Курс: Байесовские методы машинного обучения, 2011

Подход распространения ожидания (Expectation Propagation) для приближенного байесовского вывода

Дата: 7 декабря 2011

Достаточные статистики

Рассмотрим задачу оценки параметра распределения по выборке. Пусть имеется набор $X = (x_1, \ldots, x_N)$, где $x_1, \ldots, x_N$ – н.о.р.с.в. из распределения $p(x|\theta)$. Необходимо оценить значение параметра θ.

Статистикой распределения $T(X)$ будем называть произвольную функцию от наблюдений X. Статистика $T(X)$ для оценки параметра θ является достаточной, если любая другая статистика от той же выборки X не добавляет никакой новой информации о значении параметра θ. Формально это утверждение выглядит следующим образом:

$$p(X|T(X) = t, \theta) = p(X|T(X) = t).$$

В контексте байесовского подхода это выражение переходит в

$$p(\theta|X, T(X)) = p(\theta|T(X)).$$

Удобным средством для поиска достаточных статистик распределения является теорема факторизации Фишера-Неймана. Согласно этой теореме, статистика $T(X)$ является достаточной для оценки параметра θ тогда и только тогда, когда правдоподобие выборки $p(X|\theta)$ можно представить как

$$p(X|\theta) = h(X)g(\theta, T(X)),$$

где функция $h(X)$ не зависит от θ, а функция g выражает зависимость между θ и X только посредством статистики $T(X)$.

Рассмотрим для примера задачу оценивания мат.ожидания μ одномерного нормального распределения с известной дисперсией σ^2. Тогда правдоподобие выборки $p(X|\mu)$ можно представить как

$$p(X|\mu) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu, \sigma^2) = \frac{1}{\sqrt[n]{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu)^2\right) =$$

$$= \underbrace{\frac{1}{\sqrt[n]{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\sum_n x_n^2\right)}_{h(X)} \underbrace{\exp\left(\frac{\mu}{\sigma^2}\sum_n x_n - \frac{\mu^2}{2\sigma^2}\right)}_{g(\mu, \sum_n x_n)}.$$

Таким образом, статистика $T(X) = \sum_n x_n$ является достаточной для оценки мат. ожидания μ.

Экспоненциальное семейство распределений

Распределение $p(\boldsymbol{x})$ принадлежит экспоненциальному семейству распределений, если его плотность может быть представлена как

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} h(\boldsymbol{x}) \exp(\boldsymbol{\theta}^T \boldsymbol{u}(\boldsymbol{x})),$$

где $\boldsymbol{\theta}$ – набор параметров распределения, количество компонент вектора $\boldsymbol{u}(\boldsymbol{x})$ совпадает с размерностью $\boldsymbol{\theta}$, $h(\cdot)$ – некоторая функция, а $Z(\boldsymbol{\theta})$ – нормировочная константа распределения.

Многие стандартные вероятностные распределения принадлежат экспоненциальному семейству, например, нормальное, гамма, бета, Бернулли, Дирихле и многие другие. Соответствие между параметрами этих распределений и компонентами $\boldsymbol{\theta}, \boldsymbol{u}(\boldsymbol{x})$ в экспоненциальном представлении показано в таблице ниже.

Распределение	Плотность	$\boldsymbol{u}(\boldsymbol{x})$	$\boldsymbol{\theta}$
Бернулли	$q^x(1-q)^{1-x}$	x	$\log\frac{q}{1-q}$
Мультиномиальное	$\prod_k \mu_k^{x_k}$	$[x_1, \ldots, x_{K-1}]$	$\theta_i = \log\frac{\mu_i}{1-\sum_j \mu_j}$
Нормальное	$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$[x, x^2]$	$\left[-\frac{1}{2\sigma}, \frac{\mu}{\sigma^2}\right]$
Гамма	$\frac{b^a}{\Gamma(a)}x^{a-1}\exp(-bx)$	$[\log x, x]$	$[a-1, -b]$
Бета	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$	$[\log(x), \log(1-x)]$	$[a-1, b-1]$
Пуассон	$\exp(-\lambda)\frac{\lambda^x}{x!}$	$[x, \log\Gamma(x+1)]$	$[k, -1]$

Очевидно, что для экспоненциального семейства величина $T(X) = \sum_n \boldsymbol{u}(\boldsymbol{x}_n)$ является достаточной статистикой для параметра $\boldsymbol{\theta}$.

Пусть имеется набор распределений $p_1(\boldsymbol{x}), \ldots, p_N(\boldsymbol{x}), g_1(\boldsymbol{x}), \ldots, g_M(\boldsymbol{x})$ из экспоненциального семейства с набором параметров $\boldsymbol{\theta}_1^f, \ldots, \boldsymbol{\theta}_N^f, \boldsymbol{\theta}_1^g, \ldots, \boldsymbol{\theta}_M^g$ одинаковой размерности. Тогда распределение

$$p(\boldsymbol{x}) \propto \frac{p_1(\boldsymbol{x})\ldots p_N(\boldsymbol{x})}{g_1(\boldsymbol{x})\ldots g_M(\boldsymbol{x})}$$

будет также принадлежать экспоненциальному семейству с набором параметров $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^f, \ldots, \boldsymbol{\theta}_N^f, \boldsymbol{\theta}_1^g, \ldots, \boldsymbol{\theta}_M^g]$.

В экспоненциальном семействе распределений все моменты достаточных статистик $\boldsymbol{u}(\boldsymbol{x})$ могут быть вычислены путем дифференцирования логарифма нормировочной константы $Z(\boldsymbol{\theta})$. Найдем производящую функцию статистики $\boldsymbol{u}(\boldsymbol{x})$ для распределения $p(\boldsymbol{x}|\boldsymbol{\theta})$ из экспоненциального семейства:

$$M(\boldsymbol{t}) = \mathbb{E}_p \exp(\boldsymbol{t}^T\boldsymbol{u}(\boldsymbol{x})) = \frac{1}{Z(\boldsymbol{\theta})}\int \exp(\boldsymbol{t}^T\boldsymbol{u}(\boldsymbol{x}))h(\boldsymbol{x})\exp(\boldsymbol{\theta}^T\boldsymbol{u}(\boldsymbol{x}))d\boldsymbol{x} =$$
$$= \frac{Z(\boldsymbol{\theta}+\boldsymbol{t})}{Z(\boldsymbol{\theta})} = \exp(\log Z(\boldsymbol{\theta}+\boldsymbol{t}) - \log Z(\boldsymbol{\theta})).$$

Известно, что n-ый момент распределения может быть найден путем вычисления соответствующих производных производящей функции в нуле:

$$\mathbb{E}_p u_{i_1}(\boldsymbol{x})\ldots u_{i_n}(\boldsymbol{x}) = M^{(n)}_{i_1,\ldots,i_n}(\boldsymbol{0}).$$

Так как производящая функция в экспоненциальном семействе зависит только от $\log Z(\boldsymbol{\theta}+\boldsymbol{t})$, то все моменты $\boldsymbol{u}(\boldsymbol{x})$ определяются соответствующими производными $\log Z(\boldsymbol{\theta})$. В частности, можно показать, что

$$\begin{aligned}&\mathbb{E}\boldsymbol{u}(\boldsymbol{x}) = \nabla \log Z(\boldsymbol{\theta}),\\ &\operatorname{cov}(u_i(\boldsymbol{x}), u_j(\boldsymbol{x})) = (\nabla\nabla \log Z(\boldsymbol{\theta}))_{ij}.\end{aligned} \tag{1}$$

Свойство (1) можно использовать в две стороны. Предположим, что нам известно аналитическое выражение для нормировочной константы у распределения из экспоненциальном семейства (как в случае распределений из таблицы выше). Тогда для вычисления моментов достаточных статистик нет необходимости вычислять многомерные интегралы, достаточно лишь продифференцировать логарифм нормировочной константы нужное число раз. С другой стороны, в некоторых случаях оказывается удобным оценить моменты достаточных статистик и использовать их для оценок трудно вычислимой нормировочной константы и ее производных. Например, модель Изинга является представителем экспоненциального семества:

$$p(X|\theta) = \frac{1}{Z(\theta)}\exp(-\theta E(X)), \quad E(X) = -\sum_p h_p x_p - \frac{J}{2}\sum_{(i,j)\in\mathcal{E}} x_i x_j, \quad x_p \in \{-1, +1\}.$$

Тогда для оценки производной $\log Z(\theta)$ достаточно уметь находить $-\mathbb{E}E(X)$, что можно сделать, например, с помощью схемы Гиббса.

Приближенный байесовский вывод путем минимизации прямой КЛ-дивергенции

В дальнейшем будем рассматривать следующую задачу. Пусть нам известно некоторое распределение с точностью до нормировочной константы:

$$p(T) = \frac{\tilde{p}(T)}{Z}.$$

Здесь мы предполагаем, что значение $\tilde{p}(T)$ может быть легко вычислено для произвольной точки T, а значение нормировочной константы $Z = \int \tilde{p}(T)dT$ не может быть найдено. Задача состоит в поиске апроксимирующего распределения

$$q(T) \simeq p(T),$$

а также в оценке константы Z. В контексте байесовского вывода описанная задача соответствует следующей. Пусть имеется вероятностная модель $p(X, T)$, где X – множество наблюдаемых переменных, а T – множество скрытых переменных. Задача состоит в поиске приближения для апостериорного распределения $p(T|X)$, а также в оценке правдоподобия модели $p(X) = \int p(X, T)dT$.

Будем искать апроксимирующее распределение $q(T)$ путем минимизации КЛ-дивергенции:

$$\mathrm{KL}(p||q) \to \min_q.$$

Очевидно, что

$$\mathrm{KL}(p||q) = -\int p(T)\log\frac{q(T)}{p(T)}dT = -\int \log q(T)p(T)dT + \underbrace{\int \log p(T)p(T)dT}_{\text{const}} \to \min_q$$

Следовательно, исходная задача минимизации эквивалентна следующей:

$$\int \log q(T)p(T)dT = \frac{1}{Z}\int \log q(T)\tilde{p}(T)dT \to \max_q \;\Leftrightarrow \int \log q(T)\tilde{p}(T)dT \to \max_q. \tag{2}$$

Данная задача оптимизации требует, в частности, усреднения $\log q(T)$ по функции $\tilde{p}(T)$. По предположению вычисление нормировочной константы $\int \tilde{p}(T)dT$ является недоступным. Следовательно, вычислить $\int \log q(T)\tilde{p}(T)dT$ также не представляется возможным. Тем не менее, рассмотрим решение задачи максимизации (2) для случая, когда $q(T)$ принадлежит экспоненциальному семейству:

$$q(T) = \frac{1}{Z(\boldsymbol{\theta})}h(T)\exp(\boldsymbol{\theta}^T\boldsymbol{u}(T)).$$

Тогда

$$\int \log q(T)p(T)dT = -\log Z(\boldsymbol{\theta}) + \underbrace{\int \log h(T)p(T)dT}_{\text{const}} + \boldsymbol{\theta}^T\mathbb{E}_p[\boldsymbol{u}(T)] \to \max_{\boldsymbol{\theta}}.$$

Дифференцируя по $\boldsymbol{\theta}$ и приравнивая градиент к нулю, получаем:

$$\mathbb{E}_p[\boldsymbol{u}(T)] = \nabla_{\boldsymbol{\theta}}\log Z(\boldsymbol{\theta}) = \mathbb{E}_q[\boldsymbol{u}(T)].$$

Последнее равенство следует из свойства (1) для экспоненциального семейства распределений. Таким образом, минимизация КЛ-дивергенции соответствует приравниванию достаточных статистик $\mathbb{E}[\boldsymbol{u}(T)]$ распределений p и q.

Общая схема EP

Предположим, что исходное распределение $p(T)$ представимо в следующем виде:

$$p(T) = \frac{1}{Z}\prod_{j=1}^{J} f_j(T).$$

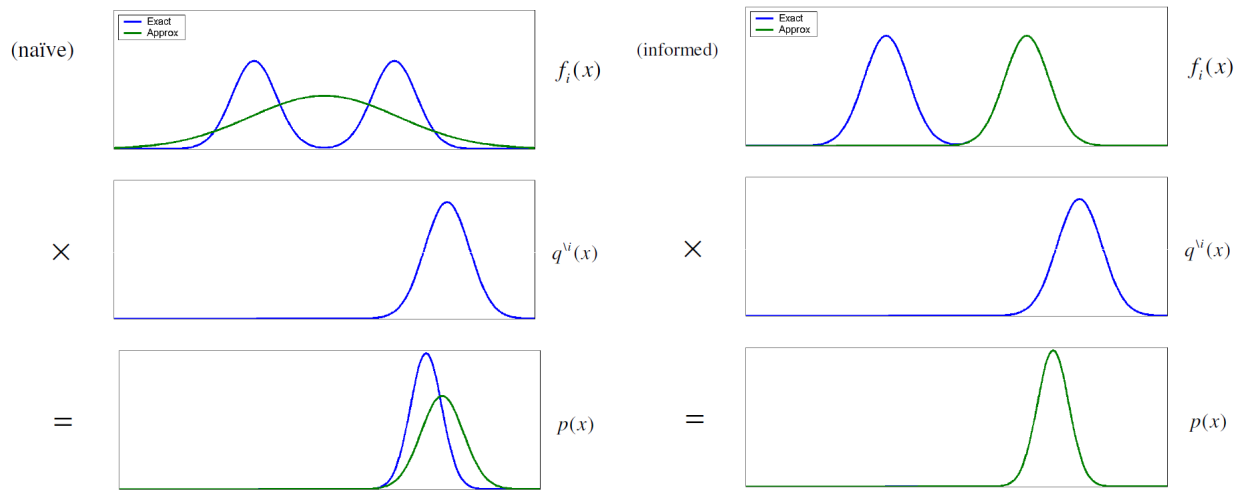

Рис. 1: Слева: индивидуальное приближение каждого фактора f_j может приводить к плохим результатам. Справа: приближение фактора f_j в контексте других факторов значительно улучшает качество итоговой апроксимации.

Здесь $f_j(T) > 0$ – некоторые факторы, а Z – нормировочная константа распределения. Такое представление для $p(T)$ часто встречается на практике. Например, функция правдоподобия для выборки X н.о.р.с.в. $\boldsymbol{x}_n$ вычисляется через произведение индивидуальных правдоподобий $p(X|T) = \prod_{n=1}^{N} p(\boldsymbol{x}_n|T)$.

Будем искать апроксимирующее распределение $q(T)$, которое также представимо в виде произведения факторов:

$$q(T) = \frac{1}{Z_q} \prod_{j=1}^{J} q_j(T).$$

Заметим, что решить задачу минимизации $KL(p(T)||q(T))$ по $q(T)$ не представляется возможным, т.к. для этого необходимо вычислять интеграл вида $\int p(T) \log q(T) dT$. Поэтому далее рассмотрим приближенные схемы решения этой задачи.

Сначала рассмотрим простейшую схему приближения. Предположим, что мы можем найти индивидуально для каждого истинного фактора f_j его оптимальное приближение q_j посредством минимизации КЛ-дивергенции $KL(f_j||q_j) \to \min_{q_j}$. Тогда возьмем в качестве $q(T)$ произведение индивидуально оптимальных факторов $q_j(T)$. Однако, такая схема апроксимации может приводить к сильно неоптимальным решениям (см. рис. 3).

Значительно лучший результат можно получить при приближении факторов f_j в контексте других факторов q_i, $i \neq j$. Предположим, что у нас имеется некоторое начальное приближение для всех факторов $q_j(X)$. Будем по очереди модифицировать один выбранный фактор q_j при фиксированных остальных факторах $q_i(X)$, $i \neq j$ путем решения следующей задачи:

$$\mathrm{KL}\left(\frac{1}{Z_p} f_j(T) \prod_{i \neq j} q_i(T) \,\middle\|\, \frac{1}{Z_q} q_j(T) \prod_{i \neq j} q_i(T)\right) \to \min_{q_j(T)}.$$

Здесь Z_p и Z_q – нормировочные константы соответствующих распределений. Обозначим произведение фиксированных факторов через $q^{\setminus j}(T) = \prod_{i \neq j} q_i(T)$. Предположим далее, что все факторы q_j принадлежат (ненормированному) экспоненциальному семейству распределений. Тогда $q^{new}(T) = \frac{1}{Z_q} q_j(T) q^{\setminus j}(T)$ также принадлежит экспоненциальному семейству. Выше было показано, что минимизация КЛ-дивергенции для экспоненциального семейства соответствует приравниваю моментов распределения $\mathbb{E}\boldsymbol{u}(T)$. Предположим, что данная операция может быть выполнена, т.е. мы можем вычислить необходимые моменты распределения $f_j(T)q^{\setminus j}(T)$ (например, в случае нормального распределения для $q^{new}(T)$ достаточно вычислить мат.ожидание и дисперсию распределения $f_j(T)q^{\setminus j}(T)$). Тогда новый фактор $q_j(T)$ может быть найден из $q^{new}(T)$ следующим

Алгоритм 1: Алгоритм Expectation Propagation

Вход: Распределение $p(X,T) = \prod_{j=1}^{J} f_j(T)$

Выход: Приближение $q(T)$ для апостериорного распределения $p(T|X)$ и оценка для обоснованности $p(X) = \int p(X,T)dT$

1: Инициализация всех факторов $q_j(T)$;
2: Инициализация приближения $q(T) \propto \prod_j q_j(T)$;
3: пока не достигнута сходимость
4: для $j = 1, \ldots, J$
5: $q^{\setminus j}(T) = \frac{q(T)}{q_j(T)}$;
6: Найти $q^{new}(T)$ путем приравнивания достаточных статистик распределения $f_j(T)q^{\setminus j}(T)$;
7: Вычислить нормировочную константу $K = \int f_j(T)q^{\setminus j}(T)dT$;
8: Обновить фактор $q_j(T) = \frac{Kq^{new}(T)}{q^{\setminus j}(T)}$;
9: Оценить обоснованность $p(X) \simeq \int \prod_j q_j(T)dT$.

образом:

$$q_j(T) = \frac{Kq^{new}(T)}{q^{\setminus j}(T)}.$$

Здесь K – некоторая константа. Заметим, что константу K можно выбрать любой, т.к. она не влияет на статистики распределения $q^{new}(T)$. Выберем ее таким образом, чтобы нулевые статистики распределений $f_j(T)q^{\setminus j}(T)$ и $q_j(T)q^{\setminus j}(T)$ также совпадали:

$$K = \int f_j(T)q^{\setminus j}(T)dT.$$

При таком выборе константы K величина $\int \prod_j q_j(T)dT$ будет оценкой нормировочной константы исходного распределения $p(X)$ (обоснованности). Факторы f_j обновляются в цикле до сходимости по статистикам распределений. Итоговая схема EP представлена как Алгоритм 1.

Пример применения EP

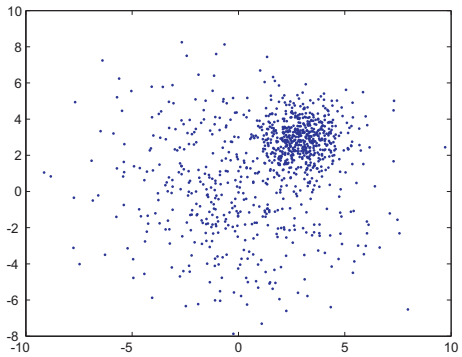

Рис. 2: Пример выборки для оценивания среднего значения в сгустке точек, помещенных в разреженное облако точек.

Рассмотрим задачу оценивания мат.ожидания $\boldsymbol{\theta}$ нормального распределения по выборке из него. Дополнительно предположим, что к выборке добавляется некоторое разреженное облако точек, также взятых из нормального распределения, но с большей дисперсией (см. рис. 2). Таким образом, модель наблюдений представляет собой смесь двух нормальных распределений

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = (1-w)\mathcal{N}(\boldsymbol{w}|\boldsymbol{\theta}, I) + w\mathcal{N}(\boldsymbol{w}|\boldsymbol{0}, aI),$$

где пропорция w и уровень шума a считаются известными. Введем также априорное распределение на параметр $\boldsymbol{\theta}$ следующим образом:

$$p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}|\mathbf{0}, bI).$$

В результате получаем вероятностную модель вида:

$$p(X, \boldsymbol{\theta}) = p(\boldsymbol{\theta}) \prod_{n=1}^{N} p(\boldsymbol{x}_n|\boldsymbol{\theta}). \tag{3}$$

Апостериорное распределение $p(\boldsymbol{\theta}|X)$ в данной модели пропорционально смеси из 2^N гауссиан, и поэтому оно не может быть вычислено аналитически уже для средних значений N.

Применим алгоритм EP для этой модели с целью получения приближения для апостериорного распределения $p(\boldsymbol{\theta}|X)$ и оценки обоснованности модели $p(X)$. Для этого рассмотрим модель (3) как модель с факторами $f_0(\boldsymbol{\theta}) = p(\boldsymbol{\theta})$ и $f_n(\boldsymbol{\theta}) = p(\boldsymbol{x}_n|\boldsymbol{\theta})$. В качестве факторов апроксимирующего распределения $q(\boldsymbol{\theta})$ возьмем ненормированные нормальные распределения вида

$$q_n(\boldsymbol{\theta}) = s_n \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{m}_n, v_n I), \quad q_0(\boldsymbol{\theta}) = p(\boldsymbol{\theta}).$$

Таким образом, итоговое апроксимирующее распределение $q(\boldsymbol{\theta}) \propto q_0(\boldsymbol{\theta}) \prod_{n=1}^{N} q_n(\boldsymbol{\theta})$ также является нормальным

$$q(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{m}, vI).$$

Заметим, что величины v_n не обязаны быть положительными (отдельные факторы q_n не обязаны быть корректными нормальными распределениями). Важно, чтобы итоговое апроксимирующее распределение $q(\boldsymbol{\theta})$ имело бы положительно-определенную матрицу ковариации, т.е. $v > 0$.

Рассмотрим процесс пересчета одного фактора q_n в схеме EP. Заметим при этом, что фактор $q_0 = p_0$ всегда остается неизменным. Сначала удалим текущий фактор q_n из апроксимирующего распределения q:

$$\begin{aligned}
&q^{\backslash n}(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{m}^{\backslash n}, v^{\backslash n} I), \\
&\boldsymbol{m}^{\backslash n} = \boldsymbol{m} + v^{\backslash n} v_n^{-1} (\boldsymbol{m} - \boldsymbol{m}_n), \\
&(v^{\backslash n})^{-1} = v^{-1} - v_n^{-1}.
\end{aligned}$$

Затем нам необходимо вычислить все достаточные статистики $\mathbb{E}\boldsymbol{u}(\boldsymbol{\theta})$ и нормировочную константу распределения $f_n(\boldsymbol{\theta}) q^{\backslash n}(\boldsymbol{\theta})$. Для нормального распределения достаточными статистиками являются мат.ожидание и дисперсия. Все эти величины могут быть вычислены аналитически:

$$\begin{aligned}
Z_n &= (1-w)\mathcal{N}(\boldsymbol{x}_n|\boldsymbol{m}^{\backslash n}, (v^{\backslash n}+1)I) + w\mathcal{N}(\boldsymbol{x}_n|\mathbf{0}, aI), \\
\boldsymbol{m} &= \boldsymbol{m}^{\backslash n} + \rho_n \frac{v^{\backslash n}}{v^{\backslash n}+1}(\boldsymbol{x}_n - \boldsymbol{m}^{\backslash n}), \\
v &= v^{\backslash n} - \rho_n \frac{(v^{\backslash n})^2}{v^{\backslash n}+1} + \rho_n(1-\rho_n)\frac{(v^{\backslash n})^2 \|\boldsymbol{x}_n - \boldsymbol{m}^{\backslash n}\|}{D(v^{\backslash n}+1)^2}, \\
\rho_n &= 1 - \frac{w}{Z_n}\mathcal{N}(\boldsymbol{x}_n|\mathbf{0}, aI).
\end{aligned}$$

Здесь D – размерность пространства $\boldsymbol{\theta}$. Величина ρ_n имеет смысл вероятности того, что $\boldsymbol{x}_n$ принадлежит сгустку точек. После получения нового апроксимирующего распределения $q(\boldsymbol{\theta})$ параметры обновленного фактора q_n определяются как

$$\begin{aligned}
v_n^{-1} &= v^{-1} - (v^{\backslash n})^{-1}, \\
\boldsymbol{m}_n &= \boldsymbol{m}^{\backslash n} + (v_n + v^{\backslash n})(v^{\backslash n})^{-1}(\boldsymbol{m} - \boldsymbol{m}^{\backslash n}), \\
s_n &= \frac{Z_n}{(2\pi v_n)^{D/2}\mathcal{N}(\boldsymbol{m}_n|\boldsymbol{m}^{\backslash n}, (v_n + v^{\backslash n})I)}.
\end{aligned}$$

Итерационный процесс пересчета всех факторов q_n по представленным выше формулам продолжается до сходимости по всем параметрам $(s_n, \boldsymbol{m}_n, v_n)$. В качестве начального приближения можно

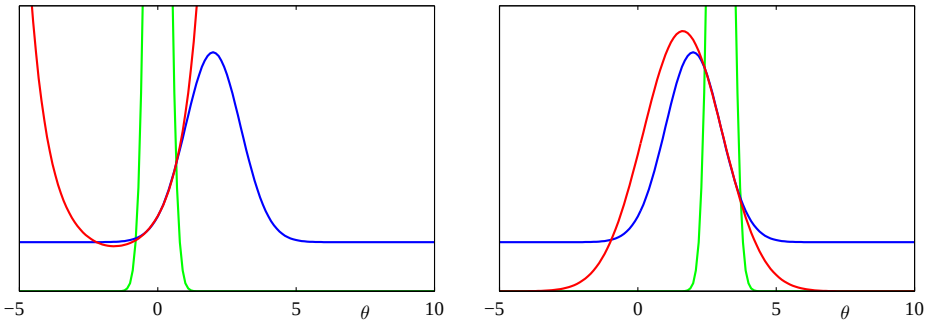

Рис. 3: Примеры контекстного приближения фактора f_n (синяя кривая) фактором q_n (красная кривая) для скалярного параметра θ. Контекст $q^{\setminus n}$ обозначен зеленой кривой.

выбрать значения $s_n = (2\pi v_n)^{D/2}, v_n \to +\infty, \boldsymbol{m}_n = \mathbf{0}$. Это соответствует ситуации $q(\boldsymbol{\theta}) = q_0(\boldsymbol{\theta})$. После сходимости итерационного процесса значение обоснованности модели $p(X)$ может быть оценено как

$$p(X) \approx (2\pi v)^{D/2} \exp(B/2) \prod_{n=1}^{N} \frac{s_n}{(2\pi v_n)^{D/2}},$$

$$B = \frac{\boldsymbol{m}^T \boldsymbol{m}}{v} - \sum_{n=1}^{N} \frac{\boldsymbol{m}_n^T \boldsymbol{m}_n}{v_n}.$$

В заключение рассмотрим важность требования того, чтобы факторы q_n необязательно были корректными нормальными распределениями, т.е. дисперсия v_n может быть отрицательной или уходить в бесконечность. На рис. 3 показаны примеры контекстной апроксимации факторов f_n факторами q_n для случая $D = 1$ (скалярный параметр θ). Как видно из левого рисунка, приближение с помощью корректного нормального распределения в данном случае будет неадекватным.