

Домашнее задание по МСМС и вариационному выводу

Курс: Байесовские методы анализа данных, 2020

Задача 1. Запишите алгоритм генерации $x \sim p(x) = C(x|a, b) = \frac{1}{\pi b[1+(\frac{x-a}{b})^2]}$ (распределение Коши) с помощью метода обратной функции.

Задача 2. Вы обучаете логистическую регрессию, используя байесовский подход, в частности, вариационный вывод. Априорное распределение на веса: $p(w) = \mathcal{N}(w|0, I)$, приближенное апостериорное распределение на веса: $q(w) = \mathcal{N}(w|\mu, \Sigma)$, $\mu \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d \times d}$, d – число признаков, свободный коэффициент не используется. Какой функционал необходимо минимизировать (и по какой величине), чтобы выполнить вариационный вывод? Запишите общий вид функционала и подставьте конкретные распределения согласно условию задачи.

Задача 3. Вы обучаете логистическую регрессию, используя байесовский подход, в частности, схему Метрополиса-Хастингса. Априорное распределение на веса: $p(w) = \mathcal{N}(w|0, I)$, $w \in \mathbb{R}^d$, d – число признаков, свободный коэффициент не используется. Предложное распределение: $r(\cdot|w^n) = \mathcal{N}(\cdot|w^n, 0.01I)$. Запишите схему генерации w^{n+1} (три шага схемы Метрополиса-Хастингса), подставьте в нее конкретные распределения согласно условию задачи.

Задача 3. Предположим, что мы обучаем байесовскую нейронную сеть с полностью факторизованным нормальным априорным распределением на веса $p(w) = \mathcal{N}(w|0, \lambda I)$, $\lambda > 0$ – гиперпараметр, и приближаем апостериорное распределение на веса с помощью полностью факторизованного нормального распределения: $q(w|\mu, \sigma) = \mathcal{N}(w|\mu, \text{diag}(\sigma))$, $\mu, \sigma \in \mathbb{R}^M$, M – число весов нейронной сети.

- Запишите трюк репараметризации для данной модели.
- Вычислите KL-регуляризатор (в ответе должна получиться формула от λ , μ и σ). Подсказка 1: сначала докажите, что KL-регуляризатор будет равен сумме регуляризаторов по весам. Подсказка 2: запишите определение KL-дивергенции, подставьте выражения для плотностей нормального распределения *под логарифмом* и воспользуйтесь формулами моментов нормального распределения.
- Запишите алгоритм обработки одного мини-батча: как выполняется проход вперед, от какого функционала вычисляются градиенты во время прохода назад, по каким величинам выполняется градиентный спуск.