Аппроксимация функции ошибки*

М. Е. Панов

panov.maxim@gmail.com

Московский физико-технический институт, ФУПМ, каф. «Интеллектуальные системы»

Введение

В работе рассматривается метод аппроксимации функции ошибки функцией многомерного нормального распределения. Рассматриваются случаи матрицы ковариации общего вида, диагональной матрицы ковариации, а также диагональной матрицы ковариации с равными значениями дисперсии. Для нормировки получившихся функций распределения используется аппроксимация Лапласа.

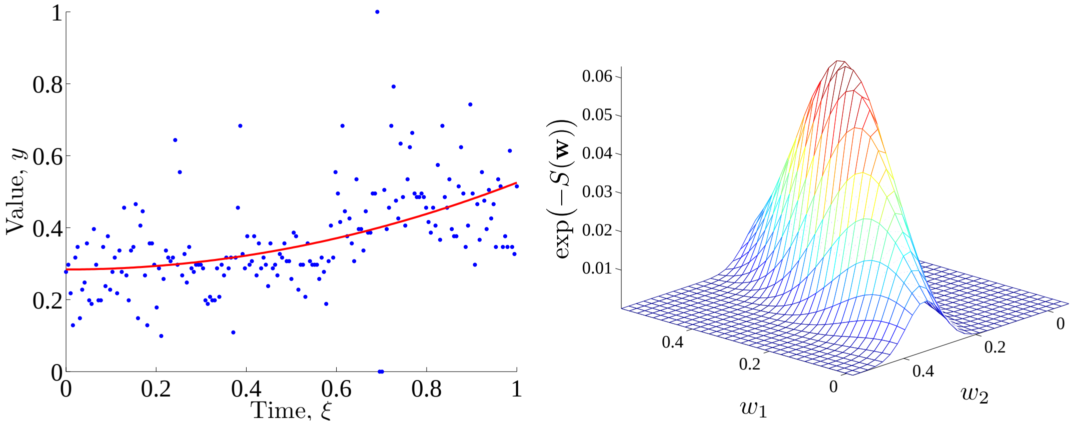

Рис. 1. Аппроксимация данных с помощью функции $f(x,w) = w_1 + w_2 * x^2$ при оптимальных значениях параметров $w = w_{MP}$ и ненормированная функции ошибки $exp(-S(w))$ в окрестности $w = w_{MP}$.

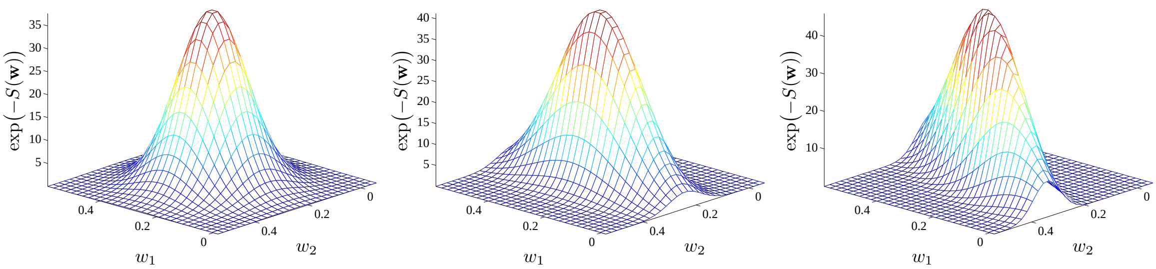

Рис. 2. Аппроксимации с помощью функции нормального распределения с различными видами матрицы ковариаций: с постоянной дисперсией, диагональной и общего вида.

Постановка задачи

Дана выборка $D = \{(x_i, y_i)\}_{i=1}^N$, где $x_i \in \mathbb{R}^n, i = 1, \ldots, N$ — вектора независимой переменной, а $y_i \in \mathbb{R}, i = 1, \ldots, N$ — значения зависимой переменной.

Научный руководитель В.В. Стрижов

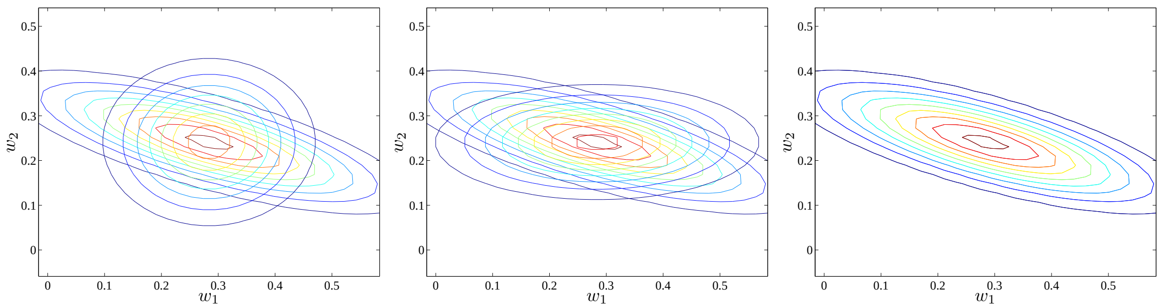

Рис. 3. Сравнение различных типов аппроксимации с реальной функцией ошибки.

Предполагается, что

$$y = f(x, w),$$

где $f(x, w)$ — некоторая параметрическая функция, $w \in W$ — вектор ее параметров. Предполагается, что задано апостериорное распределение параметров модели $p(w|D, f)$, которому соответствует функция ошибки $S(w)$: $p(w|D, f) = \frac{exp(-S(w))}{Z_S}$. Пусть $w_{MP} = \arg\max_w p(w|D, f)$ — наиболее вероятные параметры модели. Требуется найти аппроксимацию Лапласа для функции $p(w|D, f)$ в точке w_{MP}. Заметим, что в данной работе в качестве функции ошибки берется сумма квадратов ошибок аппроксимации $S(w) = \sum_{i=1}^{N}(y_i - f(x_i, w))^2$.

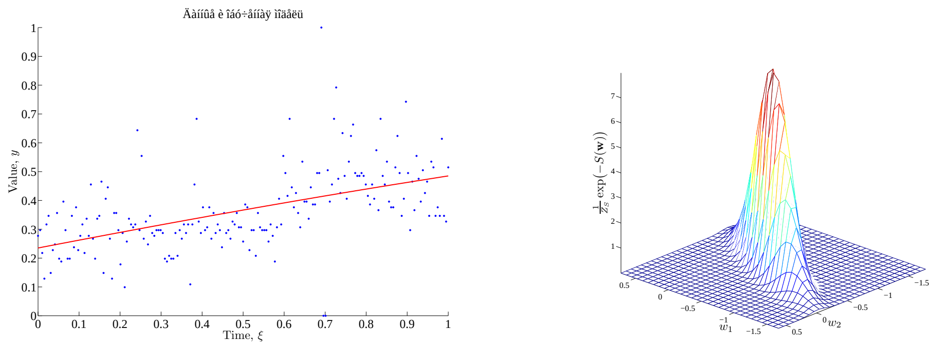

Рис. 4. Аппроксимация данных с помощью функции $f(x, w) = \frac{1-\exp(w_1+w_2*x)}{1+\exp(w_1+w_2*x)}$ при оптимальных значениях параметров $w = w_{MP}$ и ненормированная функции ошибки $exp(-S(w))$ в окрестности $w = w_{MP}$.

Описание алгоритма

Сначала находим оптимальные значения параметров модели w:

$$w_{MP} = \arg\max_w p(w|D, f).$$

Далее необходимо найти аппроксимацию Лапласа в точке w_{MP}: $p^*(w|k, A) = k * \exp(-(w - w_{MP})^T A(w - w_{MP}))$, где A — матрица, обратная к ковариационной матрице нормального распределения, а k — нормирующий коэффициент. Заметим, что в силу положительной определенности матрицы A ее можно представить в соответствии с разложением Холецкого: $A = LL^T$, где L — верхнетреугольная матрица. Параметризуем матрицу L следующим

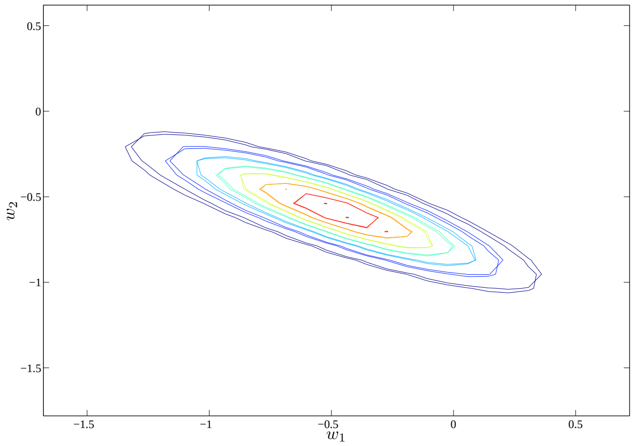

Рис. 5. Контурный график ее аппроксимации Лапласа с ковариационной матрицей общего вида.

образом: $L(i,j) = \begin{cases} exp(h_{ij}) & i=j, \\ sinh(h_{ij}) & j>i, \\ 0 & j<i, \end{cases}$ где $h_{ij} \in \mathbb{R}, i,j = 1,\ldots,N, j \geqslant i$. Также параметризуем нормирующий множитель $k = exp(h_0)$. Получаем, что $p^*(w|A,k) = p^*(w|h_{ij}, i,j = 1,\ldots,N, j \geqslant i, h_0)$. Построим обучающую выборку $D_S = (w_k, S(w_k)), k = 1,\ldots,N_S$, где точки w_k берутся равномерно из окрестности наиболее вероятных параметров w_{MP}, в которой мы хотим построить аппроксимацию. Для нахождения неизвестных параметров $h_{ij}, i,j = 1,\ldots,N, j \geqslant i, h_0$ минимизируем квадратичный критерий для точек обучающей выборки D_S:

$$\sum_{k=1}^{N_S} (S(w_k) - p^*(w_k|h_{ij}, h_0))^2 \to \min_{h_{ij}, h_0}. \quad (1)$$

Заметим, что получаемые в результате решения оптимизационной задачи [1] значения параметров могут существенно отличаться в зависимости от используемого для ее решения оптимизационного алгоритма. В данной работе рассматриваются два алгоритма оптимизации: Левенберг-Марквардт и Trust region.

После нахождения оптимальных значений параметров полученные распределения остается отнормировать в соответствии с аппроксимацией Лапласа: $Z_S = exp(-S(w_{MP})) * \sqrt{\frac{(2\pi)^n}{\det A}}$.

Вычислительный эксперимент: качество аппроксимации

В эксперименте в качестве обучающей выборки использовался временной ряд цен на хлеб из 195 точек. Для приближения использовалась модель линейной регрессии $f(x,w) = w_1 + w_2 * x^2$, а в качестве алгоритма оптимизации — алгоритм Левенберга-Марквардта [3, 4]. На картинках ниже графически представлены результаты.

Вычислительный эксперимент: устойчивость по начальным данным

Для сравнения устойчивости алгоритмов Левенберга-Марквардта и Trust region [2, 1] в качестве обучающей выборки использовался временной ряд цен на хлеб из 195 точек. Для приближения использовалась регрессионная модель $f(x,w) = \frac{1-\exp(w_1+w_2*x)}{1+\exp(w_1+w_2*x)}$. При таком

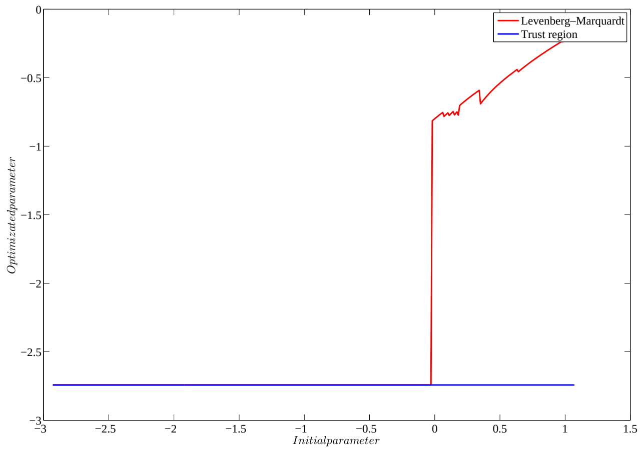

Рис. 6. Зависимость значения параметра h_0, полученного в результате оптимизации от его начального значения.

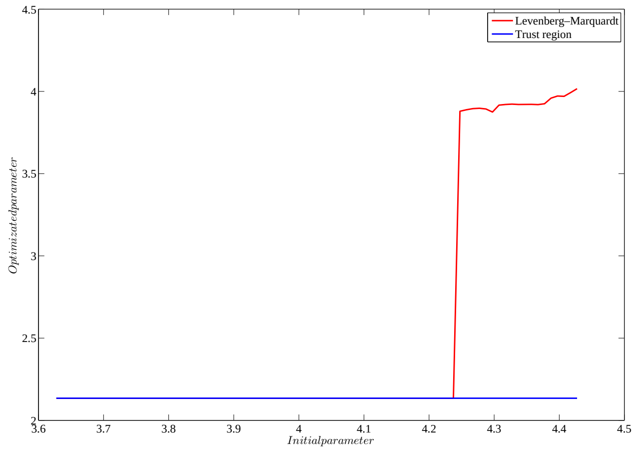

Рис. 7. Зависимость значения параметра h_{22}, полученного в результате оптимизации от его начального значения.

виде целевой функции вид функции ошибки в окрестности оптимума несколько отличается от нормального.

Рассматривалась зависимость оптимизированного значения параметров h_0 и h_{22} от начального значения.

Заключение

Функция ошибки в рассмотренных случаях хорошо аппроксимируется предложенным методом, причем качество аппроксимации возрастает с увеличением качества модели. Хорошее качество аппроксимации обусловлено тем, что функция ошибки в рассматриваемом примере принадлежит тому же классу, что и функция аппроксиматор, либо близка к нему. Сравнение алгоритмов оптимизации Левенберга-Марквардта и Trust region в применении к рассматриваемой задаче показало, что алгоритм Trust region гораздо более устойчив по начальным данным.

Литература

[1] *Coleman T., Li Y.* An interior, trust region approach for nonlinear minimization subject to bounds // *SIAM Journal on Optimization.* — 1994. — Vol. 6. — Pp. 418–445.

[2] *Coleman T., Li Y.* On the convergence of reflective newton methods for large-scale nonlinear minimization subject to bounds // *Mathematical Programming.* — 1994. — Vol. 67, no. 2. — Pp. 189–224.

[3] *Levenberg K.* A method for the solution of certain problems in least-squares // *Quarterly Applied Math.* — 1944. — Vol. 2. — Pp. 164–168.

[4] *Marquardt D.* An algorithm for least-squares estimation of nonlinear parameters // *SIAM Journal Applied Math.* — 1963. — Vol. 11. — Pp. 431–441.