Метод наименьших квадратов (МНК) в матричной форме

Имеем таблицу значений $f_k = f(x_k)$, где $k = 0,1...m$.

Обобщенный полином имеет вид

$$P_n(x) = C_o\varphi_o(x) + C_1\varphi_1(x) + ... + C_n\varphi_n(x) = \sum_{k=0}^{n} C_k\varphi_k(x) \qquad (1)$$

Целевая функция

$$S(C_o,...C_n) = \sum_{k=0}^{m} \left[P_n(x_k) - f_k\right]^2 \qquad (2)$$

Введем векторы

$$C = \begin{pmatrix} C_o \\ ... \\ C_n \end{pmatrix}_{N\times1}, \quad f = \begin{pmatrix} f_o \\ ... \\ f_m \end{pmatrix}_{M\times1}, \varphi(x) = \begin{pmatrix} \varphi_0(x) \\ ... \\ \varphi_n(x) \end{pmatrix}_{N\times1} \qquad (3)$$

где $M = m+1,\ N = n+1$.

Рассмотрим матрицу

$$X = \begin{pmatrix} \varphi_0(x_0) & \varphi_1(x_0) & ... & \varphi_n(x_0) \\ \varphi_0(x_1) & \varphi_1(x_1) & ... & \varphi_n(x_1) \\ ... & ... & ... & ... \\ \varphi_0(x_m) & \varphi_1(x_m) & ... & \varphi_n(x_m) \end{pmatrix}_{M\times N} \qquad (4)$$

Тогда целевую функцию можно записать в виде

$$S = \left(X_{M\times N}C_{N\times1} - f_{M\times1}\right)^T \left(X_{M\times N}C_{N\times1} - f_{M\times1}\right) \qquad (5)$$

Проведем перемножение

$$S=\left(XC-f\right)^{T}\left(XC-f\right)=\left(C^{T}X^{T}-f^{T}\right)\left(XC-f\right)= \\ =C^{T}X^{T}XC-f^{T}XC-C^{T}X^{T}f+f^{T}f \tag{6}$$

Учитывая, что $f^{T}XC=\left(XC\right)^{T}f=C^{T}X^{T}f$, получим

$$S=C^{T}X^{T}XC-2C^{T}X^{T}f+f^{T}f \tag{7}$$

Дифференцирование по вектору дает

$$\frac{\partial S}{\partial C}=2X^{T}XC-2X^{T}f=0 \tag{8}$$

Таким образом, получаем СЛАУ

$$A_{N\times N}C_{N\times 1}=B_{N\times 1} \tag{9}$$

где $A_{N\times N}=X_{N\times M}^{T}X_{M\times N}$, $B_{N\times 1}=X_{N\times M}^{T}f_{M\times 1}$.

Пояснения – дифференцирование по вектору ($N=2$)

$$C^{T}AC=\begin{pmatrix}C_0 & C_1\end{pmatrix}\begin{pmatrix}A_{11} & A_{12}\\ A_{12} & A_{11}\end{pmatrix}\begin{pmatrix}C_0\\ C_1\end{pmatrix}=$$

$$=\begin{pmatrix}C_0 & C_1\end{pmatrix}\begin{pmatrix}C_0A_{11}+A_{12}C_1\\ C_0A_{12}+A_{22}C_1\end{pmatrix}=C_0^2A_{11}+2A_{12}C_0C_1+C_1^2A_{22}$$

Дифференцирование

$$\frac{\partial}{\partial C_0}\left(C_0^2A_{11}+2A_{12}C_0C_1+C_1^2A_{22}\right)=2C_0A_{11}+2A_{12}C_1$$

$$\frac{\partial}{\partial C_1}\left(C_0^2A_{11}+2A_{12}C_0C_1+C_1^2A_{22}\right)=2C_0A_{12}+2A_{22}C_1$$

Таким образом $\frac{\partial}{\partial C}\left(C^{T}AC\right)=2AC$.

Соотношение $\frac{\partial}{\partial C}\left(C^{T}X^{T}f\right)=\frac{\partial}{\partial C}\left(C^{T}B\right)=B$ доказывается аналогично.

Итерационный метод наименьших квадратов (МНК)

Частный случай

Целевая функция

$$S_m(C)=\sum_{k=0}^{m}\left[C\varphi\left(x_k\right)-f_k\right]^2 \tag{10}$$

Дифференцируем

$$\frac{\partial S_m}{\partial C}=2\sum_{k=0}^{m}\left[C\varphi\left(x_k\right)-f_k\right]\varphi\left(x_k\right)=0 \tag{11}$$

Тогда

$$C^{(m)}=\frac{\sum\limits_{k=0}^{m} f_k\varphi\left(x_k\right)}{\sum\limits_{k=0}^{m}\varphi^2\left(x_k\right)} \tag{12}$$

Пусть в таблице имеем одну точку

$$C^{(0)}=\frac{f_0\varphi\left(x_0\right)}{\varphi^2\left(x_0\right)}=\frac{f_0}{\varphi\left(x_0\right)} \tag{13}$$

Прибавилась еще одна точка, тогда

$$C^{(1)}=\frac{f_0\varphi\left(x_0\right)+f_1\varphi\left(x_1\right)}{\varphi^2\left(x_0\right)+\varphi^2\left(x_1\right)} \tag{14}$$

Уравнение (14) перепишем в виде

$$C^{(1)}=C^{(0)}-C^{(0)}\frac{\varphi^2\left(x_0\right)+\varphi^2\left(x_1\right)}{\varphi^2\left(x_0\right)+\varphi^2\left(x_1\right)}+\frac{f_0\varphi\left(x_0\right)+f_1\varphi\left(x_1\right)}{\varphi^2\left(x_0\right)+\varphi^2\left(x_1\right)} \tag{15}$$

Или

$$C^{(1)} = C^{(0)} + \frac{\varphi(x_0)\Big(f_0 - C^{(0)}\varphi(x_0)\Big) + \varphi(x_1)\Big(f_1 - C^{(0)}\varphi(x_1)\Big)}{\varphi^2(x_0) + \varphi^2(x_1)} \qquad (16)$$

Более просто. Подставляя (13) в (16), получим

$$C^{(1)} = C^{(0)} + \frac{\varphi(x_0)\left(f_0 - \frac{f_0}{\varphi(x_0)}\varphi(x_0)\right) + \varphi(x_1)\Big(f_1 - C^{(0)}\varphi(x_1)\Big)}{\varphi^2(x_0) + \varphi^2(x_1)} = \qquad (17)$$

$$= C^{(0)} + \frac{\varphi(x_1)\Big(f_1 - C^{(0)}\varphi(x_1)\Big)}{\varphi^2(x_0) + \varphi^2(x_1)}$$

Или в общем случае

$$C^{(m+1)} = C^{(m)} + p_m\varphi(x_{m+1})\Big(f_{m+1} - C^{(m)}\varphi(x_{m+1})\Big) \qquad (18)$$

где $p_m = \dfrac{1}{\sum\limits_{k=0}^{m}\varphi^2(x_k)}$ и сумма $\sum\limits_{k=0}^{m}\varphi^2(x_k)$ естественно накапливается, то есть

$$\sum_{k=0}^{m}\varphi^2(x_k) = \varphi^2(x_m) + \sum_{k=0}^{m-1}\varphi^2(x_k).$$

Общий случай

Целевая функция

$$S_m = \sum_{k=0}^{m}\Big[C^T\varphi(x_k) - f_k\Big]^2 \qquad (19)$$

где

$$C = \begin{pmatrix} C_o \\ \dots \\ C_n \end{pmatrix}_{N\times 1}, \quad \varphi(x_k) = \begin{pmatrix} \varphi_0(x_k) \\ \dots \\ \varphi_n(x_k) \end{pmatrix}_{N\times 1}$$

Дифференцируя по вектору, получаем

$$\frac{\partial S_m}{\partial C} = 0 \Rightarrow A_m C^{(m)} = B_m \tag{20}$$

где

$$A_m = \sum_{k=0}^{m} \varphi(x_k)\varphi^T(x_k),\ B_m = \sum_{k=0}^{m} f_k \varphi(x_k)$$

Введем обозначение $Q_m^{-1} = A_m$, поэтому

$$C^{(m)} = Q_m B_m \tag{21}$$

Система СЛАУ (20) может быть записана в виде

$$A_m C^{(m)} = f_m \varphi(x_m) + \sum_{k=0}^{m-1} f_k \varphi(x_k) \tag{22}$$

Из СЛАУ на предыдущей итерации имеем $A_{m-1} C^{(m-1)} = B_{m-1}$ или

$$\sum_{k=0}^{m-1} f_k \varphi(x_k) = \sum_{k=0}^{m-1} \varphi(x_k)\varphi^T(x_k) C^{(m-1)} \tag{23}$$

Тогда, подставляя (23) в (22), получим

$$Q_m^{-1} C^{(m)} = f_m \varphi(x_m) + \sum_{k=0}^{m-1} \varphi(x_k)\varphi^T(x_k) C^{(m-1)} \tag{24}$$

Прибавляя и вычитая $\varphi(x_m)\varphi^T(x_m) C^{(m-1)}$ в (24), найдем

$$\begin{aligned} Q_m^{-1} C^{(m)} &= f_m \varphi(x_m) - \varphi(x_m)\varphi^T(x_m) C^{(m-1)} + \\ &+ \varphi(x_m)\varphi^T(x_m) C^{(m-1)} + \sum_{k=0}^{m-1} \varphi(x_k)\varphi^T(x_k) C^{(m-1)} = \\ &= \varphi(x_m)\left(f_m - \varphi^T(x_m) C^{(m-1)}\right) + \sum_{k=0}^{m} \varphi(x_k)\varphi^T(x_k) C^{(m-1)} \end{aligned} \tag{25}$$

Или

$$Q_m^{-1}C^{(m)} = \varphi(x_m)\left(f_m - \varphi^T(x_m)C^{(m-1)}\right) + Q_m^{-1}C^{(m-1)} \tag{26}$$

Отсюда формально

$$C^{(m)} = C^{(m-1)} + Q_m\varphi(x_m)\left(f_m - \varphi^T(x_m)C^{(m-1)}\right) \tag{27}$$

Уравнение (26) — это итерационная формула для определения $C^{(m)}$. Однако для этого матрица Q_m также должна быть получена итерационно. Имеем

$$Q_m^{-1} = A_m = \sum_{k=0}^{m}\varphi(x_k)\varphi^T(x_k) = Q_{m-1}^{-1} + \varphi(x_m)\varphi^T(x_m) \tag{28}$$

Умножая (28) слева на Q_m, получим

$$E = Q_m Q_{m-1}^{-1} + Q_m\varphi(x_m)\varphi^T(x_m) \tag{29}$$

где E - единичная матрица.

Теперь, умножая (29) на Q_{m-1} справа, найдем

$$Q_{m-1} = Q_m + Q_m\varphi(x_m)\varphi^T(x_m)Q_{m-1} \tag{30}$$

Если умножить уравнение (30) на $\varphi(x_m)$, то получим

$$\begin{aligned} Q_{m-1}\varphi(x_m) &= Q_m\varphi(x_m) + Q_m\varphi(x_m)\varphi^T(x_m)Q_{m-1}\varphi(x_m) = \\ &= Q_m\varphi(x_m)\left(1 + \varphi^T(x_m)Q_{m-1}\varphi(x_m)\right) \end{aligned} \tag{31}$$

Здесь 1 – это не единичная матрица, так как $\varphi^T(x_m)_{1\times N}\left(Q_{m-1}\right)_{N\times N}\varphi(x_m)_{N\times 1}$ есть скаляр.

Теперь умножим (31) на $\left(1 + \varphi^T(x_m)Q_{m-1}\varphi(x_m)\right)^{-1}\varphi^T(x_m)Q_{m-1}$ справа, что дает

$$Q_{m-1}\varphi(x_m)\left(1+\varphi^T(x_m)Q_{m-1}\varphi(x_m)\right)^{-1}\varphi^T(x_m)Q_{m-1}=$$
$$=Q_m\varphi(x_m)\left(1+\varphi^T(x_m)Q_{m-1}\varphi(x_m)\right)\left(1+\varphi^T(x_m)Q_{m-1}\varphi(x_m)\right)^{-1}\varphi^T(x_m)Q_{m-1}=$$
$$=Q_m\varphi(x_m)\varphi^T(x_m)Q_{m-1}$$

(32)

Из выражения (30) следует $Q_m\varphi(x_m)\varphi^T(x_m)Q_{m-1}=Q_{m-1}-Q_m$, тогда

$$Q_{m-1}\varphi(x_m)\left(1+\varphi^T(x_m)Q_{m-1}\varphi(x_m)\right)^{-1}\varphi^T(x_m)Q_{m-1}=$$
$$=Q_m\varphi(x_m)\varphi^T(x_m)Q_{m-1}=Q_{m-1}-Q_m \tag{33}$$

Откуда

$$Q_m=Q_{m-1}-Q_{m-1}\varphi(x_m)\left(1+\varphi^T(x_m)Q_{m-1}\varphi(x_m)\right)^{-1}\varphi^T(x_m)Q_{m-1} \tag{34}$$

Формула (34) не требует обращения матрицы, так как $1+\varphi^T(x_m)Q_{m-1}\varphi(x_m)$ - скаляр.

Окончательно, итерационная процедура сводится в применении двух формул (34) и (27), то есть

$$C^{(m)}=C^{(m-1)}+Q_m\varphi(x_m)\left(f_m-\varphi^T(x_m)C^{(m-1)}\right)$$

При проведении итераций полагается, что при $m=0$ имеем $C^{(-1)}=0$. А для уравнения (34) в соответствии практикой применения данной итерационной процедуры можно положить $Q_{-1}=E/\varepsilon$, где $\varepsilon<<1$. Например, $\varepsilon=0.01\div0.1$. В этом случае результат уже после нескольких итерации слабо зависит от Q_{-1} и быстро сходится к результату МНК.