 0 — more uncertain, more information

P(a) = 1.0, P(b) = 0.0 → H = 0 — less uncertain, less information

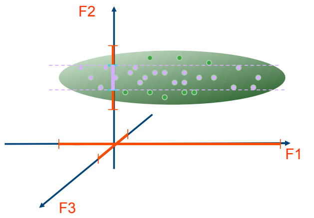
Maria Luisa Sapino (BDM 2018)

Which feature is better?



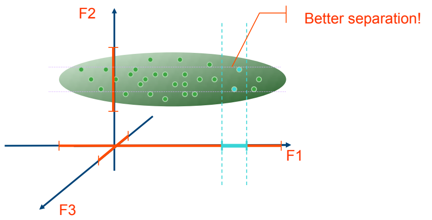
Maria Luisa Sapino (BDM 2018)

Which feature is better?



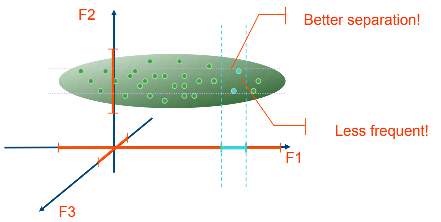
Maria Luisa Sapino (BDM 2018)

Which feature is better?



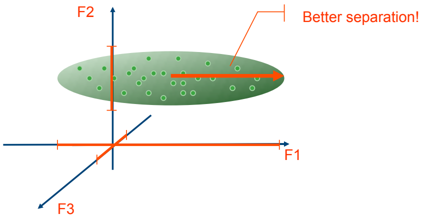
Maria Luisa Sapino (BDM 2018)

Which feature is better?



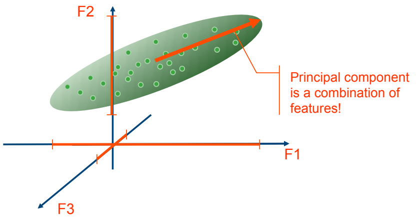
Maria Luisa Sapino (BDM 2018)

Principal component



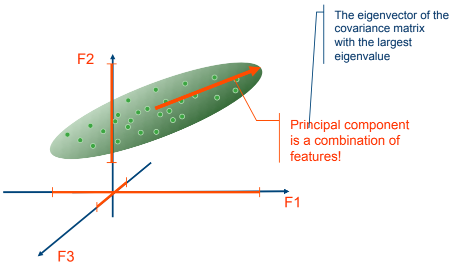
Maria Luisa Sapino (BDM 2018)

Principal component analysis



Maria Luisa Sapino (BDM 2018)

Principal component analysis

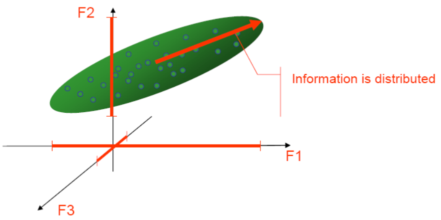


Maria Luisa Sapino (BDM 2018)

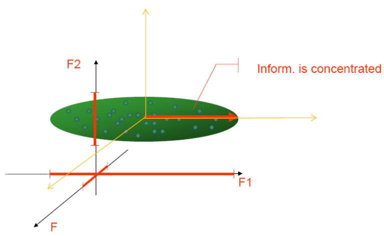
Principle Component Analysis

- ..also known as Karhunen-Loeve Transfrom
  - ..a linear transform that optimally decorrelates the input.

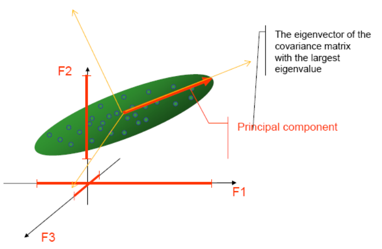
Maria Luisa Sapino (BDM 2018)



Maria Luisa Sapino (BDM 2018)

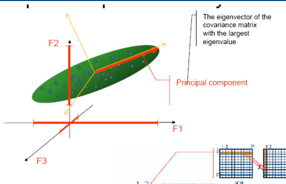


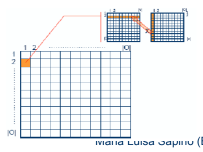
Maria Luisa Sapino (BDM 2018)



Linearly Independent Eigenvectors

- Suppose that A is and n × n square matrix. If the eigenvalues, c1....ck are distinct, then eigenvectors v1,...vk are a set of k linearly independent vectors.





The covariance between objects a and b is

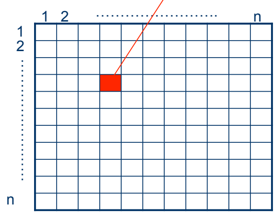
$cov(a,b) = (|V|)^{-1}\Sigma_i(a_i\text{-}avg\_a)(b_i\text{-}avg\_b)$,

where |V| is the # of features

$$A[i,j] = Cov(i,j) = E\left((Fi-\mu_i)(Fj-\mu_j)\right)$$

Eigen decomposition

A =



Covariance matrix

Maria Luisa Sapino (BDM 2018)

…reminder

- Eigenvalue and eigenvector
- Given a matrix A, let *c* (scalar) and *x* (vector) be such that

$$\underbrace{c}_{\text{Eigenvalue}}\,\vec{x} = A\,\underbrace{\vec{x}}_{\text{Eigenvector}}$$

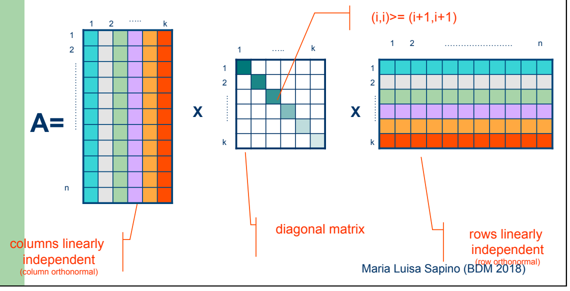
Maria Luisa Sapino (BDM 2018)

Properties of Eigenvectors

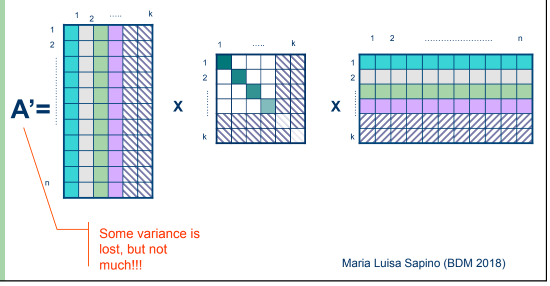
- Suppose that A is and n × n square matrix
  - if the eigenvalues, $c_1 \ldots c_k$, are distinct, then eigenvectors $v_1, \ldots v_k$ are a set of k linearly independent vectors.
    - thus they can be used as the basis of the space!!!
  - The value of c_i describes the contribution of v_i in A. Thus
    - if we pick an A that describe the variation of data,
    - c_i will describe the directions along which variation is high

Maria Luisa Sapino (BDM 2018)

eigen decomposition



eigen decomposition



How many eigenvectors shall we maintain?

- **Mean eigenvalue**: use only the dimensions whose eigenvalues are greater than or equal to the mean eigenvalue.
- **Kaiser rule:** keep only those eigenvectors whose eigenvalues are greater than 1.
- **Parallel analysis:**
  - analyze a random covariance matrix,
  - Plots cumulative eigenvalues for both random and intended matrices;
  - Find where the two curves intersect.
- **Scree test:** plot the successive eigenvalues to find a point where the plot levels off.
- **Variance explained:** keep enough dimensions to account for 95% of the initial variance

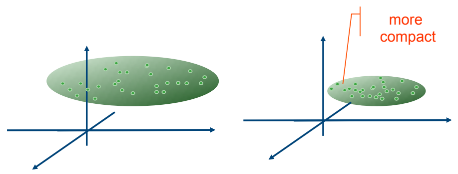
Compactness of a database

$$comp(D) = \sum_{i \neq j} similarity(o_i, o_j)$$

Maria Luisa Sapino (BDM 2018)

Compactness of a database

$$comp(D) = \sum_{i \neq j} similarity(o_i, o_j)$$

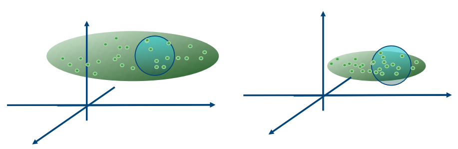


Maria Luisa Sapino (BDM 2018)

Compactness of a database

$$comp(D) = \sum_{i \neq j} similarity(o_i, o_j)$$

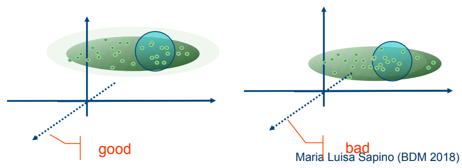
A compact database is not desirable!!!



Maria Luisa Sapino (BDM 2018)

Feature quality

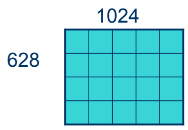
A feature is
- good if we remove it, the overall compactness increases
- bad if we remove it, the overall compactness decreases



Maria Luisa Sapino (BDM 2018)

Problem…

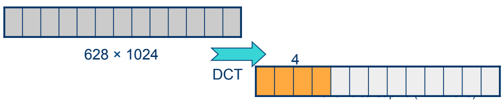
- Feature vector size: 628 × 1024
  - Dimensionality curse: high dimensions make indices unusable (10-15 dimensions max!!!)



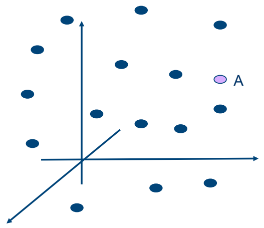
Maria Luisa Sapino (BDM 2018)

Problem…

- Feature vector size: 628 × 1024
  - Dimensionality curse: high dimensions make indices unusable (10-15 dimensions max!!!)
- Solution: Reduce # dimensions of the vector
  - use distance-preserving transforms
  - Ex: fourier trans., DCT,wavelet trans.

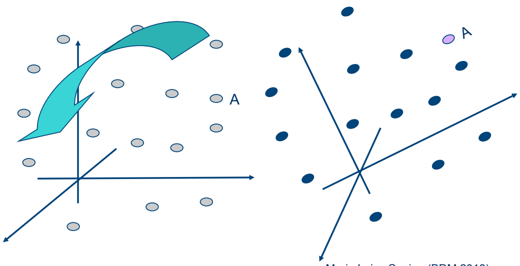


Transforms



Maria Luisa Sapino (BDM 2018)

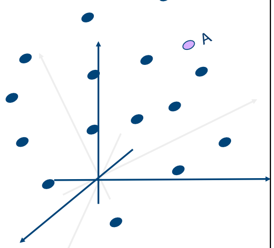
Transforms



Maria Luisa Sapino (BDM 2018)

Transforms

Distances and angles are preserved

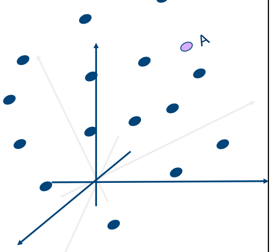


Maria Luisa Sapino (BDM 2018)

Transforms

Distances and angles are preserved

Some dimensions are more important (differentiating) than the other

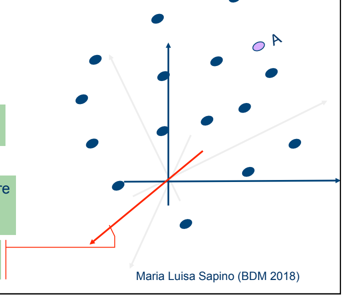


Transforms

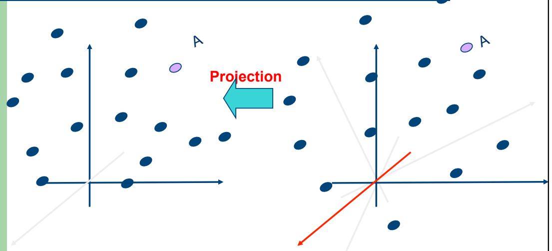
Distances and angles are preserved

Some dimensions are more important (differentiating) than the other

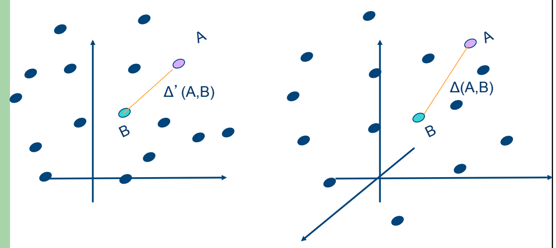
Eliminate unimportant dimensions



Transform + Projection (Compression or Feature selection)



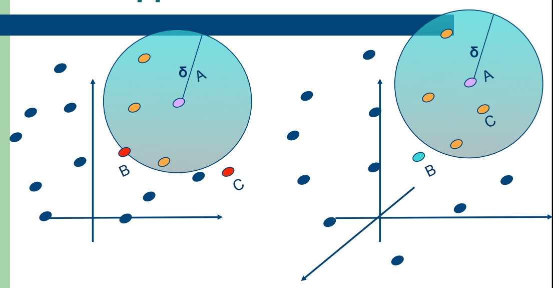
What happens to distances???



Δ(A,B) <> Δ' (A,B)

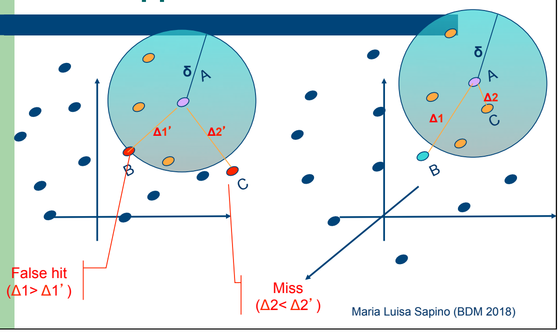
Maria Luisa Sapino (BDM 2018)

What happens to distances???



Maria Luisa Sapino (BDM 2018)

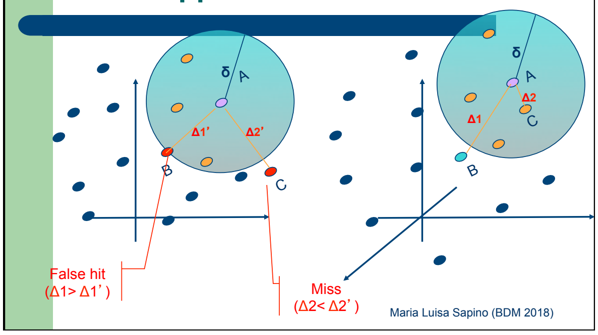
What happens to distances???



Maria Luisa Sapino (BDM 2018)

Misses are not desirable!
Can not be eliminated with postprocessing

What happens to distances???



Maria Luisa Sapino (BDM 2018)