Feature Whitening: PCA and ZCA

Thomas Pellegrini

Université de Toulouse; UPS; IRIT; Toulouse, France
Samovert Montlaur 8 juin 2017

Feature preprocessing: whitening

- Mean and variance not particularly relevant $\rightarrow$ standardize the data
- i.e. we remove the fisrt and second order statistics

- PCA and ZCA
  - Let μ be the mean, Σ the covariance matrix
  - We decompose Σ as $\Sigma = ULU^t$

PCA

$$X \leftarrow L^{-1/2}U^t(X - \mu)$$

ZCA

$$X \leftarrow UL^{-1/2}U^t(X - \mu)$$

Cost to compute L and U : $\mathcal{O}(D^3)$

PCA

$x_n \leftarrow L^{-1/2} U^t (x_n - \mu)$

mean :

$$\frac{1}{N} \sum_n x_n = \frac{1}{N} \sum_n L^{-1/2} U^t (x_n - \mu)$$
$$= L^{-1/2} U^t \left(\frac{1}{N} \sum_n x_n - \mu\right) = 0$$

covariance :

$$\frac{1}{N} \sum_n x_n x_n^t = \frac{1}{N} \sum_n L^{-1/2} U^t (x_n - \mu)(x_n - \mu)^t U L^{-1/2}$$
$$= L^{-1/2} U^t \frac{1}{N} \sum_n (x_n - \mu)(x_n - \mu)^t U L^{-1/2}$$
$$= L^{-1/2} U^t \Sigma U L^{-1/2} = L^{-1/2} U^t U L U^t U L^{-1/2} = \ldots = I$$

Special case: $N << D$

On cherche $U = [\mathbf{u_0}, \mathbf{u_1}, \ldots, \mathbf{u_{D-1}}]$ avec $\Sigma \mathbf{u_i} = \lambda_i \mathbf{u_i}$

$$
\begin{aligned}
\Sigma \mathbf{u_i} = \lambda_i \mathbf{u_i} &\Leftrightarrow N^{-1} X^t X \mathbf{u_i} = \lambda_i \mathbf{u_i} \\
&\Leftrightarrow N^{-1} X X^t (X \mathbf{u_i}) = \lambda_i (X \mathbf{u_i}) \\
&\Leftrightarrow N^{-1} X X^t (\mathbf{v_i}) = \lambda_i (\mathbf{v_i})
\end{aligned}
$$

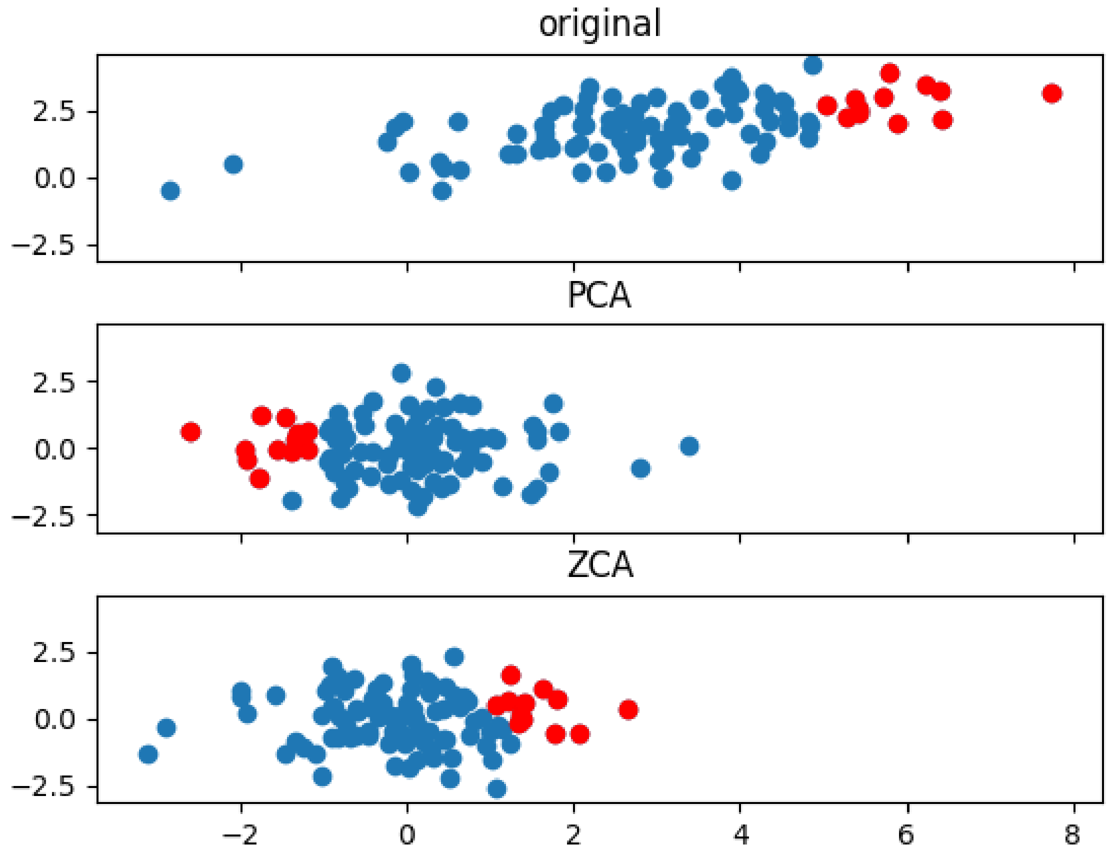
$\mathbf{v_i} = X \mathbf{u_i}$ eigen vector of $N^{-1} X X^t$ of dimension $N \times N << D \times D$

To derive $\mathbf{u_i}$, multiply the last equation by X^t on both sides:

$$
\begin{aligned}
N^{-1} X^t X (X^t (\mathbf{v_i})) &= \lambda_i (X^t \mathbf{v_i}) \\
&\Rightarrow \mathbf{u_i} = X^t \mathbf{v_i}
\end{aligned}
$$

Normed: $\mathbf{u_i} = \frac{1}{(N\lambda_i)^{1/2}} X^t \mathbf{v_i}$

Comparison between PCA and ZCA



In practice

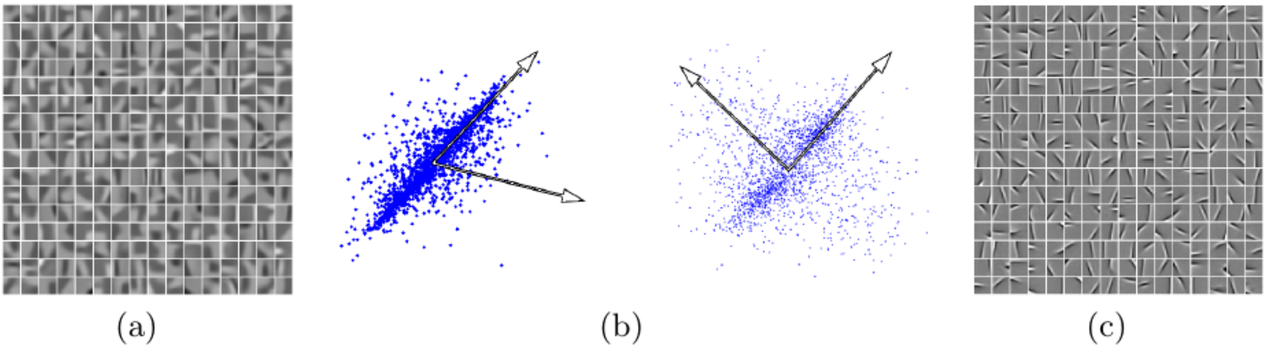
$$X \leftarrow U(L + \epsilon_{ZCA} I)^{-1/2} U^t (X - \mu)$$

Good starting values for images of

- 16×16 pixels: $\epsilon_{ZCA} = 0.01$
- 8×8: $\epsilon_{ZCA} = 0.1$

A. Coates, and A. Y. Ng. ”Learning feature representations with k-means.” Neural networks: Tricks of the trade. Springer Berlin Heidelberg, 2012. 561-580.

ZCA: centroïds through kmeans applied on images



A. Coates, and A. Y. Ng. ”Learning feature representations with k-means.” Neural networks: Tricks of the trade. Springer Berlin Heidelberg, 2012. 561-580.

ZCA on a 10-s speech signal: 13 MFCC + Δ + ΔΔ