Математические методы прогноза и восстановления зависимостей

Аппроксимация – замена одних объектов (функций) другими

Интерполяция – способ нахождения промежуточных значений величины по имеющемуся дискретному набору известных ее значений

Вид аппроксимирующей функции ?

Число и положение точек, в которых известно значение функции (узловых точек) ?

Полиномиальная аппроксимация

$$y = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n$$

Степень полинома на единицу меньше числа узловых точек

$$y_1 = a_0 + a_1 x_1 + a_2 x_1^2 + ... + a_n x_1^n$$

$$y_2 = a_0 + a_1 x_2 + a_2 x_2^2 + ... + a_n x_2^n$$

. . . .

$$y_n = a_0 + a_1 x_n + a_2 x_n^2 + ... + a_n x_n^n$$

Кусочно полиномиальная аппроксимация

Полиномы Лагранжа

$$y = \sum_{j=1}^{n+1} \left(y_j \frac{\prod\limits_{i=1 (i \neq j)}^{n+1} (x - x_i)}{\prod\limits_{i=1 (i \neq j)}^{n+1} (x_j - x_i)} \right)$$

$$y = y_1 \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)} + y_2 \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)} + y_3 \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)}$$

Многочлены Эрмита

$$y = \sum_{j=1}^{n+1} \left[H_j(x) y_j + h_j(x) y'_j \right]$$

$$h_j = \frac{(x - x_j) \prod\limits_{i=1 (i \neq j)}^{n+1} (x - x_i)^2}{\prod\limits_{i=1 (i \neq j)}^{n+1} (x_j - x_i)^2} \qquad H_j = \frac{(ax + b) \prod\limits_{i=1 (i \neq j)}^{n+1} (x - x_i)^2}{\prod\limits_{i=1 (i \neq j)}^{n+1} (x_j - x_i)^2}$$

$$a = -2 \sum_{i=1}^{n+1} \frac{1}{x_j - x_i} \qquad b = 1 - a x_j$$

Полиномы Ньютона

$$y = y_0 + (x - x_0)[x_1, x_0] + (x - x_0)(x - x_1)[x_2, x_1, x_0] + \dots$$
$$+ (x - x_0)(x - x_1)\dots(x - x_{n-1})[x_n, \dots x_n]$$

$$[x_j, x_i] = \frac{y_j - y_i}{x_j - x_i} \qquad [x_{i+1}, x_i, x_{i-1}] = \frac{[x_{i+1}, x_i] - [x_i, x_{i-1}]}{x_{i+1} - x_{i-1}}$$

Многочлены Чебышева

$$y = b_0 + b_1 T_1 + b_2 T_2 + \ldots + b_m T_m$$

$$T_0 = 1 \qquad T_1 = x \qquad T_{n+1} = 2xT_n - T_{n-1}$$

Аппроксимация рациональными функциями

$$y = \frac{N(x)}{D(x)}$$

Сплайн-интерполяция

Квадратичный сплайн

$$y = a_i + b_i(x - x_{i-1}) + c_i(x - x_{i-1})^2$$

$$x_{i-1} \le x \le x_i$$

$$a_i = y_{i-1} \qquad y_i = a_i + b_i(x_i - x_{i-1}) + c_i(x_i - x_{i-1})^2$$

$$y'(x_i - 0) = y'(x_i + 0)$$

$$y'(x_0) = 0$$

Кубический сплайн

$$y = a_i + b_i(x - x_{i-1}) + c_i(x - x_{i-1})^2 + d_i(x - x_{i-1})^3$$

$$x_{i-1} \le x \le x_i$$

$$a_i = y_{i-1} \qquad y_i = a_i + b_i(x_i - x_{i-1}) + c_i(x_i - x_{i-1})^2 + d_i(x_i - x_{i-1})^3$$

$$y'(x_{i-1} - 0) = y'(x_{i-1} + 0) \qquad y'(x_i - 0) = y'(x_i + 0)$$

$$y''(x_{i-1} - 0) = y''(x_{i-1} + 0) \qquad y''(x_i - 0) = y''(x_i + 0)$$

$$y''(x_0) = y''(x_n) = 0$$

$$a_i = y_{i-1} \quad (1)$$

$$a_i + b_i h + c_i h^2 + d_i h^3 = y_i \quad (2)$$

$$b_i + 2c_i h + 3d_i h^2 - b_{i+1} = 0 \quad (3)$$

$$c_i + 3d_i h - c_{i+1} = 0 \quad (4)$$

$$c_1 = c_{n+1} = 0$$

Метод прогонки

$$A_i c_{i-1} - D_i c_i + B_i c_{i+1} = -G_i$$

$$c_i = \alpha_{i+1} c_{i+1} + \beta_{i+1}$$

$$\alpha_{i+1} = \frac{B_i}{D_i - A_i \alpha_i} \qquad \beta_{i+1} = \frac{G_i - A_i \beta_i}{D_i - A_i \alpha_i}$$

Интерполяция с использованием базисного сплайна

$$y_i = \sum_{i=1}^{n} s_i B_i$$

$$x_{i-2} \le x \le x_{i-1} \qquad B_i = \frac{1}{4h^3}(x - x_{i-2})^3$$

$$x_{i-1} \le x \le x_i \qquad B_i = \frac{1}{4} + \frac{3}{4h}(x - x_{i-1}) + \frac{3}{4h^2}(x - x_{i-1})^2 - \frac{3}{4h^3}(x - x_{i-1})^3$$

$$x_i \le x \le x_{i+1} \qquad B_i = \frac{1}{4} + \frac{3}{4h}(x_{i+1} - x) + \frac{3}{4h^2}(x_{i+1} - x)^2 - \frac{3}{4h^3}(x_{i+1} - x)^3$$

$$x_{i+1} \le x \le x_{i+2} \qquad B_i = \frac{1}{4h^3}(x_{i+2} - x)^3$$

Аппроксимация Фурье

$$y = a_0/2 + \sum_{k=1}^{n-1} a_k \cos(\pi k x/n) + \sum_{k=1}^{n-1} b_k \sin(\pi k x/n) + a_n/2\cos(\pi x)$$

$$N = 2n \qquad\qquad h = 1$$

$$a_k = 1/n \sum_{x=0}^{2n-1} y(x)\cos(\pi k x/n) \qquad b_k = 1/n \sum_{x=0}^{2n-1} y(x)\sin(\pi k x/n)$$

$$na_k = \left(\cos\frac{\pi k}{n}\right)U_{2n-1} - U_{2n-2} + f(0)$$

$$nb_k = \left(\sin\frac{\pi k}{n}\right)U_{2n-1}$$

$$U_0 = 0$$

$$U_1 = f(2n-1)$$

$$U_m = \left(2\cos\frac{\pi k}{n}\right)U_{m-1} - U_{m-2} + f(2n-m)$$

Использование усеченного ряда (M гармоник)

$$\sum_x (f - f_M)^2 = \sum_x f^2 - n\left[\frac{a_0^2}{2} + \sum_{k=1}^{M} (a_k^2 + b_k^2)\right]$$

Ряд по косинусам $\quad -N+1 \le x \le N-1$

$$a_k = \frac{2}{N}\left[\frac{f_0}{2} + \sum_{x=1}^{N-1} f(x)\cos\frac{\pi k}{N}x + \frac{(-1)^k}{2} f_N\right]$$

Медленное изменение по времени

$$-N+1 \le x \le N$$

$$a_m(t) = 1/N \sum_{x=-N+1}^{N} y(x+t)\cos(\pi m x / N)$$

$$b_m(t) = 1/N \sum_{x=-N+1}^{N} y(x+t)\sin(\pi m x / N)$$

Смешение частот

$$A_0 = a_0 + 2\sum_{j=1}^{\infty} a_{2jN}$$

$$A_k = a_k + \sum_{j=1}^{\infty} (a_{2(j-k)N} + a_{2(j+k)N})$$

$$B_k = b_k + \sum_{j=1}^{\infty} (-b_{2(j-k)N} + b_{2(j+k)N})$$

Сигма-множители Ланцоша.

- Прямоугольная волна с периодом 2π

$$y = 1/2 + 2/\pi \sum_{k=1}^{N} \frac{1}{2k-1} \sin((2k-1)x)$$

$$\overline{y} = \frac{N}{\pi} \int_{x-\pi/2N}^{x+\pi/2N} y(\tau)d\tau = \frac{1}{2} + \frac{2}{\pi} \sum_{k=1}^{N} \frac{1}{2k-1} \frac{\sin[(2k-1)\pi/2N]}{(2k-1)\pi/2N} \sin((2k-1)x)$$

Ряд экспонент

$$y = A_0 e^{a_0 x} + A_1 e^{a_1 x} + \ldots + A_{n-1} e^{a_{n-1} x}$$

Метод наименьших квадратов
(Регрессионный анализ)

$$L = \sum_{i=1}^{n} (y - y_i)^2$$

$$y = a_0 \qquad a_0 = \frac{1}{n} \sum_{1}^{n} y_i \quad a_0 = \frac{y_{\max} + y_{\min}}{2}$$

$$y = a_0 + a_1 x$$

$$a_0 = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \qquad a_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$y = a_0 + a_1 x + a_2 x^2$$

Ортогональные многочлены –
Лежандра, Лагерра, Эрмита, Чебышева

$$y = a_0 + a_1 F_1 + a_2 F_2 + \ldots$$

Полиномы Лежандра

$$P_0 = 1 \qquad P_1 = x \qquad P_2 = \frac{1}{2}(3x^2 - 1)$$

$$P_3 = \frac{1}{2}(5x^3 - 3x) \qquad P_4 = \frac{1}{8}(35x^4 - 30x^2 + 3)$$

$$y = a_0^{'} + a_1(x - \bar{x})$$

Оценка дисперсии

$$V(a_0^{'}) = \frac{\sigma^2}{n} \qquad V(a_1) = \frac{\sigma^2}{S_{xx}}$$

$$\sigma^2 = \frac{SS_{ош}}{n-2} \qquad SS_{ош} = \sum_j (y_j - \hat{y}_j)^2 \qquad S_{xx} = \sum_j (x_j - \bar{x})^2$$

Множественный коэффициент детерминации

$$R^2 = \frac{\sum_j (\hat{y}_j - \bar{y})^2}{\sum_j (y_j - \bar{y})^2}$$

Многомерная регрессия

$$y_i = \beta_0 + \beta_{1i}x_{1i} + \beta_{2i}x_{2i} + \ldots + \beta_{11i}x_{1i}^2 + \beta_{22i}x_{2i}^2 + \ldots + \beta_{12i}x_{1i}x_{2i} + \ldots$$

$$L = \sum_i (y_i - \beta_0 - \sum_j \beta_j x_{ij})^2 + \tau(\beta_0 + \sum_j \beta_j)^2$$

$$y_t = \beta \sum_i x_{it} \qquad y_t = \sum_i (\beta_i x_{it}) \qquad y_t = \sum_i (\beta x_{it} + f_i)$$

$$\beta_i = \gamma \sum_j z_j + \beta + \alpha_i$$

$$\Delta y_t = \beta \sum_i x_{it} + \rho(y_{t-1} - \gamma \sum_i x_{i,t-1})$$

Множество исходных данных разбивается на три множества: обучающее, проверочное и контрольное

$$y_{j\lambda} = a_{0j\lambda} + a_{1j\lambda}x_j + a_{2j\lambda}x_\lambda + a_{3j\lambda}x_j x_\lambda + a_{4jj}x_j^2 + a_{5\lambda\lambda}x_\lambda^2$$

$$z_{l\gamma} = a_{0l\gamma} + a_{1l\gamma}y_l + a_{2l\gamma}y_\gamma + a_{3l\gamma}y_l y_\gamma + a_{4ll}y_l^2 + y_{5\gamma\gamma}y_\gamma^2$$

$$y = \sum_i (a_{0i} + a_{1i}x_i + a_{2i}x_i^2)$$

λ

Минимальное число экспертов

$$n_{\min} = 0{,}5\left(\frac{3}{\varepsilon} + 5\right) \qquad \varepsilon = \left(\frac{\bar{x}_i - \bar{x}_i'}{\bar{x}_{i\max}}\right)_{\max}$$

Максимальное число экспертов $$n_{\max} \le \frac{3\sum_{i=1}^{n} K_i}{2K_{\max}}$$

Компетентность эксперта $$K = 0{,}5\left(\frac{\sum_{j=1}^{m} x_j}{\sum_{j=1}^{m} x_{j\max}} + \frac{\lambda}{P}\right)$$

λ - самооценка P – максимальная самооценка

Среднее взвешенное $$\bar{x}_i = \sum_{k=1}^{l} \sum_{j=1}^{n} q_k K_j x_{ij}^k$$

Вес $$W_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}}$$

$$W_i = \frac{\sum_{j=1}^{n} W_{ij}}{n}$$

Вариационный размах $$R_i = x_{i\max} - x_{i\min}$$

Среднеквадратичное отклонение $$\sigma_i = \sqrt{\frac{\sum_{j=1}^{n} \left(x_{ij} - \bar{x}_i\right)^2}{n-1}}$$

Коэффициент вариации $$V = \frac{\sigma}{\bar{x}_i}$$

Коэффициент ранговой корреляции Спирмена

$$\rho = 1 - \frac{6\sum_{i=1}^{m}\left(x_{ij} - x_{if}\right)^2}{m\left(m^2 - 1\right)}$$

Коэффициент конкордации

$$W_K = \frac{12\sum_{i=1}^{m}\left(\sum_{j=1}^{n} x_{ij} - \frac{1}{2}n\left(m+1\right)\right)^2}{n^2 m\left(m^2 - 1\right)}$$

Скользящее среднее

1) сглаживание
$$y_t = \frac{\sum_{i=t-p}^{i=t+p} d_i}{2p+1}$$

сглаживание с весом
$$y_t = \frac{\sum_{i=t-p}^{i=t+p} d_i f_i}{\sum_{i=t-p}^{i=t+p} f_i}$$

$$\sum_{i=t-p}^{i=t+p} f_i = 1$$

2) прогноз $$y_{i+1} = \frac{1}{p+1}\sum_{i=t-p}^{i=t} d_i$$

$$y_{t+1} = y_t + \frac{1}{p+1}(d_t - d_{t-p})$$

Экспоненциально взвешенное среднее

$$y_{t+1} = \alpha d_t + \alpha(1-\alpha)d_{t-1} + \alpha(1-\alpha)^2 d_{t-2} + \alpha(1-\alpha)^3 d_{t-2} + \ldots$$

$$y_{t+1} = \alpha d_t + (1-\alpha)y_{t-1}$$

Прогноз при наличии тренда

Аддитивный тренд

Метод Холта

$$y_{t+\tau} = u_t + b_t \tau$$

$$u_t = Ad_t + (1 - A)(u_{t-1} + b_{t-1})$$

$$b_t = B(u_t - u_{t-1}) + (1 - B)b_{t-1}$$

Модификация Муира

$$y_{t+\tau} = u_t + b_t[\tau + (1 - A)/A]$$

$$u_t = Ad_t + (1 - A)u_{t-1}$$

$$b_t = B(u_t - u_{t-1}) + (1 - B)b_{t-1}$$

Метод Бокса-Дженнингса

$$y_{t+1} = u_t$$

$$u_t = u_{t-1} + \gamma_{-1}(e_t - e_{t-1}) + \gamma_0 e_t + \gamma_1 \sum_{t}^{t-\infty} e_t$$

$$e_t = d_t - y_t$$

Двойное сглаживание Брауна

$$y_{t+\tau} = f_t + b_t \tau$$

$$f_t = 2u_t - \overline{u}_t$$

$$b_t = \frac{\alpha}{1-\alpha}(u_t - \overline{u}_t)$$

$$u_t = \alpha d_t + (1-\alpha)(u_{t-1} + b_{t-1})$$

$$\overline{u}_t = \alpha u_t + (1-\alpha)\overline{u}_{t-1}$$

Адаптивное сглаживание Брауна

$$y_{t+\tau} = f_t + b_t\tau$$

$$u_t = u_{t-1} + b_{t-1} + (1-\gamma^2)e_t$$

$$b_t = b_{t-1} + (1-\gamma)^2 e_t$$

$$e_t = d_t - y_t$$

Мультипликативный тренд

Метод Муира

$$y_{t+\tau} = v_t r_t^{\tau}$$

$$v_t = \alpha d_t + (1-\alpha) r_t v_{t-1}$$

$$r_t = \alpha \frac{d_t}{v_{t-1}} + (1-\alpha) r_{t-1}$$

Сезонномультипликативный тренд

Метод Холта - Винтера

$$f_{t+\tau} = (u_t + b_t\tau)F_{t-L+\tau}$$

$$u_t = A\frac{d_t}{F_t} + (1-A)(u_{t-1} + b_{t-1})$$

$$b_t = B(u_t - u_{t-1}) + (1-B)b_{t-1}$$

$$F_t = C\frac{d_t}{u_t} + (1-C)F_{t-L}$$

$$\frac{1}{n}\sum_{k=1}^{n} F_k = 1$$

$$f_{t+\tau} = u_t (1 + b_t)^{\tau} F_{t-L+\tau}$$

$$u_t = \frac{Ad_t}{F_{t+L}} + (1 - A)(1 + b_{t-1})u_{t-1}$$

$$b_t = \frac{B(u_t - u_{t-1})}{u_{t-1}} + (1 - B)b_{t-1}$$

$$F_t = \frac{Cd_t}{u_t} + (1 - C)F_{t-L}$$

Сглаживание ряда (скользящее среднее или медианное) по периоду

Выделение тренда и определение коэффициентов сезонности аддитивных или мультипликативных

Нормирование коэффициентов сезонности

Фиктивные переменные

$$y_t = a_0 + a_1 t + \sum_{i=1}^{n} c_i D_i$$

$$D_i = \begin{cases} 1 - событие\_произошло \\ \hline 0 - событие\_не\_произошло \end{cases}$$

Метод Кростона (предсказание редких событий)

$$u_t = Ad_t + (1 - A)u_{t-1}$$

$$L_t = B\Delta_t + (1 - B)L_{t-1}$$

Вероятность того, что суммарные продажи за n шагов на превысят z

$$F(z) = q^n + \sum_{m=1}^{n} C_n^m p^m q^{n-m} \int_0^z f^m(x)dx$$

Метод задержек

$$\left(\begin{bmatrix} x_p \\ x_{p-1} \\ \bullet \\ \bullet \\ x_1 \end{bmatrix} \begin{bmatrix} x_{p-1} \\ x_{p-2} \\ \bullet \\ \bullet \\ x_2 \end{bmatrix} \bullet \quad \bullet \quad \bullet \begin{bmatrix} x_N \\ x_{N-1} \\ \bullet \\ \bullet \\ x_{N-p+1} \end{bmatrix} \right)$$

$$x_{N+1} = a_0 + \sum_{i=1}^{p} a_i x_{N-p+1} + \sum_{j=1}^{p} \sum_{k \geq j}^{p} b_{jk} x_{N-p+j} x_{N-p+k}$$

Адаптивные методы

Метод Чоу $$y_{t+1} = \alpha d_t + (1-\alpha)y_{t-1}$$

α уточняется после каждого шага (выбором из трех близких значений)

Метод Тамара $$y_{t+\tau} = u_t + b_t\tau$$

$$u_t = Ad_t + (1-A)(u_{t-1} + b_{t-1})$$

$$b_t = B(u_t - u_{t-1}) + (1-B)b_{t-1}$$

$$A = 0{,}02\delta$$

Метод Тригга и Лича

$$f_{t+\tau} = \widetilde{u}_t = |T_t| d_t + (1 - |T_t|)\widetilde{u}_{t-1}$$

$$T_t = \frac{\overline{e}_t}{MAD_t}$$

$$\overline{e}_t = \alpha e_t + (1-\alpha)\overline{e}_{t-1}$$

$$e_t = d_t - y_t$$

$$MAD = \alpha |e_t| + (1-\alpha) MAD_{t-1}$$

Автокорреляционный анализ

$$r_k = \frac{\sum_{t=k+1}^{n}(d_t - \bar{d})(d_{t-k} - \bar{d})}{\frac{1}{n-1}\sum_{t=1}^{n}(d_t - \bar{d})^2}$$

$$\bar{d} = \frac{1}{n}\sum_{t=1}^{n} d_t$$

$$u_{t+1} = r_1 d_t + \theta e_t$$

Метод кумулятивных сумм

$$Q_k = \sum_{t=1}^{k} e_t$$

$$L_{1,2} = \pm\frac{a}{b}(t - n - b)$$

$$|L_{1,2}| \geq |Q| \Rightarrow u_t = u_t + \Delta u$$