ML Week

Face and Handwriting Recognition

Jeff Abrahamson

20–22 juillet 2016

PCA

Principle component analysis

Analyse en composantes principales

Motivation

Remember the Curse of Dimensionality?

Principle

- Linear transformations have axes
- Find them (eigenvectors of the covariance matrix)
- Pick the biggest ones

Principle

- Linear transformations have axes
- Find them (eigenvectors of the covariance matrix)
- Pick the biggest ones

Fitting an *n*-dimensional ellipsoid to the data

Uses

- Exploratory data analysis
- Compression

Also known as

- Discrete Kosambi-Karhunen–Loève transform (KLT) (signal processing)
- Hotelling transform (multivariate quality control)
- Proper orthogonal decomposition (POD) (ME)
- Singular value decomposition (SVD), Eigenvalue decomposition (EVD) (linear algebra)
- Etc.

History

- Invented by Karl Pearson in 1901
- Invented (again) and named by Harold Hotelling in 1930's
- Also known as…

Also known as

- It's a long list, every field uses a different name...

Face Recognition

Eigenfaces

- Sirovich and Kirby (1987)
- Turk and Pentland (1991)

Turk, Matthew A and Pentland, Alex P. Face recognition using eigenfaces. Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on 1991.

Eigenfaces

Want: a low-dimensional representation of a face

Plan: cluster simplified faces

Eigenfaces

Viewed as compression:

- Use PCA on face images to form a set of basis features
- Use eigenpictures to reconstruct original faces

Eigenfaces

Eigenfaces algorithm

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a random vector with observations $x_i \in \mathbb{R}^d$.

Compute

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

OpenCV

Eigenfaces algorithm

Compute the covariance matrix S:

$$S = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T$$

Eigenfaces algorithm

Compute the eigenvectors of S:

$$Sv_i = \lambda_i v_i \qquad i = 1, 2, \ldots, n$$

Sort the eigenvectors in decreasing order.

We want the k principal components, so take the first k.

Eigenfaces algorithm

Compute the eigenvectors of S:

$$Sv_i = \lambda_i v_i \qquad i = 1, 2, \ldots, n$$

Sort the eigenvectors in decreasing order.

We want the k principal components, so take the first k.

This is PCA.

Eigenfaces algorithm

The k principal components of the observed vector x are then given by

$$y = W^T(x - \mu)$$

where

$$W = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_k \\ | & | & & | \end{bmatrix}$$

Eigenfaces algorithm

The reconstruction from the PCA basis is then

$$x = Wy + \mu$$

Eigenfaces algorithm

So the plan is this:

- Project all training samples in the PCA subspace
- Project the query into the PCA subspace
- Find the nearest neighbour to the projected query image among the projected training images

Eigenfaces algorithm

Eigenfaces algorithm

Some advantages:

- Easy, relatively inexpensive
- Recognition cheaper than preprocessing
- Reasonably large database possible

Eigenfaces algorithm

Some problems:

- Need controlled environment
- Needs straight-on view
- Sensitive to expression changes
- If lots of variance is external (e.g., lighting)…

Handwriting Recognition

Choices

- Online
- Offlne

Introduction to Handwriting Recognition

Choices

- Get path information
- Get time data
- Get pressure information
- Only get image

Major techniques

- Clustering (not great performance)
- SVM (until 2006 or so)
- Convolutional neural networks

Questions?

`purple.com/talk-feedback`