Przybliżanie danych przez funkcje będące kawałkami wielomianami

Na podstawie rozdziałów 5.1 i 5.2 książki *The Elements of Statistical Learning* (Trevor Hastie, Robert Tibshirani, Jerome Friedman)

Mateusz Łyczko

Instytut Matematyczny UWr

26.05.2023

Funkcja regresji

Regresja liniowa, liniowa analiza dyskryminacyjna i regresja logistyczna opierają się na modelu liniowym. Zazwyczaj jednak funkcja regresji $f(X) = E(Y|X)$ jest nieliniowa. Oczywiście reprezentowanie $f(X)$ przez model liniowy jest zwykle wygodnym, a czasem nawet koniecznym przybliżeniem. Przedstawimy jednak inne podejście do tego problemu.

Nowy sposób

Główną ideą metody jest zastąpienie wektora wejść X zmiennymi, które są transformacjami X, a następnie wykorzystanie modeli liniowych dla takich zmiennych.
Oznaczmy przez $h_m(X) : \mathbb{R}^p \to \mathbb{R}$ m-tą transformację X, $m = 1, \dots, M$.
Modelujemy $f(X)$ jako

$$f(X) = \sum_{m=1}^{M} \beta_m h_m(X),$$

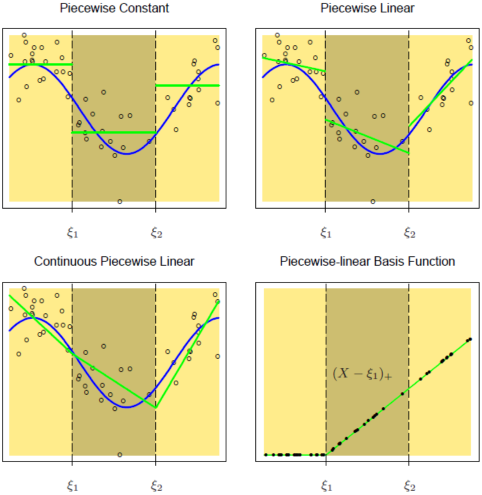
czyli liniowe rozwinięcie bazowe w X.

Funkcje bazowe

Niektóre proste i szeroko stosowane przykłady funkcji bazowych h_m są następujące:

- $h_m(X) = X_m, m = 1, \ldots, p$,
- $h_m(X) = X_j^2$ lub $h_m(X) = X_j X_k$,
- $h_m(X) = \log(X_j)$, $h_m(X) = \sqrt{X_j}$ lub $h_m(X) = ||X||$,
- $h_m(X) = I(L_m \leqslant X_k \leqslant U_m)$.

Funkcje kawałkami wielomianowe, rysunek 1



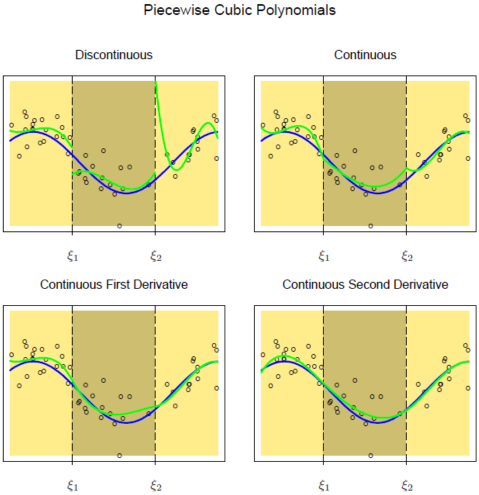
Omówienie rysunku 1

Funkcję $f(X)$ będącą kawałkami wielomianową uzyskuje się poprzez podział dziedziny X na sąsiadujące przedziały. Następnie reprezentujemy f przez oddzielny wielomian w każdym przedziale.
Poprzez podanie odpowiednich funkcji bazowych można znaleźć pewne funkcje.
Notacja: $X_+ = \max(0, X)$.

Omówienie rysunku 1, c.d.

- Pierwsze podejście jest takie, że używamy 3 bazowych funkcji:
$$h_1(X) = I(X < \xi_1), \quad h_2(X) = I(\xi_1 \leq X < \xi_2), \quad h_3(X) = I(\xi_2 \leq X).$$
Okazuje się, że w takim przypadku estymator OLS dla
$$f(X) = \sum_{m=1}^{3} \beta_m h_m(X)$$
to będzie $\widehat{\beta}_m = \bar{Y}_m$, czyli średnia w m-tym regionie. Zatem przybliżonie to jest kawałkami stałe.
- W drugim podejściu dodajemy następne 3 funkcje bazowe: $h_{m+3}(X) = h_m(X)X, m = 1, 2, 3$. Wtedy uzyskamy przybliżenie kawałkami liniowe.
- Zwykle chcemy, aby przybliżenie było ciągłe, dlatego w trzecim podejściu dodajemy jeszcze warunek $f(\xi_1^-) = f(\xi_1^+)$, co daje $\beta_1 + \xi_1\beta_4 = \beta_2 + \xi_1\beta_5$, podobnie robimy to dla ξ_2. Innym sposobem w tym przypadku jest użycie bazy (h_3 jest na ostatnim rysunku)
$$h_1(X) = 1, \quad h_2(X) = X, \quad h_3(X) = (X - \xi_1)_+, \quad h_4(X) = (X - \xi_2)_+$$

Funkcje kawałkami wielomianowe, rysunek 2



Omówienie rysunku 2

Okazuje się, że poniższa baza reprezentuje funkcję sklejaną 3. stopnia z 2 węzłami:

$$h_1(X) = 1, \quad h_3(X) = X^2, \quad h_5(X) = (X - \xi_1)_+^3,$$
$$h_2(X) = X, \quad h_4(X) = X^3, \quad h_6(X) = (X - \xi_2)_+^3.$$

Jest ona nazywana też funkcją sklejaną 4. rzędu.
Dla funkcji sklejanej rzędu M funkcje bazowe to

$$\begin{aligned} h_j(X) &= X^{j-1},\ j = 1, \ldots, M, \\ h_{M+\ell}(X) &= (X - \xi_\ell)_+^{M-1},\ \ell = 1, \ldots, K. \end{aligned}$$

Omówienie rysunku 2, c.d.

Ogólnie, funkcja sklejana rzędu M z węzłami ξ_j, $j = 1, \ldots, K$ jest kawałkami wielomianem stopnia $M - 1$ i ma ciągłe pochodne aż do rzędu $M - 2$.
Uważa się, że funkcje sklejane stopnia 3. (czyli rzędu 4.) są funkcjami sklejanymi najniższego stopnia, dla których te zmiany blisko węzłów nie są widoczne dla ludzkiego oka.
W ogólności, gdy stajemy przed problemem, należy wybrać rząd funkcji sklejanej, liczbę węzłów oraz ich rozmieszczenie.

Naturalna funkcja sklejana 3. stopnia

Dodaje ona dodatkowe ograniczenia do funkcji sklejanej 3. stopnia, a mianowicie, że funkcja jest liniowa poza węzłami ograniczającymi. Funkcja ta z K węzłami jest reprezentowana przez K funkcji bazowych:

$$N_1(X) = 1, \quad N_2(X) = X, \quad N_{k+2}(X) = d_k(X) - d_{K-1}(X),$$

gdzie

$$d_k(X) = \frac{(X - \xi_k)_+^3 - (X - \xi_K)_+^3}{\xi_K - \xi_k}.$$

Naturalna funkcja sklejana 3. stopnia VS Naturalna interpolacyjna funkcja sklejana 3. stopnia

Różnica między tymi funkcjami polega na tym, że ta druga przechodzi przez wszystkie wyznaczone punkty, a ta pierwsza niekoniecznie.

Bibliografia

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*