

Punktowa aproksymacja średniokwadratowa

Jeśli węzły $\{(x_i, y_i)\}$ użyte do interpolacji w poprzednich zadaniach są wynikami rzeczywistego eksperymentu fizycznego i w konsekwencji obarczone niepewnością pomiaru $\{(x_i \pm \Delta x_i, y_i \pm \Delta y_i)\}$ to prowadzenie przez nie wielomianu interpolacyjnego zaczyna tracić sens. Zbyt duża liczba węzłów prowadzi też do konieczności konstruowania wielomianu bardzo wysokich rzędów (co próbowaliśmy ominąć budując „sklejki”).

Dlatego wygodniej jest *przybliżyć* (aproksymować) funkcję $y(x)$ funkcją $F(x)$ minimalizując pewną odległość między funkcją aproksymowaną $y(x)$ a aproksymującą $F(x)$ (normę $\|F - y\|$). Aproksymacja punktowa funkcji zadanej węzłami (x_i, y_i) polega na wyznaczeniu współczynników c_i takich, by

$$F(x) = \sum_{i=0}^m c_i \varphi_i(x), \quad (1)$$

gdzie funkcje $\varphi_i(x)$ stanowią bazę. Bazę mogą stanowić funkcje trygonometryczne $(1, \sin x, \cos x, \sin 2x, \cos 2x, \dots, \sin kx, \cos kx)$ czy np. jednomiany stopnia nie większego niż k : $(1, x^2, x^3, x^4, \dots, x^k)$. W tym ostatnim przypadku mówimy o aproksymacji wielomianowej.

Jeśli normę $\|\cdot\|$ mamy zadaną przepisem

$$\|F(x) - y(x)\|^2 = \sum_{i=1}^N [F(x_i) - y_i]^2 = \chi^2 \quad (2)$$

mówimy o aproksymacji *średniokwadratowej* (metodą najmniejszych kwadratów). Wówczas jeśli funkcja aproksymująca $F(x)$ dana jest równaniem (1) to minimum wyrażenia (2) odpowiada układowi $(m+1)$ równań liniowych z $(m+1)$ niewiadomymi c_j dla $j = 0, \dots, m$

$$2 \sum_{i=1}^N \left[\sum_{j=0}^m c_j \varphi_j(x_i) - y_i \right] \varphi_k(x_i) = 0, \quad (3)$$

gdzie $k = 0, \dots, m$, który można rozwiązać jedną z poznanych wcześniej metod.

W bibliotece *Numerical Recipes* procedura `fit`:

```
SUBROUTINE fit(x,y,ndata,sig,mwt,a,b,siga,sigb,chi2,q)
  INTEGER mwt,ndata
  REAL a,b,chi2,q,siga,sigb,sig(ndata),x(ndata),y(ndata)
  ...
```

dokonyje regresji liniowej `ndata` punktów (x,y) . Kładąc `mwt=0` zakładamy, że *nie* podano niepewności wartości y . Procedura zwraca współczynniki fitu `a`, `b` i ich niepewności `siga`, `sigb`, wartość χ^2 — `chi2` oraz ilościową informację o jakości dopasowania `q`.

Biblioteka *GNU Scientific Library (GSL)* zawiera funkcję

```
int gsl_fit_linear (const double * x, const size_t xstride, const double * y,
                  const size_t ystride, size_t n, double * c0, double * c1,
                  double * cov00, double * cov01, double * cov11, double * sumsq)
```

wyznaczającą współczynniki `c0` i `c1` prostej $y = c_0 + c_1 x$.

- Zbiory danych dostarczane są w `x` i `y` (tablicach o rozmiarze `n`, przetwarzanych z krokiem odpowiednio `xstride` i `ystride`).
- Do `cov00`, `cov01` i `cov11` wstawiane są odpowiednie elementy macierzy kowariancji – elementy diagonalne określają wariancje wielkości `c0` i `c1`, a ich pierwiastki kwadratowe są błędami tych wielkości.
- Wszystkie dane uwzględniane są z jednakowymi wagami, a wartość χ^2 umieszczana jest w `sumsq`.

Aproksymacja punktów branych z różnymi wagami wymaga użycia funkcji `gsl_fit_wlinear`.

Z kolei fitowanie prostą $y = cx$ umożliwiają funkcje `gsl_fit_mul` i `gsl_fit_wmul`.