Глава 1

Интерполяция

1.1 Интерполяция кубическими сплайнами

Имеется $n+1$ равноотстоящих вдоль оси x точек с координатами $y_0, y_1, y_2, \dots, y_n$. Нужно провести через них гладкую кривую без разрыва не только первых, но и вторых производных.

Так как по оси x интервалы равные, изменением мрср они могут быть сделаны равными единице – на каждом участке x изменяется от 0 до 1.

Полином определяется четырьмя константами:

$$f_i(x) = a_i + b_i x + c_i x^2 + d_i x^3;$$

$$f_i'(x) = b_i + 2c_i x + 3d_i x^2;$$

$$f_i''(x) = 2c_i + 6d_i x.$$

При этом, подставляя $x = 0$ и $x = 1$ (левая и правая границы участка), получаем:

$$y_{i-1} = a_i; \quad y_i = a_i + b_i + c_i + d_i; \quad ;$$

Из этих выражений коэффициенты $i+1$-го участка выражаются через коэффициенты i-го:

$$b_{i+1} = b_i + 2c_i + 3d_i;$$

$$c_{i+1} = c_i + 3d_i.$$

Коэффициент d_{i+1} выражается через y_{i+2}:

$$y_{i+2} = y_{i+1} + b_{i+1} + c_{i+1} + d_{i+1},$$

откуда

$$d_{i+1} = y_{i+2} - y_{i+1} - b_{i+1} - c_{i+1} = y_{i+2} - y_{i+1} - b_i - 3c_i - 6d_i.$$

Таким образом, все коэффициенты $i+1$-го участка выражаются через коэффициенты i-го.

1.1.1 Уравнения на концах

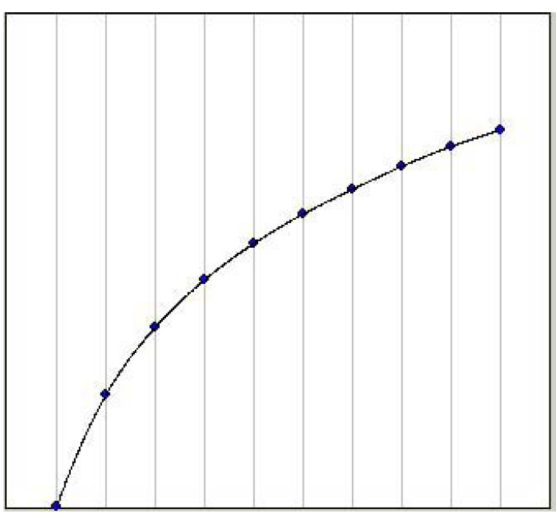
Для определения коэффициентов на первом и последнем участках нам не хватает данных (y_{-1} и y_{n+1}), поэтому на этих участках нужно ограничиться квадратичной параболой. Поэтому на самом левом участке полагаем $d_1 = 0$. на левом участке $y = y_0 + a_1 x + b_1 x^2$, причем при $x = 1 \quad y = y_1$, откуда $b_1 = y_1 - y_0 - a_1$ и неопределен всего один коэффициент a. для того, чтобы добиться $d_n = 0$, воспользуемся линейностью связи коэффициентов последующего интервала с коэффициентами предыдущего. Коэффициент d_n линейно зависит от a_1:

$$d_n = A + Ba_1,$$

где A и B пока неизвестные константы. Из этого соотношения видно, что d_n обращается в нуль при $a_1 = -A/B$. Чтобы найти A и B, положим сначала $a_1 = 0$ и найдем $d_n = d_n^{(0)}$, значение которого и равно константе A. Теперь положим $a_1 = 1$ и снова вычислим $d_n = d_n^{(1)}$. Тогда $B = d_n^{(1)} - d_n^{(0)}$, после чего находим на левом конце

$$a_1 = \frac{d_n^{(0)}}{d_n^{(0)} - d_n^{(1)}}.$$

Теперь на левом конце все коэффициенты определены. Сплайны такого типа называются *сплайнами со свободными границами*.



Это интерполяция функции $Ln(x)$ по значениям в целочисленных точках 1, 2, ... 10. При правильной работе с точками на концах интерполяция проходит без выбросов.

1.2 Регрессия сплайнами

Нужно иметь массив значений функции $(\bar{y}_i)$ при смещении по x с постоянным шагом.

Сплайновая регрессия может быть получена из вариационного принципа. Кривизна кривой при малых отклонениях ее от оси x приблизительно равна второй производной. Минимизируя интеграл по x от квадрата второй производной при задании точного прохождения кривой через заданных точки, получается кривая минимальной кривизны, проходящая через заданные точки. Однако, при регрессии мы можем заменить условие точного прохождения кривой через узловые точки условием одновременной с минимумом критичны минимизацией суммы квадратов отклонений кривой в узлах от заданных значений:

$$Q = \frac{1}{2}\int_0^l (y'')^2 dx + \frac{\lambda}{2}\sum_0^n (y_i - \bar{y}_i)^2,$$

где константр λ определяет

$$y'''' + \lambda \sum_0^n (y_i - \bar{y}_i)\delta(x - x_i) = 0.$$

Интегрируя это уравнение в окрестности i-й границы, получим связь между скачком третьей производной в узле и смещением кривой по вертикали от заданного узла:

$$y'''_{i\leftarrow} - y'''_{\rightarrow i} = \lambda(\bar{y}_i - y_i). \tag{1.1}$$

На интервалах между заданными точками кривая подчиняется уравнению $y'''' = 0$, то есть является кубическим полиномом, как и в случае сплайновой интерполяции:

$$y_k(x) = a_k + b_k x + c_k x^2 + d_k x^3;$$

$$y'_k(x) = b_k + 2c_k x + 3d_k x^2;$$

$$y''_k(x) = 2c_k + 6d_k x; \quad y'''(x) = 6d.$$

Сшивание вплоть до вторых производных приводит к связи коэффициентов на $(k-1)$ - м и k - м участках:

$$a_k = a_{k-1} + b_{k-1} + c_{k-1} + d_{k-1};$$

$$b_k = b_{k-1} + 2c_{k-1} + 3d_{k-1};$$

$$c_k = c_{k-1} + 3d_{k-1}.$$

А из уравнения (1.1) следует связь и на коэффициенты d_k:

$$d_k = (1 + \lambda)d_{k-1} + \lambda(a_{k-1} + b_{k-1} + c_{k-1}). \tag{1.2}$$

Эти четыре соотношения, как и в случае сплайновой интерполяции определяют коэффициенты полинома на k - м участке через коэффициенты на $(k-1)$ - м и заданное значение функции на их границе. Интегрируя это уравнение в окрестности i-й границы, получим связь между скачком третьей производной в узле и смещением кривой по вертикали от заданного узла:

$$y'''_{i\leftarrow} - y'''_{\rightarrow i} = 6\lambda(\bar{y}_i - y_i). \tag{1.3}$$

На интервалах между заданными точками кривая подчиняется уравнению $y'''' = 0$, то есть является кубическим полиномом, как и в случае сплайновой интерполяции:

$$y_k(x) = a_k + b_k x + c_k x^2 + d_k x^3;$$

$$y'_k(x) = b_k + 2c_k x + 3d_k x^2;$$

$$y''_k(x) = 2c_k + 6d_k x; \quad y'''(x) = 6d.$$

Сшивание вплоть до вторых производных приводит к связи коэффициентов на $(k-1)$ - м и k - м участках:

$$\begin{aligned} a_k &= a_{k-1} + b_{k-1} + c_{k-1} + d_{k-1}; \qquad (1.4)\\ b_k &= b_{k-1} + 2c_{k-1} + 3d_{k-1};\\ c_k &= c_{k-1} + 3d_{k-1}. \end{aligned}$$

А из уравнения (1.3) следует связь и на коэффициенты d_k:

$$d_k = d_{k-1} + \lambda(\bar{y}_k - a_k). \tag{1.5}$$

Эти четыре соотношения, как и в случае сплайновой интерполяции определяют коэффициенты полинома на k - м участке через коэффициенты на $(k-1)$ - м и заданное значение функции на их границе.

1.2.1 Граничные участки

Как и при интерполяции сплайнами граничные участки (самый левый и самый правый) нужно рассматривать отдельно.

$$\delta Q = \int y''\,\delta y''\,dx + 6\,\lambda \sum_{k=0}^{n} (y_k - \bar{y}_k)\delta y(x_k) =$$

$$\sum_{k=0}^{n} (y''\,\delta y' - y'''\,\delta y +$$

$$c_0 = 0; \quad c_n = 0; \quad d_0 + \lambda\,(a_0 - \bar{y}_0) = 0.$$

$$c_n = 0; \quad \delta = d_n + \lambda(a_n - \bar{y}_n) = 0$$

Из-за линейности соотношений (1.4-1.5), связывающих коэффициенты на соседних участках, коэффициенты на последнем участке линейно выражаются через коэффициенты на первом:

$$\delta = d_n + \lambda(a_n - \bar{y}_n) = A\,a_0 + B\,b_0 + K;$$

$$c \equiv c_n = C\,a_0 + D\,b_0 + L.$$

Вычислим их
при $a_0 = 0,\ b_0 = 0$ и обозначим рассогласования при этом

$$\delta_{00} = K; \quad c_{00} = L;$$

При $a_0 = 1,\ b_0 = 0$:

$$\delta_{10} = K + A; \quad c_{10} = C + L;$$

При $a_0 = 0,\ b_0 = 1$:

$$\delta_{01} = K + B; \quad c_{01} = D + L,$$

откуда

$$K = \delta_{00}; \quad A = \delta_{10} - \delta_{00}; \quad B = \delta_{01} - \delta_{00};$$

$$L = c_{00} \quad C = c_{10} - c_{00}; \quad D = c_{01} - c_{00}.$$

$$DD = A\,D - B\,C = (\delta_{10}c_{01} - \delta_{01}c_{10}) + \delta_{00}(c_{10} - c_{01}) + c_{00}(\delta_{01} - \delta_{10}).$$

$$a_0 = \frac{B\,L - D\,K}{A\,D - B\,C} = \frac{\delta_{01}c_{00} - c_{01}\delta_{00}}{DD};$$

$$b_0 = \frac{C\,K - A\,L}{A\,D - B\,C} = \frac{\delta_{00}c_{10} - c_{00}\delta_{10}}{DD}.$$