

Aproksymacja

Zadanie 1. Wykonaj aproksymację średniokwadratową punktową populacji Stanów Zjednoczonych w przedziale $[1900, 1980]$ wielomianami stopnia m dla $0 \leq m \leq 6$.

- (a) Dla każdego m dokonaj ekstrapolacji wielomianu do roku 1990. Porównaj otrzymaną wartość z prawdziwą wartością dla roku 1990, wynoszącą 248 709 873. Ile wynosi błąd względny ekstrapolacji dla roku 1990? Dla jakiego m błąd względny był najmniejszy?
- (b) Zbyt niski stopień wielomianu oznacza, że model nie jest w stanie uwzględnić zmienności danych (duże obciążenie). Zbyt wysoki stopień wielomianu oznacza z kolei, że model uwzględnia szum lub błędy danych (duża wariancja), co w szczególności obserwowaliśmy w przypadku interpolacji. Wielomian stopnia m posiada $k = m + 1$ parameterów. Stopień wielomianu, m , jest hiperparametrem modelu. Do wyboru optymalnego stopnia wielomianu można posłużyć się kryterium informacyjnym Akaikego (ang. Akaike information criterion):

$$\text{AIC} = 2k - k \ln \left(\frac{\sum_{i=1}^n [y_i - \hat{y}(x_i)]^2}{k} \right),$$

gdzie y_i ($i = 1, \dots, n$) oznacza prawdziwą liczbę osób w roku x_i , natomiast $\hat{y}(x_i)$ liczbę osób przewidywaną przez model, tzn wartość wielomianu $\hat{y}(x)$. Ponieważ rozmiar próbki jest niewielki (dane z dziewięciu lat, $n = 9$), $n/k < 40$, należy użyć wzoru ze składnikiem korygującym:

$$\text{AIC}_c = \text{AIC} + \frac{2k(k+1)}{n-k-1}.$$

Mniejsze wartości kryterium oznaczają lepszy model. Czy wyznaczony w ten sposób stopień m , odpowiadający najmniejszej wartości AIC_c , pokrywa się z wartością z poprzedniego podpunktu?

Zadanie 4. Wykonaj aproksymację średniokwadratową ciągłą funkcji $f(x) = \sqrt{x}$ w przedziale $[0, 2]$ wielomianem drugiego stopnia, używając wielomianów Czebyszewa. Aproksymacja ta jest tańszym obliczeniowo zamiennikiem aproksymacji jednostajnej.

Literatura

- [1] Xin-She Yang Introduction to Algorithms for Data Mining and Machine Learning, 2019.