Humboldt-Universität zu Berlin
Institute for Theoretical Biology

Exercise Set 3
May–8–2006

Computational Neuroscience IV: Analysis of Neural Systems

Dr. R. Kempter, Prof. Dr. A.V.M. Herz

Principal Component Analysis and Whitening

Whitening means that we linearly transform a zero-mean random vector $\mathbf{x}$ with n elements by multiplying it with some $n \times n$-matrix $\mathbf{V}$ so that the n-vector $\mathbf{z} = \mathbf{V}\,\mathbf{x}$ has a covariance matrix $\mathbf{C}_{\mathbf{z}} = \mathbf{I}$.

Let $\mathbf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ be the matrix whose columns are the unit-norm eigenvectors of $\mathbf{C}_{\mathbf{x}}$, and let $\mathbf{D} = \text{diag}(d_1, \ldots, d_n)$ be the diagonal matrix of (positive) eigenvalues of $\mathbf{C}_{\mathbf{x}}$. An important instance of $\mathbf{V}$ is the matrix $\mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T$, which is called the inverse square root $\mathbf{C}_{\mathbf{x}}^{-1/2}$ of the covariance matrix $\mathbf{C}_{\mathbf{x}}$.

1. The whitening matrix $\mathbf{V}$ is not unique. Show that $\mathbf{U}\mathbf{V}$ is also a whitening matrix if $\mathbf{U}$ is an orthogonal matrix ($\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}$) and $\mathbf{V}$ is an arbitrary whitening matrix.

2. The vector $\hat{\mathbf{x}} = \sum_{i=1}^m (\mathbf{e}_i^T\mathbf{x})\mathbf{e}_i$ is called the projection of $\mathbf{x}$ onto the subspace spanned by the first m eigenvectors ($m \leq n$). Show that the mean-square (reconstruction) error is

$$E\{|\mathbf{x} - \hat{\mathbf{x}}|^2\} = \sum_{i=m+1}^{n} d_i \ .$$

Hint: $\text{tr}\{\mathbf{C}_{\mathbf{x}}\} = E\{\mathbf{x}^T\mathbf{x}\} = \sum_{i=1}^n d_i$.

3. An important practical problem in PCA is how to choose m, i.e. to find a limit below which the eigenvalues are so small as to be insignificant. Sometimes an optimal m can be found from prior information about the vector $\mathbf{x}$. For instance, assume that $\mathbf{x}$ is of length n and obeys a signal-noise model

$$\mathbf{x} = \sum_{i=1}^{m} \mathbf{a}_i s_i + \mathbf{n}$$

where $m < n$. The $\mathbf{a}_i$ are some fixed, pairwise orthogonal vectors that span an m-dimensional signal subspace, and the random vector $\mathbf{s} = (s_1, \ldots, s_m)^T$ is white. The term $\mathbf{n}$ is white noise, for which $E\{\mathbf{n}\mathbf{n}^T\} = \sigma^2\mathbf{I}$ holds. Show that

a) the covariance matrix of $\mathbf{x}$ is $\mathbf{C}_{\mathbf{x}} = \sum_{i=1}^m \mathbf{a}_i\mathbf{a}_i^T + \sigma^2\mathbf{I}$.

b) the eigenvalues are constant beyond index m: $d_1 \geq d_2 \geq \ldots \geq d_m > d_{m+1} = \ldots = d_n = \sigma^2$.

Higher-Order Moments

4. Show that for two zero-mean, independent random variables s_1 and s_2,

a) $\text{kurt}(s_1 + s_2) = \text{kurt}(s_1) + \text{kurt}(s_2)$ and

b) $\text{kurt}(\alpha\, s_1) = \alpha^4\text{kurt}(s_1)$, where α is constant.

5. Nongaussianity can be measured by the kurtosis. To use kurtosis for ICA estimation of sources $\mathbf{s}$, we need to prove that the maxima of the function

$$F_q(\mathbf{q}) = |\mathrm{kurt}(\mathbf{q}^T\mathbf{s})| = |q_1^4\,\mathrm{kurt}(s_1) + q_2^4\,\mathrm{kurt}(s_2)|$$

for $|\mathbf{q}| = 1$ are obtained when only one of the components of $\mathbf{q} = (q_1, q_2)^T$ is nonzero.

a) Make the change of variables $t_i = q_i^2$. What is the geometrical form of the constraint set of $\mathbf{t} = (t_1, t_2)^T$?

b) Assume that $\mathrm{kurt}(s_1) > \mathrm{kurt}(s_2) > 0$. What is the geometrical shape of sets $F_t(\mathbf{t}) = \mathrm{const.}$? By a geometrical argument, show that the maximum of $F_t(\mathbf{t})$ is obtained when one of the t_i is one and the other one is zero.

c) Assume that $\mathrm{kurt}(s_1) < \mathrm{kurt}(s_2) < 0$. Use the same logic as in b).

d) Assume that the kurtoses have different signs. What is the shape of sets $F_t(\mathbf{t}) = \mathrm{const.}$ now?

Information Theory

The differential entropy H of a continuous-valued random vector $\mathbf{x}$ with density $p_\mathbf{x}$ is defined as

$$H(\mathbf{x}) = -\int p_\mathbf{x}(\boldsymbol{\xi})\,\log p_\mathbf{x}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}\ .$$

The negentropy is defined as

$$J(\mathbf{x}) = H(\mathbf{x}_{gauss}) - H(\mathbf{x})$$

where $\mathbf{x}_{gauss}$ is a gaussian random vector of the same covariance matrix $\boldsymbol{\Sigma}$ as $\mathbf{x}$.

6. Show that the entropy is not scale-invariant, i.e. it changes as we linearly transform to a different coordinate system $\mathbf{y} = \mathbf{A}\,\mathbf{x}$, where $\mathbf{A}$ is invertible. Hint: Use the density of a transformation, $p_\mathbf{y}(\boldsymbol{\eta}) = |\mathrm{det}\mathbf{A}|^{-1}\,p_\mathbf{x}(\mathbf{A}^{-1}\boldsymbol{\eta})$, and prove $H(\mathbf{y}) = H(\mathbf{x}) + \log|\mathrm{det}\mathbf{A}|$.

7. Prove $H(\mathbf{x}_{gauss}) = \frac{1}{2}\log|\mathrm{det}\boldsymbol{\Sigma}| + \frac{n}{2}[1+\log 2\pi]$. Hint: use the definition of a gaussian pdf from Exercises 2.

8. Show that the negentropy is scale-invariant: $J(\mathbf{A}\,\mathbf{x}) = J(\mathbf{x})$.

Mutual information ("Transinformation") is a measure of the information that members of a set of random variables have about the other random variables in the set. The mutual information I between n scalar random variables, x_i, $i = 1, \ldots, n$, is

$$I(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} H(x_i) - H(\mathbf{x})$$

9. Show that the following relation holds:

$$I(x_1, x_2, \ldots, x_n) = \int p_\mathbf{x}(\boldsymbol{\xi})\,\log\frac{p_\mathbf{x}(\boldsymbol{\xi})}{p_1(\xi_1)\,p_2(\xi_2)\cdots p_n(\xi_n)}\,\mathrm{d}\boldsymbol{\xi}\ .$$

DR. R. KEMPTER, phone 2093-8925, room 2315, `r.kempter(AT)biologie.hu-berlin.de`
PROF. DR. A.V.M. HERZ, phone 2093-9112, room 2325, `a.herz(AT)biologie.hu-berlin.de`

Problems handed out on Monday, 08.05.2006.
Solutions to be handed in by Monday, 15.05.2006, 12^{15} pm.
Discussion and presentation of the problems on Friday, 19.05.2006, $8^{30} - 10^{00}$ am in front of room 2317 I-W (Zwischengeschoß).