Unsupervised Learning and Dimensionality Reduction

Data Science Decal

Hosted by Machine Learning at Berkeley

Unsupervised Learning and Dimensionality Reduction

Background

Algorithms

Questions

Background

Curse of Dimensionality

- Coined by Bellman in 1961. Problems of geometric increase of computational cost from increasing dimensionality (number of variables).

Algorithms

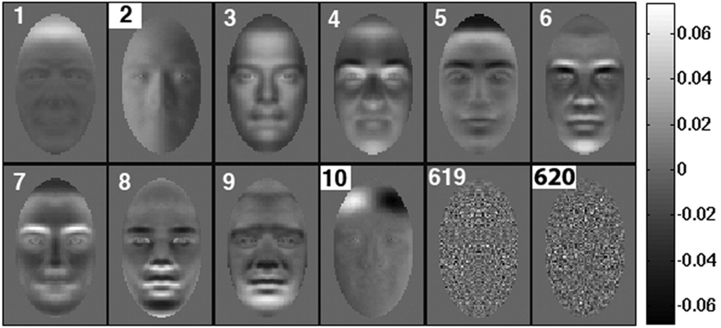
Dimensionality Reduction: Principal Components Analysis (PCA)

ML@B

- Goal: Reduce dimension while preserving randomness from HD space.
- More specifically: Pick orthogonal vectors to represent the maximum variation in the data.
- The most useful directions explain the most variation.

When to use PCA

- feature selection

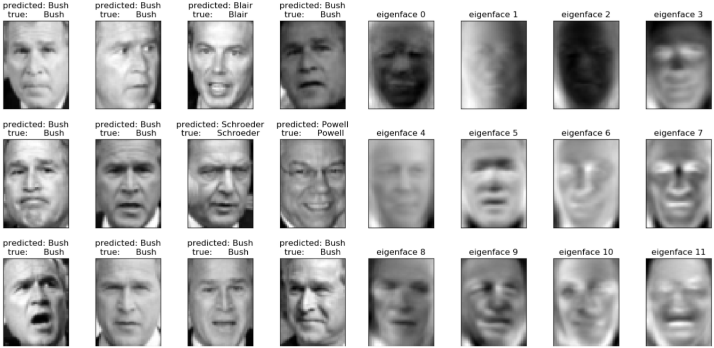


When to use PCA

- classification

Expected results for the top 5 most represented people in the dataset:

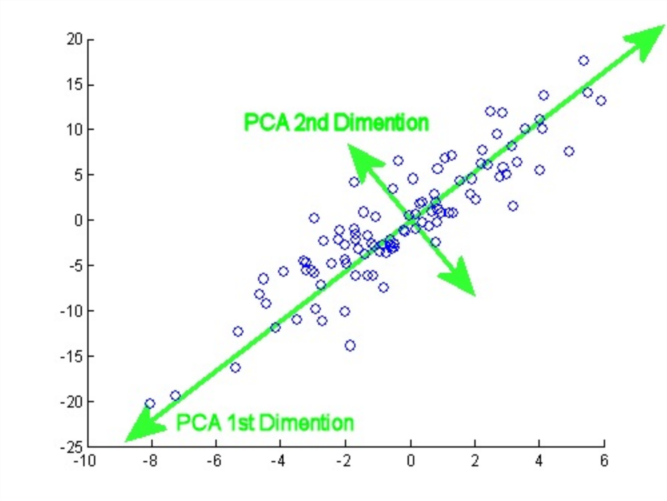
Ariel Sharon	0.67	0.92	0.77	13
Colin Powell	0.75	0.78	0.76	60
Donald Rumsfeld	0.78	0.67	0.72	27
George W Bush	0.86	0.86	0.86	146
Gerhard Schroeder	0.76	0.76	0.76	25
Hugo Chavez	0.67	0.67	0.67	15
Tony Blair	0.81	0.69	0.75	36
avg / total	0.80	0.80	0.80	322



Dimensionality Reduction: Principal Components Analysis (PCA)

- Data matrix $X \in \mathbf{R}^{mxn}$. assume X is centered.
- Covariance matrix $\sum$ is $X^T X \in \mathbf{R}^{nxn}$ and ≥ 0
- Find a set of $d \leqslant n$ linearly independent vectors (principal components) that best explain the variation. Alternatively, minimize the sum-squared error with X.
- The principal components are eigenvector θ_i and corresponding eigenvalue λ_i for $i \in 1, ..., d$.

Dimensionality Reduction: Principal Components Analysis (PCA)



Dimensionality Reduction: Principal Components Analysis (PCA)

- Demo: http://setosa.io/ev/principal-component-analysis/
- The eigenvector with the largest eigenvalue is the direction along which the data set has the maximum variance. Intuitively, this is because the first eigenvector has to lie along where points are scattered the most.
- Adding more principal components explains more variation.
- sklearn.decomposition.PCA

Dimensionality Reduction: Linear Discriminant Analysis (LDA)

ML@B

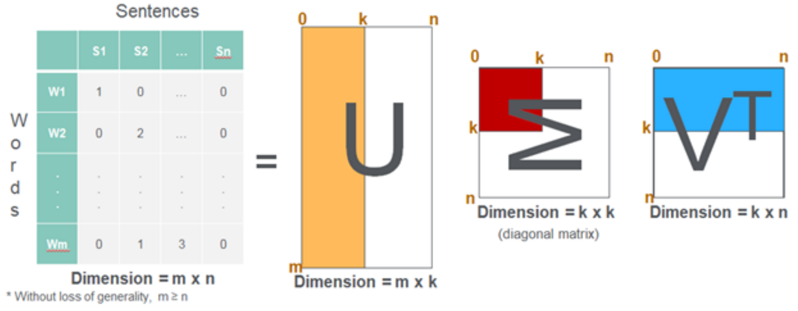
- Pick the first principal component and its eigenvalue (θ_1, λ_1).
- Project X down to one dimension. $X' = X\theta_1$. Predict: x_i is a positive example if $x_i' > 0$; x_i is a negative example if $x_i' \leqslant 0$.
- Revisiting the Demo: http://setosa.io/ev/principal-component-analysis/

- A classification algorithm.
- Setup:
- Data matrix $X \in \mathbf{R}^{mxn}$. assume X is centered.
- Covariance matrix $\sum$ is $X^T X \in \mathbf{R}^{nxn}$ and $\succeq 0$
- The principal components are eigenvector θ_i and corresponding eigenvalue λ_i for $i \in 1, ..., d$.

Unsupervised Learning: Singular Value Decomposition (SVD)

- Goal: Low rank approximation of any matrix.
- Uses:
  - Left singular vectors $\equiv$ pc components.
  - Remove redundant data.
  - Image compression.

Unsupervised Learning: Singular Value Decomposition (SVD)

- 

Unsupervised Learning: Singular Value Decomposition (SVD)

- Any matrix $A \in \mathbb{R}^{m,n}$ can be expressed as

$$A = \sum_{i=1}^{r} \sigma_i u_i {v_i}^T = U_r \sum {V_r}^T$$

  where r = rank A and the singular vectors in matrices U, V are orthonormal.

- σ_i = singular value
- σ^2 = eigenvalue
- u_i = left singular vector and eigenvector of $A^T A$
- v_i = right singular vector and eigenvector of AA^T

Unsupervised Learning: Singular Value Decomposition (SVD)

- Finding SVD by
  - hand: `https://www.scss.tcd.ie/~dahyotr/CS1BA1/SolutionEigen.pdf`
  - sklearn: sklearn.decomposition.TruncatedSVD
- Clustering demo: `http://www.cs.cornell.edu/courses/cs4300/2016sp/Demos/demo20_2.html`

Questions

Questions?