AI Helsinki Academia

Principal Components and Whitening

Srikanth Gadicherla
srikanth.gadicherla@aalto.fi

October 31, 2016

Content

Gaussianity as the basis for PCA

- The probability model related to PCA
- PCA as generative model
- Image Synthesis results

Power spectrum of natural images

- Fourier amplitude or power spectrum
- Connection between power spectrum and covariances

Mathematics of PCA

The probability model related to PCA
Gaussianity as the basis for PCA

- Variances and Covariances are central aspect of data in PCA and whitening techniques. So, Gaussian probability distribution is quite appropriate here as it is completely defined in defined in terms of covariances (if mean is zero) which goes as:

$$p(x_1, ..., x_2) = \frac{1}{(2\pi)^{n/2}|C|^{1/2}} exp\left(-\frac{1}{2}\sum_{i,j} x_i, x_j[C^{-1}]_{ij}\right) \qquad (1)$$

  where $[C^{-1}]_{ij}$ is the *i*-th row and *j*-th column of inverse of covariance matrix.
- If the data is distributed normally, then covariance based methods usually suffice, but in reality the image data far from being Gaussian.
- Bayesian inference could be applied to image data if we consider image data as Gaussian likelihood and interpret PCA as statistical model. But, constructing the prior distribution $p(x_1, ..., x_n)$ is huge challenge.

The probability model related to PCA

Gaussianity as the basis for PCA

- A generative model learns about the (joint)density of the data (unlike a discriminative model which learns about conditional density)
- Here, we can represent the data in terms of linear transformation of the principal components.

$$I(x, y) = \sum_i W_i(x, y) s_i \tag{2}$$

- It is possible as W_i vectors are orthogonal, then inverse of the matrix is its matrix transpose.

The probability model related to PCA
Gaussianity as the basis for PCA

- Therefore, feature detector weights from PCA (eigen vectors) are same as feature vectors in the generative model.
- So, therefore the distribution of s_i is Gaussian with variance equal to variance of i-th principal component(i.e., the eigen value)
- s_i's are statistically independent of each other.

Image Synthesis results

Gaussianity as the basis for PCA

- We sample from the model.
- This gives smooth component which belong to largest principal components.

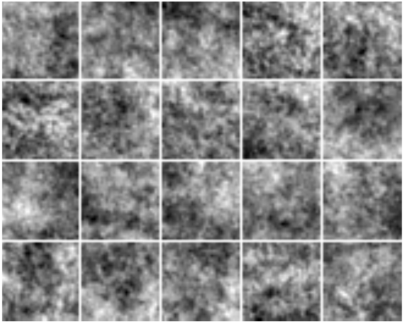

Figure: 20 images generated from PCA model. [1]

Fourier amplitude or power spectrum
Power spectrum of natural images

- From frequency based representation of natural images,

$$Powerspectrum \propto \frac{1}{f^2}$$
$$\propto Fourier\ Amplitude^2$$

- Therefore,

$$Power\ Amplitude \propto \frac{1}{f}$$

- In logarithms,

$$log\ Fourier\ amplitude = -log\ f\ +\ constant$$

or

$$log\ power\ spectrum = -2log\ f\ +\ constant$$

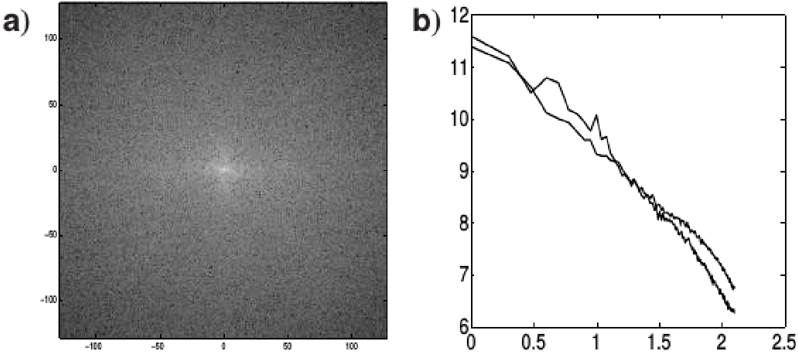

Figure: The average over orientations of one-dimensional cross-sections of the power spectrum of the two images.[1]

- From the Wiener-Khinchin theorem, in a stochastic process, the average power spectrum is the Fourier Transform of the autocorrelation function.
- Autocorrelation is nothing but correlation as a function of vertical and horizontal distances.

$$\begin{pmatrix} \mathrm{cov}(I(x_0,y_0),I(1,1)) & \ldots & \mathrm{cov}(I(x_0,y_0),I(1,n)) \\ \vdots & & \\ \mathrm{cov}(I(x_0,y_0),I(n,1)) & \ldots & \mathrm{cov}(I(x_0,y_0),I(n,n)) \end{pmatrix}$$

Figure: Autocorrelation matrix where (x_0, y_0) is a centre to avoid border effect. [1]

Fourier amplitude or power spectrum

Power spectrum of natural images

- From Wiener-Khinchin theorem, the Fourier transformation of C(x0, y0), the Fourier amplitudes equal the average power spectrum of the original image.
- The power spectrum has the similar information as the covariance matrix, and using covariances only is equivalent to using a Gaussian model.

Eigen Value Decomposition - covariance matrix

Mathematics of PCA

- As already told, complete data can be described by covariance matrix.

$$C(x,y;x'y') = E\{I(x,y)I(x',y')\}$$

- Covariance can be expanded as:

$$E\left\{\left[\sum_{x,y} W_1(x,y)I(x,y)\right]\left[\sum_{x,y} W_2(x,y)I(x,y)\right]\right\}$$

$$= E\left\{\left[\sum_{x,y,x',y'} W_1(x,y)I(x,y)W_2(x',y')I(x',y')\right]\right\}$$

$$= \sum_{x,y,x',y'} W_1(x,y)W_2(x',y')E\{I(x,y)I(x',y')\}$$

$$= \sum_{x,y,x',y'} W_1(x,y)W_2(x',y')C(x,y,x',y')\}$$

- A property of symmetric matrix says that it can decomposed into:

$$C = UDU^T$$

where U is orthogonal matrix and D is diagonal matrix $diag(\lambda_1, .., \lambda_m)$

- U's columns are the eigen vectors, whiles the λ_i's are eigen values.
- PCA problem: Maximization of variance (using EVD)

$$\max_{w:||w||=1} w^T C w \implies \max v^T D v \implies \max \sum_i v_i^2 \lambda_i$$

Properties of eigen values and eigen vectors

1. The eigen values are in descending order in D matrix(principle of maximization of variance)
2. Correspondingly, the eigen vectors(columns of U matrix) are placed.
3. The eigen values are the variances of the principle components.
4. The principal components are uncorrelated. $E\{ss^T\} = D$ (so to find whithened data matrix just divide the principal components with sqaure root corresponding eigen values)

Thank you for your attention!

References

[1] Aapo Hyvarinen, Jarmo Hurri, and Patrick O. Hoyer.(2009) *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision.*, Computational Imaging and Vision, Volume 39 2009 , Sp ringer, ISBN: 978-1-84882-490-4 (Print), 978-1-84882-491-1 (Online).