Neural networks

Sparse coding - ZCA preprocessing

PREPROCESSING

Topics: ZCA

- Before running a sparse coding algorithm, it is beneficial to remove "obvious" structure from the data
  - normalize such that mean is 0 and covariance is the identity (whitening)
  - this will remove 1st and 2nd order statistical structure

- ZCA preprocessing
  - let the empirical mean be $\widehat{\boldsymbol{\mu}}$ and the empirical covariance matrix be $\widehat{\Sigma} = \mathbf{U}\Lambda\mathbf{U}^\top$ (in its eigenvalue/eigenvector representation)
  - ZCA transforms each input $\mathbf{x}$ as follows:
    - $\mathbf{x} \Longleftarrow \mathbf{U}\,\Lambda^{-\frac{1}{2}}\,\mathbf{U}^\top(\mathbf{x} - \widehat{\boldsymbol{\mu}})$

PREPROCESSING

Topics: ZCA

- After this transformation
  - the empirical mean is 0

$$
\begin{aligned}
& \frac{1}{T}\sum_t \mathbf{U}\,\Lambda^{-\frac{1}{2}}\,\mathbf{U}^\top(\mathbf{x}^{(t)} - \widehat{\boldsymbol{\mu}}) \\
= \quad & \mathbf{U}\,\Lambda^{-\frac{1}{2}}\,\mathbf{U}^\top \left(\left(\frac{1}{T}\sum_t \mathbf{x}^{(t)} \right) - \widehat{\boldsymbol{\mu}} \right) \\
= \quad & \mathbf{U}\,\Lambda^{-\frac{1}{2}}\,\mathbf{U}^\top (\widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\mu}}) \\
= \quad & 0
\end{aligned}
$$

PREPROCESSING

Topics: ZCA

- After this transformation
  - the empirical covariance matrix is the identity

$$\frac{1}{T-1}\sum_t \left(\mathbf{U}\,\Lambda^{-\frac{1}{2}}\,\mathbf{U}^\top(\mathbf{x}^{(t)}-\widehat{\boldsymbol{\mu}})\right)\left(\mathbf{U}\,\Lambda^{-\frac{1}{2}}\,\mathbf{U}^\top(\mathbf{x}^{(t)}-\widehat{\boldsymbol{\mu}})\right)^\top$$

$$= \quad \mathbf{U}\,\Lambda^{-\frac{1}{2}}\,\mathbf{U}^\top \left(\frac{1}{T-1}\sum_t (\mathbf{x}^{(t)}-\widehat{\boldsymbol{\mu}})(\mathbf{x}^{(t)}-\widehat{\boldsymbol{\mu}})\right)^\top \mathbf{U}\,\Lambda^{-\frac{1}{2}}\,\mathbf{U}^\top$$

$$= \quad \mathbf{U}\,\Lambda^{-\frac{1}{2}}\,\mathbf{U}^\top\,\widehat{\Sigma}\,\mathbf{U}\Lambda^{-\frac{1}{2}}\,\mathbf{U}^\top$$

$$= \quad \mathbf{U}\,\Lambda^{-\frac{1}{2}}\,\mathbf{U}^\top\,\mathbf{U}\,\Lambda\,\mathbf{U}^\top\,\mathbf{U}\,\Lambda^{-\frac{1}{2}}\,\mathbf{U}^\top$$

$$= \quad \mathrm{I}$$