§7. Сглаживание и фильтрация

Задачами, непосредственно связанными со спектральным анализом, являются проблемы *сглаживания* и *фильтрации* данных. Они заключаются в построении для исходных экспериментальных точек $y(x_i)$ некоторой (непрерывной или дискретной) зависимости $f(x)$, приближающей определенным образом эти точки. При этом, необходимо учитывать, что данные (x_i,y_i) получены с некоторой погрешностью, выражающей шумовую компоненту измерений. Как правило, функция $f(x)$ с помощью того или иного алгоритма уменьшает погрешность, присутствующую в данных (x_i,y_i). Такого типа задачи называют задачами *фильтрации*.

Наиболее часто целью фильтрации является подавление быстрых вариаций, которые чаще всего обусловлены шумом. В результате из быстроосциллирующей зависимости $y(x_i)$ получается другая, сглаженная зависимость, в которой доминирует более низкочастотная составляющая. В связи с этим, считают, что *сглаживание* является частным случаем более общей задачи фильтрации сигнала, которая связана с подавлением определенных диапазонов частот спектра.

Часто рассматривают и задачу, противоположную сглаживанию – устранение медленно меняющихся вариаций в целях исследования высокочастотной составляющей. В этом случае говорят о задаче *устранения тренда*. Иногда интерес представляют смешанные задачи выделения среднемасштабных вариаций путем подавления как более быстрых, так и более медленных вариаций. Одна из возможностей решения связана с применением полосовой фильтрации.

Регрессия

Наиболее простым и эффективным рецептом *сглаживания* можно считать регрессию различного вида. Следует иметь в виду, что регрессия часто уничтожает информативную составляющую данных, оставляя лишь наперед заданную исследователем зависимость. В связи с этим, регрессия очень эффективна, когда из каких-либо соображений заранее известен (или, по крайней мере, хорошо угадывается) закон распределения данных (x_i,y_i).

Задачи математической регрессии имеют смысл приближения выборки данных (x_i,y_i) некоторой функцией f(x), определенным образом минимизирующей совокупность ошибок

$$|f(x_i)-y_i)|\sim\min. \tag{38}$$

Регрессия сводится к подбору неизвестных коэффициентов, определяющих аналитическую зависимость f(x).

Самый простой и наиболее часто используемый вид регрессии – линейная. Приближение данных (x_i,y_i) осуществляется линейной функцией

$$y(x)==b+ax. \tag{39}$$

На координатной плоскости (x,y) линейная функция, как известно, представляется прямой линией (рис. 71). Еще линейную регрессию часто называют *методом наименьших квадратов*, поскольку коэффициенты a и b вычисляются из условия минимизации суммы квадратов ошибок $|b+a\cdot(x_i-y_i)|^2$.

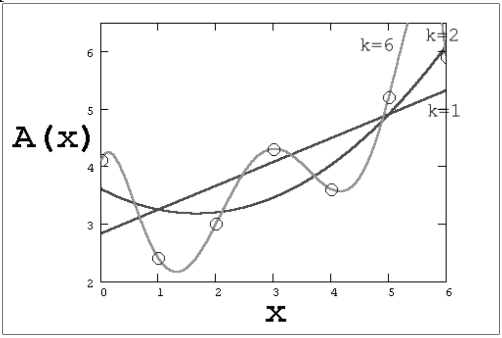

Рис. 71. *Полиномиальная регрессия*

Чаще всего такое же условие ставится и в других задачах регрессии: приближения массива данных (x_i,y_i) зависимостью y(x). В частности, полиномиальная регрессия означает приближение данных (x_i,y_i) полиномом k-й степени:

$$A(x)=a+b\cdot x+c\cdot x^2+d\cdot x^3+...+h\cdot x^k. \tag{40}$$

При $k=1$ полином является прямой линией, при $k=2$ – параболой, при $k=3$ – кубической параболой и т. д. (рис. 71). Как правило, на практике применяются $k<5$. Для построения регрессии полиномом k-й степени необходимо наличие, по крайней мере, $(k+1)$ точек данных.

Аналогично можно определить регрессию общего вида, следует только параметризовать искомую функцию и аккуратно поставить задачу минимизации.