CK0146: Exercise 03b (2017.1)

Exercise Q.1 (The Gaussian distribution). You are given data $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ in which the observations $\mathbf{x}_n \in \mathcal{R}^D$ are assumed to be drawn independently from a multivariate Gaussian distribution $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with unknown mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

- Scatterplot the N observations, for all $(D \times D)$ pairs of variables ;
- Estimate the parameters of the distribution by maximum likelihood.

Let $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ denote the estimated parameters. Let $\mathbf{x}_a$ be a random vector consisting of the first M components of $\mathbf{x}$ and let random vector $\mathbf{x}_b$ comprise the remaining $(D - M)$ components.

- Calculate the (estimated) mean vector $\hat{\boldsymbol{\mu}}_{a|b}$, covariance matrix $\hat{\boldsymbol{\Sigma}}_{a|b}$ and precision matrix $\hat{\boldsymbol{\Lambda}}_{a|b}$ of the Gaussian distribution $p(\mathbf{x}_a|\mathbf{x}_b)$, for $M = 2$ and and for $M = 1$. You are free to set $\mathbf{x}_b$;
- Plot the two resulting conditional Gaussian distributions.

The data are available for download here.

Exercise Q.2 (Density estimation). You are given data $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ in which the observations $\mathbf{x}_n \in \mathcal{R}^D$ are assumed to be drawn independently from an unknown distribution $p(\mathbf{x})$.

- Estimate the unknown distribution using a mixture of K Gaussians and plot the result. You are free to choose the number K of Gaussian components in the mixture. Fit the distribution using either existing code or implement the maximum likelihood solution of the model yourself.

  Comment on the results;
- Plot the unknown distribution as estimated using k-NN with values of $k = \{1, 2, 4, 8, 16, 32, 64\}$. Fit the distribution using your implementation of the nearest-neighbour density estimator.

  Comment on the results.

The data are available for download here.