Statistical Models and Data Analysis

Summer term 2017

Problem Set 7 **28.06.2017**

The solutions to this exercise should be ready by 11 am on 5.7.2017. If you have any questions, please send me an email at stemmler@bio.lmu.de. You may submit your exercises by email in one file, but please be legible!

---

1. (Principal components analysis) You are given the following data $\mathbf{x}^\mu$:

$$\mathbf{x}^1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \mathbf{x}^2 = \begin{pmatrix} 3 \\ 2 \end{pmatrix} \quad \mathbf{x}^3 = \begin{pmatrix} 4 \\ 5 \end{pmatrix} \quad \mathbf{x}^4 = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$$

Let $\mathbf{X}$ be the matrix consisting of the column vectors $\mathbf{x}^\mu$.

(a) (Computing the mean)

Compute the average $\langle \mathbf{x}^\mu \rangle$, where the triangular brackets $\langle \cdots \rangle$ denote the average over the set. Subtract $\langle \mathbf{x}^\mu \rangle$ from each vector $\mathbf{x}^\mu$ and form a new matrix $\widetilde{\mathbf{X}}$ with the mean-subtracted vectors as the columns.

(b) (Covariance)

Compute $\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}}$. What relationship does $\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}}$ have to the covariance of the data?

(c) (Eigensystem)

Derive the eigenvectors $\mathbf{e_1}$ and $\mathbf{e_2}$ of $\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}}$ and the corresponding eigenvalues. Normalize the eigenvectors to have unit length.

Why are these eigenvectors orthogonal to each other? Moreover, the eigenvalues of any such matrix have to be positive—explain why.

(d) (Projection)

Project the data $\mathbf{x}^\mu$ onto the eigenvectors, i.e.

$$c_1^\mu = \mathbf{e_1^T} \mathbf{x}^\mu$$
$$c_2^\mu = \mathbf{e_2^T} \mathbf{x}^\mu$$

Compute the variance of the coefficients c_1^μ and c_2^μ. What is the relationship of these variances to the eigenvalues computed earlier?