Advanced Machine Learning (GR5242)
Fall 2017

Homework 2

Due: Wednesday 10 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.

For conciseness, we assume that you have imported the `numpy` package as `np`, and the `matplotlib.pyplot` package as `plt`.

Problem 1 (HMMs and topics)

This is a former exam problem.

Suppose we have to model text data which is streamed from a news feed; each news item is part of a single topic. After a (random) number of words, the new item ends and the next item begins, which in general has a different topic. Over time, topics may repeat. Suppose we have estimated empirically that:

- There are K topics, and we have estimated the probability vectors $\theta_1, \dots, \theta_K$ (where θ_k is the parameter vector of a multinomial which models text with topic k).
- At any given word, the probability of remaining within the current topic is 0.99.
- The probability of switching to a different topic is 0.01. For simplicity, assume all topics are equally probable, so the probability to switch to a specific new topic is $q := \frac{0.01}{K-1}$ for each topic.

a) Define a hidden Markov model to model the word sequence $X_1, X_2, \dots$. Please make sure that you specify:

- The state space of your model.
- The observed and hidden variables.
- The transition and emission probabilities.

b) Are the variables X_i and X_{i+2} (for any $i \in \mathbb{N}$) stochastically dependent?

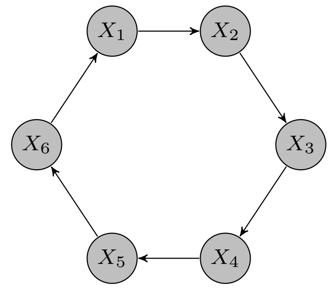
Problem 2 (Gibbs sampling)

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

$$x_{-i} = \{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d\} .$$

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 $i \in \{1, \dots, d\}$.

b) Consider a directed graphical model with graph:



If $i = 1$, we read X_{i-1} as X_6. Suppose each variable X_i takes values in $\{0, 1\}$, with conditional

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

where σ is the sigmoid function, given by $\sigma(y) = \frac{1}{1-\exp(-y)}$. What are the full conditionals $P(X_i = \bullet|X_{-i})$?

Problem 3 (Implementation)

- Implement a Gibbs sampler for the d-dimensional Gaussian.
- Run the sampler for $d = 2$ on a Gaussian with mean vector and covariance matrix

$$\mu = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 5 \end{pmatrix} \ ,$$

  and visualize the results: Plot the contour lines of Gaussian (for one standard deviation) against, say, 1000 samples of the Gibbs sampler.
- Implement Gibbs sampler for distribution in 2b above, with $\theta_1 = \ldots = \theta_6 = \frac{1}{3}$. Compare before and after burn-in.