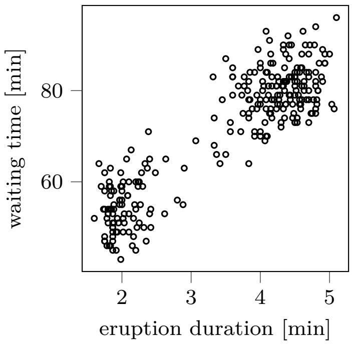
Probabilistic Machine Learning

Exercise Sheet #11

Variational Inference

1. **EXAMple Question — Gaussian Mixture Models.**

   The plot on the right shows a classic dataset: The interval between eruptions of the *Old Faithful* geyser in Yellowstone National Park in the US, plotted against the duration of the eruption following the waiting time. Datasets like this, which show a "cluster" structure, are often modelled with *Gaussian mixture models*: Each datum $\boldsymbol{x}_n \in \mathbb{R}^D$ for $n = 1, \dots, N$ is a real vector. The data are assumed to come from K separate Gaussian distributions (in this concrete case, $N = 272, D = 2, K = 2$), according to the following generative process:
   For each datum $n \in [1, \dots, N]$:



   - draw a discrete cluster identity $\boldsymbol{c}_n \in \{0, 1\}^K, \sum_k c_{nk} = 1$ ("one-hot") with

$$p(\boldsymbol{c}_n \mid \pi) = \prod_{k=1}^{K} \pi_k^{c_{nk}}$$

   - draw the datum $x_n \in \mathbb{R}^D$ from one of K Gaussian distributions, selected by $\boldsymbol{c}_n$ with probability

$$p(\boldsymbol{x}_n \mid \boldsymbol{c}_n, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{k=1}^{K} \mathcal{N}(\boldsymbol{x}_n; \mu_k, \Sigma_k)^{c_{nk}}$$

   This process, evidently, involves parameters: a probability $\pi \in \mathbb{R}^K_{+,0}, \sum_k \pi_k = 1$ and cluster parameters $\mu_k \in \mathbb{R}^d, \Sigma_k \in \mathbb{R}^{d \times d}$, spd.

   Your task: Draw a directed graphical model representing this generative process (with variables $\boldsymbol{x}_n, \boldsymbol{c}_n$ and the above parameters).

2. **Theory Question — EM for Gaussian Mixtures** Consider the Gaussian mixture model defined above. Write down an EM (expectation maximization) algorithm that finds a maximum likelihood assignment for the parameters π_k, μ_k, Σ_k defined above. To do so, proceed as in the latent Dirichlet allocation example in the lecture: For the E-step, compute the expectation of the complete-data likelihood $p(\boldsymbol{c}, \boldsymbol{x} \mid \mu, \Sigma, \pi)$ under the posterior $p(\boldsymbol{c} \mid \mu, \Sigma, \boldsymbol{x})$. For the M-step, analytically maximize this expression in π, μ, Σ.

3. **Practical Question.** In this week's coding exercise, we will implement the Kernel Topic Model, a latent Gaussian process model for the topic distributions of documents with meta-data. Use `Exercise_12.ipynb` as your workspace.