

# МОДИФИКАЦИЯ АЛГОРИТМА ЛЕВЕНБЕРГА-МАРКВАРДТА ДЛЯ ЗАДАЧ НЕЛИНЕЙНОЙ РЕГРЕССИИ С УЧЕТОМ ПОГРЕШНОСТЕЙ КАК В ЗАВИСИМЫХ, ТАК И В НЕЗАВИСИМЫХ ДАННЫХ

Г. И. Рудой

## 1 Введение

Для известной задачи нахождения оптимальных коэффициентов некоторой фиксированной регрессионной модели, представленной в виде формулы, по набору экспериментальных данных широко применяется алгоритм Левенберга-Марквардта [?]. Однако, данный алгоритм построен и статистически обоснован в предположении о нормальности распределения регрессионных остатков и точно измеренных независимых переменных — иными словами, учитываются и рассматриваются только ошибки измерения зависимой переменной. Более того, предполагается, что ошибки для всех точек принадлежат одному и тому же распределению с одними и теми же параметрами.

В ряде физических приложений это предположение не выполняется. Например, в задаче нахождения дисперсионной зависимости прозрачного полимера (то есть, зависимости коэффициента преломления  $n$  от длины волны  $\lambda$ ) погрешности измерения различных физических параметров, вообще говоря, различны. Так, например, если для измерения длины волны  $\lambda$  используется дифракционная решетка, то постоянной является относительная погрешность определения длины волны  $\frac{\sigma_{\lambda_i}}{\lambda_i} \approx \text{const}$ , и, следовательно, погрешность определения длины волны зависит от самой длины волны.

Таким образом, возникает задача поиска оптимальных коэффициентов регрессионной формулы с учетом отличающихся погрешностей различных экспериментальных точек. Для некоторых частных случаев эта задача уже была решена: например, в работе [?] вводится предположение, что зависимые переменные  $y_i$  измеряются неточно, и каждая переменная  $y_i$  имеет свою собственную погрешность измерения  $\sigma_{y_i}$ . Затем в работе показывается, что обычный функционал суммы квадратов регрессионных остатков, где каждый остаток нормирован на соответствующую величину  $\sigma_{y_i}^2$ , корректен и статистически состоятелен.

В настоящей работе вводятся дополнительные предположения о том, что независимые переменные также измеряются неточно, и каждая переменная имеет свою собственную погрешность измерения. Предлагается функционал качества и модифицированный алгоритм Левенберга-Марквардта, позволяющий найти оптимальные параметры согласно этому функционалу качества и опирающийся на классический алгоритм Левенберга-Марквардта (в дальнейшем будем называть их мАЛМ и АЛМ соответственно). Доказывается сходимость модифицированного алгоритма и приводятся результаты на экспериментальных данных по измерению насыщения лазерного излучателя.

## 2 Постановка задачи

Дана обучающая выборка  $D = \{\mathbf{x}_i, y_i\} | i \in \{1, \dots, \ell\}, \mathbf{x}_i \in \mathbb{R}^m, y_i \in \mathbb{R}$ . Для каждой зависимой переменной переменной  $y_i$  известно стандартное отклонение ошибки ее измерения  $\sigma_{y_i}$ , а для соответствующего вектора независимых переменных  $\mathbf{x}_i$  аналогично известны стандартные отклонения его компонент  $\sigma_{x_{ij}} | j \in \{1, \dots, m\}$ . Пусть, кроме того, дана

некоторая параметрическая регрессионная модель  $y = f(\mathbf{x}, \boldsymbol{\omega})$ .

Для удобства обозначим вектор, составленный из ошибок измерений зависимых переменных  $\sigma_{y_i}$  как  $\boldsymbol{\sigma}_y$ :

$$\boldsymbol{\sigma}_y = \{\sigma_{y_1}, \dots, \sigma_{y_\ell}\}.$$

Аналогично обозначим матрицу, составленную из ошибок измерений компонент независимых переменных  $\sigma_{x_{ij}}$  как  $\Sigma_x$ :

$$\Sigma_x = \|\sigma_{x_{ij}}\|_{i \in \{1, \dots, \ell\}, j \in \{1, \dots, m\}}.$$

Требуется построить функционал ошибки  $S(\boldsymbol{\omega})$  вектора параметров  $\boldsymbol{\omega}$  модели  $f$ , учитывающий ошибки измерений  $\sigma_{y_i}$  и  $\sigma_{x_{ij}}$ :

$$S(\boldsymbol{\omega}) = S(\boldsymbol{\omega}, \boldsymbol{\sigma}_y, \Sigma_x), \quad (1)$$

и, кроме того, найти вектор параметров  $\boldsymbol{\omega}$ , минимизирующий функционал  $S$ :

$$\hat{\boldsymbol{\omega}} = \arg \min_{\boldsymbol{\omega}} S(\boldsymbol{\omega}) \quad (2)$$

### 3 Модифицированный функционал качества

Вводятся следующие физические соображения:

1. Чем больше погрешность определения зависимой переменной, тем меньше соответствующий регрессионный остаток должен учитываться при оптимизации параметров модели.
2. Чем больше погрешность определения некоторой независимой переменной, тем, аналогично, меньше должен учитываться соответствующий регрессионный остаток.
3. При прочих равных сильнее должны учитываться те точки, в которых производная регрессионной модели  $\frac{\partial f}{\partial x_j}$  по соответствующей компоненте  $x_j$  меньше.

Рассмотрим случай  $x \in \mathbb{R}$ , то есть, независимая переменная всего одна. Пользуясь вышеупомянутыми соображениями (в частности, первым и вторым), введем следующее определение расстояния  $\rho(x, i)$  от точки  $(x_i, y_i)$  из обучающей выборки до некоторой точки  $(x, f(x, \boldsymbol{\omega}))$  на кривой, описываемой регрессионной моделью:

$$\rho(x, i) = \frac{(x_i - x)^2}{\sigma_{x_i}^2} + \frac{(y_i - y)^2}{\sigma_{y_i}^2}, \quad (3)$$

где  $y = f(x, \boldsymbol{\omega})$ .

Линеаризуем  $f(x, \boldsymbol{\omega})$  в окрестности точки  $x_i$ :

$$f(x, \boldsymbol{\omega}) = f(x_i, \boldsymbol{\omega}) + (x - x_i) \frac{\partial f}{\partial x}(x_i, \boldsymbol{\omega}). \quad (4)$$

Введем для удобства обозначение  $k = \frac{\partial f}{\partial x}(x_i, \boldsymbol{\omega})$ .

Тогда расстояние (3) можно выразить через линеаризованную функцию (4) следующим образом:

$$\rho(x, i) = \frac{(x_i - x)^2}{\sigma_{x_i}^2} + \frac{(y_i - f(x_i, \boldsymbol{\omega}) - k(x - x_i))^2}{\sigma_{y_i}^2}. \quad (5)$$

Минимизируем это расстояние (5) по  $x$ , для этого возьмем соответствующую производную и приравняем ее нулю:

$$2\frac{x_i - x}{\sigma_{x_i}^2} + 2k\frac{y_i - f(x_i, \boldsymbol{\omega}) - k(x - x_i)}{\sigma_{y_i}^2} = 0.$$

Отсюда путем несложных преобразований можно получить выражение на  $x - x_i$ :

$$x - x_i = \frac{\sigma_{x_i}^2 k(y_i - f(x_i, \boldsymbol{\omega}))}{\sigma_{y_i}^2 + k^2 \sigma_{x_i}^2}$$

Подставив это выражение в (5), можно получить выражение для расстояния:

$$\rho(x, i) = \frac{(y_i - f(x_i, \boldsymbol{\omega}))^2}{\sigma_{y_i}^2 + k^2 \sigma_{x_i}^2}.$$

Аналогичным образом можно получить выражение для случая, когда  $\mathbf{x}$  — вектор  $m$ -мерном пространстве  $\mathbb{R}^m$ :

$$\rho(\mathbf{x}, i) = \frac{(y_i - f(\mathbf{x}_i, \boldsymbol{\omega}))^2}{\sigma_{y_i}^2 + \sum_{j=1}^m (\frac{\partial f}{\partial x_j}(\mathbf{x}_i, \boldsymbol{\omega}))^2 \sigma_{x_{ij}}^2}.$$

Таким образом, функционал, минимизирующий сумму введенных согласно (3) расстояний, выглядит следующим образом:

$$S(\boldsymbol{\omega}) = \sum_{i=1}^{\ell} \frac{(y_i - f(\mathbf{x}_i, \boldsymbol{\omega}))^2}{\sigma_{y_i}^2 + \sum_{j=1}^m (\frac{\partial f}{\partial x_j}(\mathbf{x}_i, \boldsymbol{\omega}))^2 \sigma_{x_{ij}}^2}. \quad (6)$$

Отметим, что этот функционал соответствует классической сумме квадратов регрессионных остатков при условии нормировки квадрата каждого остатка на сумму квадратов погрешности определения зависимой величины  $\sigma_{y_i}$  и произведения частной производной регрессионной модели по  $j$ -ой компоненте вектора независимых величин на погрешность определения соответствующей компоненты  $\sigma_{x_{ij}}$ .

Кроме того, если все независимые переменные измерены точно, то есть,  $\forall i, j : \sigma_{x_{ij}} = 0$ , то предложенный функционал переходит в предложенный в [?]. Если же, кроме того, все зависимые переменные имеют одну и ту же погрешность, то предложенный функционал переходит в известную сумму квадратов регрессионных остатков (с точностью до некоторого множителя  $\sigma_y$ ).

## 4 Модифицированный алгоритм Левенберга-Марквардта

Для минимизации функционала (6) предлагается следующий итеративный алгоритм, предназначенный для использования с уже имеющимися реализациями АЛМ <sup>1</sup>:

1. Выбирается некоторое начальное приближение вектора параметров  $\boldsymbol{\omega}$ .

---

<sup>1</sup>Предполагается, что реализация АЛМ «принимает на вход» массив значений  $y_i$ , функцию вычисления значения  $f$  в точках  $\mathbf{x}_i$  с вектором параметров  $\boldsymbol{\omega}$ , и критерий останова в виде числа итераций. Примером такой реализации, использовавшейся авторами, может служить [?].

2. Для каждой пары  $(\mathbf{x}_i, y_i)$  из обучающей выборки рассчитывается значение частной производной  $\frac{\partial f}{\partial x}$  в точке  $(\mathbf{x}_i, \boldsymbol{\omega})$ .
3. Каждое значение зависимой переменной  $y_i$  и значение функции  $f(\mathbf{x}_i, \boldsymbol{\omega})$  нормируется на соответствующую величину

$$\sigma_{y_i}^2 + \sum_{j=1}^m \left( \frac{\partial f}{\partial x_j}(\mathbf{x}_i, \boldsymbol{\omega}) \right)^2 \sigma_{x_{ij}}^2.$$

4. Выполняется итерация АЛМ для таким образом модифицированных значений функции  $f$  и зависимых переменных  $y_i$ , таким образом получается новое значение вектора  $\boldsymbol{\omega}$ .
5. Если критерий останова мАЛМ не достигнут, алгоритм продолжает выполнение с пункта 2.

Отметим следующее:

- Критерием останова мАЛМ могут служить обычные критерии вроде достижения некоторого числа итераций, нормы изменения вектора  $\boldsymbol{\omega}$ , и т. п.
- Если известно, что производная  $\frac{\partial f}{\partial x}$  достаточно гладка в окрестности  $(\mathbf{x}_i, \boldsymbol{\omega}) \mid i \in \{1, \dots, \ell\}$ , на шаге 4 алгоритма мАЛМ представляется разумным выполнить сразу несколько итераций классического АЛМ во избежание потенциально ресурсоемкого пересчета производных и перенормировки значений  $y_i$  и  $f$ .

Сходимость предложенного алгоритма можно показать, сведя его к классическому АЛМ. Для этого вместо объектов  $(\mathbf{x}_i, y_i)$  в обучающей выборке будем формально рассматривать объекты  $(\tilde{\mathbf{x}}_i, \tilde{y}_i)$ , где  $\tilde{y}_i = 0$ , а  $\tilde{\mathbf{x}}_i$  — вектор  $\mathbf{x}_i$  с дополнительно приписанным к нему значением  $y_i$ . Примем

$$\tilde{f}(\tilde{\mathbf{x}}_i, \boldsymbol{\omega}) = \frac{f(\mathbf{x}_i, \boldsymbol{\omega}) - y_i}{\sigma_{y_i}^2 + \sum_{j=1}^m \left( \frac{\partial f}{\partial x_j}(\mathbf{x}_i, \boldsymbol{\omega}) \right)^2 \sigma_{x_{ij}}^2}.$$

Тогда минимизация функционала (6) возможна средствами классического АЛМ, так как легко видеть, что (6) в этом случае эквивалентен

$$S(\boldsymbol{\omega}) = \sum_{i=1}^{\ell} (\tilde{y}_i - \tilde{f}(\tilde{\mathbf{x}}_i, \boldsymbol{\omega}))^2.$$

Кроме того, при возможности выполнить аналитическое дифференцирование функции  $f$ , этот метод показывает еще один способ практической минимизации функционала (6) без постоянной корректировки значений  $y_i$  и  $f$ , как в предложенном выше алгоритме.

## Список литературы

- [1] Marquardt, D. W.: *An algorithm for least-squares estimation of non-linear parameters*. Journal of the Society of Industrial and Applied Mathematics, 11(2):431–441, 1963.
- [2] King, Davis E.: *Dlib-ml: A Machine Learning Toolkit*. Journal of Machine Learning Research, 10:1755–1758, 2009.