Advanced Machine Learning (GR5242)
Fall 2017

Homework 3

Due: Friday 27th of October at 4pm (for both sections of the class)

Homework submission: please submit your homework by publishing a notebook that cleanly displays your code, results and plots to pdf or html. You may wish to include another pdf file containing typeset or neatly scanned answers to the non-coding questions.

Problem 1 (Approximating in KL divergence by a Gaussian)

Let p be a target density on $\mathbb{R}^d$, and choose the approximating class for variational inference as the set of Gaussians densities

$$\mathcal{Q} := \left\{ q(x|\mu, \Sigma) \,\middle|\, \mu \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d\times d} \text{ positive definite } \right\} .$$

Now consider the optimization problem

$$\min_{q\in\mathcal{Q}} D_{\text{KL}}(p\|q)$$

Note we are not computing the expectation with respect to q, as in variational inference, but with respect to the target p. Show the solution is given by the Gaussian $q(x|\hat{\mu}, \hat{\Sigma})$ with

$$\hat{\mu} := \mathbb{E}_p[x] \qquad \text{and} \qquad \hat{\Sigma} := \text{Cov}_p[x]$$

(If you compare this solution to the illustration on slide 119, you will notice that it explains the large support of the approximating distribution.)

Problem 2 (Normal distribution)

For a set $x_1, \ldots, x_n$, we write x_{-i} for the set with the ith element removed,

$$x_{-i} = \{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\} .$$

a) Let p be a Gaussian density on $\mathbb{R}^d$, with mean μ and covariance Σ. Derive the full conditionals $p(x_i|x_{-i})$, for $i \in \{1, \ldots, d\}$.

b) Implement a Gibbs sampler for the d-dimensional gaussian distribution with mean μ and covariance Σ.

c) Implement variational inference where the target is a d-dimensional gaussian, and the approximating family is the family of d-dimensional spherical gaussians.

d) Let $d = 2$, and consider

$$\mu = (0, 1) \text{ and } \Sigma = \begin{pmatrix} 1.25 & -0.75 \\ -0.75 & 1.25 \end{pmatrix} .$$

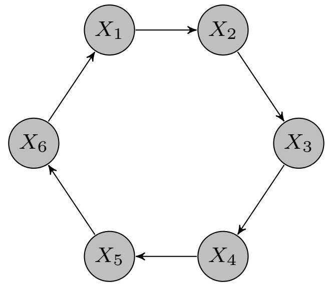
Draw 10000 samples using your Gibbs sampler, and 10000 samples from the approximating spherical gaussian. Compute the sample mean and sample covariance of each sample, and plot each one of them using a contour plot.

Problem 3 (Gibbs sampling)

For a set $x_1, \ldots, x_n$, we write x_{-i} for the set with the ith element removed,

$$x_{-i} = \{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\} .$$

a) Consider a directed graphical model with graph:



If $i = 6$, we read X_{i+1} as X_1. Suppose each variable X_i takes values in $\{-1, 1\}$, with conditional

$$P(X_{i+1} = 1 | X_i) = \sigma(\theta_i X_i) \qquad \text{for some } \theta_i \in [0, 1] \ ,$$

where σ is the sigmoid function. Compute an approximation to the full conditionals $P(X_i | X_{-i})$, assuming that $X_i \perp\!\!\!\perp X_{i+2} \mid X_{i+1}$ for all i (where again, the indices are taken modulo 6).

b) Implement a Gibbs sampler for the distribution using the approximation to the full conditionals you have derived.

c) Let $\theta = (2, -2, 2, -2, 2, -2)$. Draw 1000 samples from your Gibbs sampler, and estimate $P(X_1 = X_2)$.