Школа анализа данных

Семинар 8. Методы Монте Карло по схеме марковских цепей (MCMC)

Курс: Байесовские методы в машинном обучении, осень 2015

1. Рассмотрим схему Гиббса для генерации выборки из трёхмерного распределения $p(x, y, z)$, в которой один шаг генерации выполняется как

$$\hat{x} \sim p(x|y, z),$$
$$\hat{y} \sim p(y|\hat{x}, z),$$
$$\hat{z} \sim p(z|\hat{x}, \hat{y}).$$

Таким образом, введена марковская цепь с вероятностью перехода

$$q(\hat{x}, \hat{y}, \hat{z}|x, y, z) = p(\hat{x}|y, z)p(\hat{y}|\hat{x}, z)p(\hat{z}|\hat{x}, \hat{y}).$$

Доказать, что распределение $p(x, y, z)$ является инвариантным относительно введённой марковской цепи.

2. Рассмотрим схему Метрополиса-Хастингса, в которой в качестве предложного распределения используется схема Гиббса. Доказать, что вероятность принятия новых точек в такой схеме равна 100%.

3. Рассмотрим следующую байесовскую модель смеси распределений:

$$p(X, T, \boldsymbol{w}|\Theta, \alpha_0) = \left[\prod_{n,k=1}^{N,K} p(\boldsymbol{x}_n|\boldsymbol{\theta}_k)^{t_{nk}} w_k^{t_{nk}}\right] p(\boldsymbol{w}|\alpha_0),$$
$$p(\boldsymbol{w}|\alpha_0) = \mathrm{Dir}(\boldsymbol{w}|\alpha_0) = \frac{\Gamma(\alpha_0 K)}{\Gamma(\alpha_0)^K} w_1^{\alpha_0-1} \dots w_K^{\alpha_0-1}.$$

Здесь $t_{nk} \in \{0, 1\}$, $\sum_k t_{nk} = 1$ и $\sum_k w_k = 1$, $w_k \geq 0$. Требуется выписать формулы для генерации выборки из апостериорного распределения $p(T, \boldsymbol{w}|X, \Theta, \alpha_0)$ по схеме Гиббса.

4. Для модели смеси распределений из предыдущей задачи найти коллапсированное распределение $p(X, T|\Theta, \alpha_0)$, а также выписать формулы для генерации выборки из апостериорного распределения $p(T|X, \Theta, \alpha_0)$ по схеме Гиббса.