

Statystyka i rachunek prawdopodobieństwa

Wykład 16: Regresje nieliniowa i wielowymiarowa

Mariusz Tarnopolski

Instytut Astronomii UMK

Statystyka ©2026



Na [wykładzie 14](#) omówiliśmy dopasowanie prostej do danych pomiarowych $\{x_i, y_i\}_{i=1}^n$ —tzw. regresję liniową. Pojęcie to jest jednak szersze i obejmuje sytuacje, gdy dopasowywana funkcja nie jest liniowa w swoich argumentach, ale jest liniowa w parametrach. Przykładem jest dopasowanie wielomianu stopnia m :

$$y(x) = a_0 + a_1x + a_2x^2 + \dots + a_mx^m \equiv \sum_{k=0}^m a_kx^k$$

który ma $m + 1$ parametrów.

Na [wykładzie 14](#) omówiliśmy dopasowanie prostej do danych pomiarowych $\{x_i, y_i\}_{i=1}^n$ —tzw. regresję liniową. Pojęcie to jest jednak szersze i obejmuje sytuacje, gdy dopasowywana funkcja nie jest liniowa w swoich argumentach, ale jest liniowa w parametrach. Przykładem jest dopasowanie wielomianu stopnia m :

$$y(x) = a_0 + a_1x + a_2x^2 + \dots + a_mx^m \equiv \sum_{k=0}^m a_kx^k$$

który ma $m + 1$ parametrów.

W ogólności, można rozważać dowolne funkcje postaci

$$y(x) = \sum_{k=0}^m a_k f_k(x)$$

gdzie dla wielomianów $f_k(x) = x^k$, ale równie dobrze $f_k(x) = \sin(kx)$ itp.—pod warunkiem, że $f_k(x)$ nie zależą od żadnego a_j .

Dopasowanie metodą najmniejszych kwadratów odbywa się poprzez minimalizację wyrażenia (por. slajd 14, [wykład 15](#))

$$S \equiv S(\mathbf{a}) \equiv \sum_{i=1}^n r_i^2(\mathbf{a}) = \sum_{i=1}^n \left[\frac{1}{\sigma_i} \left(y_i - \sum_{k=0}^m a_k f_k(x_i) \right) \right]^2 = \min.$$

Dopasowanie metodą najmniejszych kwadratów odbywa się poprzez minimalizację wyrażenia (por. slajd 14, [wykład 15](#))

$$S \equiv S(\mathbf{a}) \equiv \sum_{i=1}^n r_i^2(\mathbf{a}) = \sum_{i=1}^n \left[\frac{1}{\sigma_i} \left(y_i - \sum_{k=0}^m a_k f_k(x_i) \right) \right]^2 = \min.$$

Różniczkując po poszczególnych parametrach a_l :

$$\frac{\partial S}{\partial a_l} = -2 \sum_{i=1}^n \left[\frac{f_l(x_i)}{\sigma_i^2} \left(y_i - \sum_{k=0}^m a_k f_k(x_i) \right) \right] = 0$$

Dopasowanie metodą najmniejszych kwadratów odbywa się poprzez minimalizację wyrażenia (por. slajd 14, [wykład 15](#))

$$S \equiv S(\mathbf{a}) \equiv \sum_{i=1}^n r_i^2(\mathbf{a}) = \sum_{i=1}^n \left[\frac{1}{\sigma_i} \left(y_i - \sum_{k=0}^m a_k f_k(x_i) \right) \right]^2 = \min.$$

Różniczkując po poszczególnych parametrach a_l :

$$\frac{\partial S}{\partial a_l} = -2 \sum_{i=1}^n \left[\frac{f_l(x_i)}{\sigma_i^2} \left(y_i - \sum_{k=0}^m a_k f_k(x_i) \right) \right] = 0$$

Otrzymujemy tedy układ $m + 1$ równań liniowych na parametry a_l , gdzie $l = 0, \dots, m$:

$$\sum_{i=1}^n y_i \frac{f_l(x_i)}{\sigma_i^2} = \sum_{k=0}^m \left[a_k \sum_{i=1}^n \left(\frac{f_l(x_i) f_k(x_i)}{\sigma_i^2} \right) \right]$$

Najklarowniej jest powyższy układ równań zapisać w formie macierzowej:

$$\alpha \mathbf{a} = \beta \quad (1)$$

gdzie elementy wektora β to

$$\beta_l = \sum_{i=1}^n y_i \frac{f_l(x_i)}{\sigma_i^2}$$

elementy wektora \mathbf{a} to parametry a_k , zaś macierz kwadratowa α (nazywana macierzą krzywizny, bo $\frac{\partial^2 S}{\partial a_l \partial a_k} = 2\alpha_{lk}$) ma elementy

$$\alpha_{lk} = \sum_{i=1}^n \frac{f_l(x_i) f_k(x_i)}{\sigma_i^2}$$

Najklarowniej jest powyższy układ równań zapisać w formie macierzowej:

$$\alpha \mathbf{a} = \beta \quad (1)$$

gdzie elementy wektora β to

$$\beta_l = \sum_{i=1}^n y_i \frac{f_l(x_i)}{\sigma_i^2}$$

elementy wektora \mathbf{a} to parametry a_k , zaś macierz kwadratowa α (nazywana macierzą krzywizny, bo $\frac{\partial^2 S}{\partial a_l \partial a_k} = 2\alpha_{lk}$) ma elementy

$$\alpha_{lk} = \sum_{i=1}^n \frac{f_l(x_i) f_k(x_i)}{\sigma_i^2}$$

Rozwiązanie równ. (1) uzyskujemy mnożąc je lewostronnie przez α^{-1} .
Otrzymujemy

$$\mathbf{a} = \alpha^{-1} \beta$$

pod warunkiem, że α jest odwracalna, tj. $\det \alpha \neq 0$.

Wariancja $\sigma_{a_j}^2$ parametru a_j to suma wariancji σ_i^2 pomnożonych przez kwadrat wkładu, jaki i -ty punkt pomiarowy ma na parametr a_j (por. slajd 8, [wykład 15](#)). Podobnie, kowariancja parametrów a_j i a_l to

$$\text{Cov}(a_j, a_l) = \sum_{i=1}^n \sigma_i^2 \frac{\partial a_j}{\partial y_i} \frac{\partial a_l}{\partial y_i}$$

skąd wariancję $\sigma_{a_j}^2$ dostaniemy biorąc $j = l$.

Wariancja $\sigma_{a_j}^2$ parametru a_j to suma wariancji σ_i^2 pomnożonych przez kwadrat wkładu, jaki i -ty punkt pomiarowy ma na parametr a_j (por. slajd 8, [wykład 15](#)). Podobnie, kowariancja parametrów a_j i a_l to

$$\text{Cov}(a_j, a_l) = \sum_{i=1}^n \sigma_i^2 \frac{\partial a_j}{\partial y_i} \frac{\partial a_l}{\partial y_i}$$

skąd wariancję $\sigma_{a_j}^2$ dostaniemy biorąc $j = l$.

Pochodne cząstkowe to

$$\frac{\partial a_l}{\partial y_i} = \sum_{k=0}^m \alpha_{lk}^{-1} \frac{f_k(x_i)}{\sigma_i^2}$$

gdzie $\alpha_{lk}^{-1} \equiv (\alpha^{-1})_{lk}$ to lk -ty element macierzy α^{-1} .

Podstawiając do wyrażenia na $\text{Cov}(a_j, a_l)$:

$$\begin{aligned}
Cov(a_j, a_l) &= \sum_{i=1}^n \left[\sigma_i^2 \left(\sum_{k=0}^m \alpha_{jk}^{-1} \frac{f_k(x_i)}{\sigma_i^2} \right) \left(\sum_{p=0}^m \alpha_{lp}^{-1} \frac{f_p(x_i)}{\sigma_i^2} \right) \right] \\
&= \sum_{k=0}^m \left\{ \alpha_{jk}^{-1} \sum_{p=0}^m \left[\alpha_{lp}^{-1} \sum_{i=1}^n \left(\frac{f_p(x_i) f_k(x_i)}{\sigma_i^2} \right) \right] \right\} \\
&= \sum_{k=0}^m \alpha_{jk}^{-1} \sum_{p=0}^m \alpha_{lp}^{-1} \alpha_{pk} \\
&= \sum_{k=0}^m \left(\alpha_{jk}^{-1} \delta_{lk} \right) = \alpha_{jl}^{-1}
\end{aligned}$$

gdzie pozmienialiśmy kolejność sumowania oraz lk -ty element macierzy jednostkowej \mathbf{I} oznaczyliśmy symbolem Kroneckera $\delta_{lk} = \begin{cases} 1, & l = k \\ 0, & l \neq k \end{cases}$.

$$\begin{aligned}
Cov(a_j, a_l) &= \sum_{i=1}^n \left[\sigma_i^2 \left(\sum_{k=0}^m \alpha_{jk}^{-1} \frac{f_k(x_i)}{\sigma_i^2} \right) \left(\sum_{p=0}^m \alpha_{lp}^{-1} \frac{f_p(x_i)}{\sigma_i^2} \right) \right] \\
&= \sum_{k=0}^m \left\{ \alpha_{jk}^{-1} \sum_{p=0}^m \left[\alpha_{lp}^{-1} \sum_{i=1}^n \left(\frac{f_p(x_i) f_k(x_i)}{\sigma_i^2} \right) \right] \right\} \\
&= \sum_{k=0}^m \alpha_{jk}^{-1} \sum_{p=0}^m \alpha_{lp}^{-1} \alpha_{pk} \\
&= \sum_{k=0}^m \left(\alpha_{jk}^{-1} \delta_{lk} \right) = \alpha_{jl}^{-1}
\end{aligned}$$

gdzie pozmienialiśmy kolejność sumowania oraz lk -ty element macierzy jednostkowej \mathbf{I} oznaczyliśmy symbolem Kroneckera $\delta_{lk} = \begin{cases} 1, & l = k \\ 0, & l \neq k \end{cases}$.

Zatem macierz kowariancji parametrów a_j to α^{-1} .

Funkcje nieliniowe transformowalne w funkcje liniowe

Przykład. Żeby dopasować parametry a i b funkcji wykładniczej:

$$y(x) = ae^{-bx}$$

można ją obustronnie zlogarytmować:

$$\ln y = \ln a - bx$$

która jest liniowa, postaci

$$y' = a' - bx$$

gdzie $y' = \ln y$, $a' = \ln a$ (pod warunkiem, że $a > 0$ i $y > 0$).

Funkcje nieliniowe transformowalne w funkcje liniowe

Przykład. Żeby dopasować parametry a i b funkcji wykładniczej:

$$y(x) = ae^{-bx}$$

można ją obustronnie zlogarytmować:

$$\ln y = \ln a - bx$$

która jest liniowa, postaci

$$y' = a' - bx$$

gdzie $y' = \ln y$, $a' = \ln a$ (pod warunkiem, że $a > 0$ i $y > 0$).

Żeby uzyskać dopasowanie minimalizujemy wyrażenie

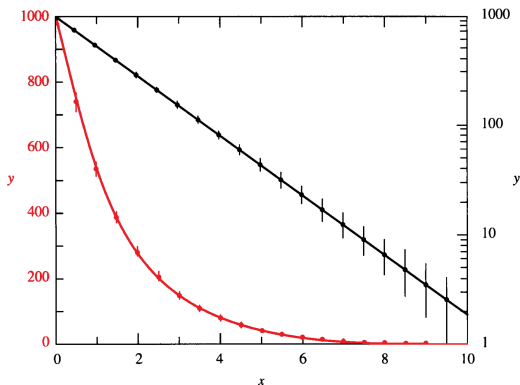
$$\sum_{i=1}^n \frac{(y'_i - a' + bx_i)^2}{\sigma_i'^2} = \sum_{i=1}^n \frac{(\ln y_i - \ln a + bx_i)^2}{\sigma_i'^2} = \min.$$

gdzie konieczne jest propagowanie błędów ([wykład 15](#)) wartości y_i na błędy wartości $\ln y_i$:

$$\sigma'_i = \left| \frac{dy'_i}{dy_i} \right| \sigma_i = \left| \frac{d \ln y_i}{dy_i} \right| \sigma_i = \frac{\sigma_i}{y_i}$$

Funkcje nieliniowe transformowalne w funkcje liniowe

To, jak istotna jest poprawna transformacja niepewności jest zilustrowane na poniższym przykładzie, gdzie $\sigma_i = \sqrt{y_i}$, tj. σ rosną wraz z y . W skali logarytmicznej jednak maleją wraz z y , co powoduje, że pominięcie (lub niepoprawne propagowanie) niepewności przy dopasowaniu funkcji $y' = a' - bx$ przeszacuje wpływ niepewności dla małych y , zakłamując ostateczny wynik.



Funkcje nieliniowe transformowalne w funkcje liniowe

Przykład. Model empiryczny tzw. log-parabola (widmo energetyczne):

$$y(E) = y_0 \left(\frac{E}{E_0} \right)^{-[\alpha + \beta \log(E/E_0)]}$$

Funkcje nieliniowe transformowalne w funkcje liniowe

Przykład. Model empiryczny tzw. log-parabola (widmo energetyczne):

$$y(E) = y_0 \left(\frac{E}{E_0} \right)^{-[\alpha + \beta \log(E/E_0)]}$$

Logarytmując obustronnie oraz przyjmując E_0 jako znane (parametr kalibracyjny; można go rozumieć jako wybór jednostek):

$$\log y = \log y_0 - [\alpha + \beta \log(E/E_0)] \log(E/E_0)$$

Funkcje nieliniowe transformowalne w funkcje liniowe

Przykład. Model empiryczny tzw. log-parabola (widmo energetyczne):

$$y(E) = y_0 \left(\frac{E}{E_0} \right)^{-[\alpha + \beta \log(E/E_0)]}$$

Logarytmując obustronnie oraz przyjmując E_0 jako znane (parametr kalibracyjny; można go rozumieć jako wybór jednostek):

$$\log y = \log y_0 - [\alpha + \beta \log(E/E_0)] \log(E/E_0)$$

zaś podstawiając $x = E/E_0$:

$$\log y = \log y_0 - (\alpha + \beta \log x) \log x$$

Funkcje nieliniowe transformowalne w funkcje liniowe

Przykład. Model empiryczny tzw. log-parabola (widmo energetyczne):

$$y(E) = y_0 \left(\frac{E}{E_0} \right)^{-[\alpha + \beta \log(E/E_0)]}$$

Logarytmując obustronnie oraz przyjmując E_0 jako znane (parametr kalibracyjny; można go rozumieć jako wybór jednostek):

$$\log y = \log y_0 - [\alpha + \beta \log(E/E_0)] \log(E/E_0)$$

zaś podstawiając $x = E/E_0$:

$$\log y = \log y_0 - (\alpha + \beta \log x) \log x$$

oraz podstawiając $y' = \log y$, $y'_0 = \log y_0$, $x' = \log x$:

$$y' = y'_0 - \alpha x' - \beta x'^2$$

Funkcje nieliniowe transformowalne w funkcje liniowe

Przykład. Model empiryczny tzw. log-parabola (widmo energetyczne):

$$y(E) = y_0 \left(\frac{E}{E_0} \right)^{-[\alpha + \beta \log(E/E_0)]}$$

Logarytmując obustronnie oraz przyjmując E_0 jako znane (parametr kalibracyjny; można go rozumieć jako wybór jednostek):

$$\log y = \log y_0 - [\alpha + \beta \log(E/E_0)] \log(E/E_0)$$

zaś podstawiając $x = E/E_0$:

$$\log y = \log y_0 - (\alpha + \beta \log x) \log x$$

oraz podstawiając $y' = \log y$, $y'_0 = \log y_0$, $x' = \log x$:

$$y' = y'_0 - \alpha x' - \beta x'^2$$

co jest de facto wielomianem stopnia $m = 2$. Parametry α i β dostajemy bezpośrednio z dopasowania, zaś $y_0 = 10^{y'_0}$.

Rozważmy zmienną losową

$$Y = \sum_{i=1}^k \beta_i x_i + \xi \quad (2)$$

gdzie czynniki x_i (zwane *zmiennymi objaśniającymi*) są nielosowe, β_i są stałymi współczynnikami (zwanymi *parametrami strukturalnymi modelu*), zaś ξ jest zmienną losową taką, że $\mathbb{E}(\xi) = 0$, $V(\xi) = \sigma^2$.

Rozważmy zmienną losową

$$Y = \sum_{i=1}^k \beta_i x_i + \xi \quad (2)$$

gdzie czynniki x_i (zwane *zmiennymi objaśniającymi*) są nielosowe, β_i są stałymi współczynnikami (zwanymi *parametrami strukturalnymi modelu*), zaś ξ jest zmienną losową taką, że $\mathbb{E}(\xi) = 0$, $V(\xi) = \sigma^2$.

Powyższy model składa się z sumy składników deterministycznego oraz stochastycznego (losowego), co odróżnia go od modelu regresji z [wykładu 14](#)—tam dopasowywano funkcje deterministyczne (np. $y = ax + b$); tutaj ξ zawiera w sobie cały (acz nieznan) wpływ fluktuacji statystycznych (poza wpływem czynników x_i) na zmienną losową Y .

Zakładamy też, że rozkład zmiennej losowej ξ nie zależy od czynników x_i .

Z równ. (2) wynika więc, że

$$\mathbb{E}(Y|x_1, x_2, \dots, x_k) = \sum_{i=1}^k \beta_i x_i$$

tzn. warunkowa wartość oczekiwana zmiennej losowej Y nie zależy od składnika stochastycznego oraz

$$\mathbb{E}\left[(Y - \mathbb{E}(Y|x_1, x_2, \dots, x_k))^2\right] = \mathbb{E}(\xi^2) = \sigma^2$$

czyli wariancja składnika stochastycznego jest miarą odchyień zmiennej losowej Y od jej warunkowej wartości oczekiwanej.

Z równ. (2) wynika więc, że

$$\mathbb{E}(Y|x_1, x_2, \dots, x_k) = \sum_{i=1}^k \beta_i x_i$$

tzn. warunkowa wartość oczekiwana zmiennej losowej Y nie zależy od składnika stochastycznego oraz

$$\mathbb{E}\left[(Y - \mathbb{E}(Y|x_1, x_2, \dots, x_k))^2\right] = \mathbb{E}(\xi^2) = \sigma^2$$

czyli wariancja składnika stochastycznego jest miarą odchyień zmiennej losowej Y od jej warunkowej wartości oczekiwanej.

Rozważać będziemy problem estymacji modelu, tj. dopasowania modelu (2) do danych. Posługiwać się będziemy formalizmem macierzowym:

Regresja liniowa wielowymiarowa

- Niech Y_i oznacza zmienną losową, której realizację oznaczmy przez y_i , zaś przez \mathbf{Y} wektor kolumnowy $n \times 1$ zmiennych losowych Y_i , z kolei \mathbf{y} niech będzie wektorem kolumnowym $n \times 1$ realizacji y_i .
- Niech x_{li} oznacza l -tą z kolei ustaloną w próbie wartość zmiennej objaśniającej x_i i niech \mathbf{X} oznacza macierz wymiaru $n \times k$ wartości wszystkich zmiennych objaśniających, gdzie $n > k$; niech też rząd macierzy \mathbf{X} wynosi k .
- Niech ξ oznacza wektor kolumnowy $n \times 1$ zmiennych losowych ξ_i (przyjmujemy, że są iid).
- Niech β będzie wektorem $k \times 1$ prawdziwych (nieznanych) wartości współczynników β_i , zaś
- niech \mathbf{B} będzie wektorem kolumnowym $k \times 1$ estymatorów współczynników β_i (czyli B_i są zmiennymi losowymi), oraz
- niech \mathbf{b} będzie wektorem kolumnowym $k \times 1$ wartości estymatorów b_i (czyli b_i jest realizacją zmiennej losowej B_i).

Regresja liniowa wielowymiarowa

- Niech Y_i oznacza zmienną losową, której realizację oznaczymy przez y_i , zaś przez \mathbf{Y} wektor kolumnowy $n \times 1$ zmiennych losowych Y_i , z kolei \mathbf{y} niech będzie wektorem kolumnowym $n \times 1$ realizacji y_i .
- Niech x_{li} oznacza l -tą z kolei ustaloną w próbie wartość zmiennej objaśniającej x_i i niech \mathbf{X} oznacza macierz wymiaru $n \times k$ wartości wszystkich zmiennych objaśniających, gdzie $n > k$; niech też rząd macierzy \mathbf{X} wynosi k .
- Niech ξ oznacza wektor kolumnowy $n \times 1$ zmiennych losowych ξ_i (przyjmujemy, że są iid).
- Niech β będzie wektorem $k \times 1$ prawdziwych (nieznanych) wartości współczynników β_i , zaś
- niech \mathbf{B} będzie wektorem kolumnowym $k \times 1$ estymatorów współczynników β_i (czyli B_i są zmiennymi losowymi), oraz
- niech \mathbf{b} będzie wektorem kolumnowym $k \times 1$ wartości estymatorów b_i (czyli b_i jest realizacją zmiennej losowej B_i).

Regresja liniowa wielowymiarowa

- Niech Y_i oznacza zmienną losową, której realizację oznaczmy przez y_i , zaś przez \mathbf{Y} wektor kolumnowy $n \times 1$ zmiennych losowych Y_i , z kolei \mathbf{y} niech będzie wektorem kolumnowym $n \times 1$ realizacji y_i .
- Niech x_{li} oznacza l -tą z kolei ustaloną w próbie wartość zmiennej objaśniającej x_i i niech \mathbf{X} oznacza macierz wymiaru $n \times k$ wartości wszystkich zmiennych objaśniających, gdzie $n > k$; niech też rząd macierzy \mathbf{X} wynosi k .
- Niech ξ oznacza wektor kolumnowy $n \times 1$ zmiennych losowych ξ_i (przyjmujemy, że są iid).
- Niech β będzie wektorem $k \times 1$ prawdziwych (nieznanych) wartości współczynników β_i , zaś
- niech \mathbf{B} będzie wektorem kolumnowym $k \times 1$ estymatorów współczynników β_i (czyli B_i są zmiennymi losowymi), oraz
- niech \mathbf{b} będzie wektorem kolumnowym $k \times 1$ wartości estymatorów b_i (czyli b_i jest realizacją zmiennej losowej B_i).

Regresja liniowa wielowymiarowa

- Niech Y_i oznacza zmienną losową, której realizację oznaczmy przez y_i , zaś przez \mathbf{Y} wektor kolumnowy $n \times 1$ zmiennych losowych Y_i , z kolei \mathbf{y} niech będzie wektorem kolumnowym $n \times 1$ realizacji y_i .
- Niech x_{li} oznacza l -tą z kolei ustaloną w próbie wartość zmiennej objaśniającej x_i i niech \mathbf{X} oznacza macierz wymiaru $n \times k$ wartości wszystkich zmiennych objaśniających, gdzie $n > k$; niech też rząd macierzy \mathbf{X} wynosi k .
- Niech ξ oznacza wektor kolumnowy $n \times 1$ zmiennych losowych ξ_i (przyjmujemy, że są iid).
- Niech β będzie wektorem $k \times 1$ prawdziwych (nieznanych) wartości współczynników β_i , zaś
 - niech \mathbf{B} będzie wektorem kolumnowym $k \times 1$ estymatorów współczynników β_i (czyli B_i są zmiennymi losowymi), oraz
 - niech \mathbf{b} będzie wektorem kolumnowym $k \times 1$ wartości estymatorów b_i (czyli b_i jest realizacją zmiennej losowej B_i).

Regresja liniowa wielowymiarowa

- Niech Y_i oznacza zmienną losową, której realizację oznaczmy przez y_i , zaś przez \mathbf{Y} wektor kolumnowy $n \times 1$ zmiennych losowych Y_i , z kolei \mathbf{y} niech będzie wektorem kolumnowym $n \times 1$ realizacji y_i .
- Niech x_{li} oznacza l -tą z kolei ustaloną w próbie wartość zmiennej objaśniającej x_i i niech \mathbf{X} oznacza macierz wymiaru $n \times k$ wartości wszystkich zmiennych objaśniających, gdzie $n > k$; niech też rząd macierzy \mathbf{X} wynosi k .
- Niech ξ oznacza wektor kolumnowy $n \times 1$ zmiennych losowych ξ_i (przyjmujemy, że są iid).
- Niech β będzie wektorem $k \times 1$ prawdziwych (nieznanych) wartości współczynników β_i , zaś
- niech \mathbf{B} będzie wektorem kolumnowym $k \times 1$ estymatorów współczynników β_i (czyli B_i są zmiennymi losowymi), oraz
- niech \mathbf{b} będzie wektorem kolumnowym $k \times 1$ wartości estymatorów b_i (czyli b_i jest realizacją zmiennej losowej B_i).

Regresja liniowa wielowymiarowa

- Niech Y_i oznacza zmienną losową, której realizację oznaczmy przez y_i , zaś przez \mathbf{Y} wektor kolumnowy $n \times 1$ zmiennych losowych Y_i , z kolei \mathbf{y} niech będzie wektorem kolumnowym $n \times 1$ realizacji y_i .
- Niech x_{li} oznacza l -tą z kolei ustaloną w próbie wartość zmiennej objaśniającej x_i i niech \mathbf{X} oznacza macierz wymiaru $n \times k$ wartości wszystkich zmiennych objaśniających, gdzie $n > k$; niech też rząd macierzy \mathbf{X} wynosi k .
- Niech ξ oznacza wektor kolumnowy $n \times 1$ zmiennych losowych ξ_i (przyjmujemy, że są iid).
- Niech β będzie wektorem $k \times 1$ prawdziwych (nieznanych) wartości współczynników β_i , zaś
- niech \mathbf{B} będzie wektorem kolumnowym $k \times 1$ estymatorów współczynników β_i (czyli B_i są zmiennymi losowymi), oraz
- niech \mathbf{b} będzie wektorem kolumnowym $k \times 1$ wartości estymatorów b_i (czyli b_i jest realizacją zmiennej losowej B_i).

Wprowadzone wektory i macierze wyglądają następująco:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix},$$

$$\boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_k \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix}$$

Regresja liniowa wielowymiarowa

Losowanie obserwacji do próby interpretujemy jako proces n -krotnego obserwowania relacji stochastycznej (2). Wynikający z tego układ n równań można zapisać macierzowo jako

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$$

Regresja liniowa wielowymiarowa

Losowanie obserwacji do próby interpretujemy jako proces n -krotnego obserwowania relacji stochastycznej (2). Wynikający z tego układ n równań można zapisać macierzowo jako

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$$

Mając n -elementową próbę, realizację powyższego modelu można przedstawić jako

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$$

Regresja liniowa wielowymiarowa

Losowanie obserwacji do próby interpretujemy jako proces n -krotnego obserwowania relacji stochastycznej (2). Wynikający z tego układ n równań można zapisać macierzowo jako

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$$

Mając n -elementową próbę, realizację powyższego modelu można przedstawić jako

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$$

Estymacji wektora $\boldsymbol{\beta}$, czyli wyznaczenie wektora \mathbf{b} dokonuje się metodą najmniejszych kwadratów:

$$\begin{aligned} S &= \sum_{i=1}^n (y_i - (\mathbf{X}\mathbf{b})_i)^2 = (\mathbf{y} - \mathbf{X}\mathbf{b})^\top (\mathbf{y} - \mathbf{X}\mathbf{b}) \\ &= \mathbf{y}^\top \mathbf{y} - \mathbf{b}^\top \mathbf{X}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{X}\mathbf{b} + \mathbf{b}^\top \mathbf{X}^\top \mathbf{X}\mathbf{b} = \min. \end{aligned}$$

które to minimum znajdziemy różniczkując S :

$$\frac{\partial S}{\partial \mathbf{b}} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \mathbf{b} = 0 \quad (3)$$

$$\frac{\partial^2 S}{\partial \mathbf{b}^2} = 2\mathbf{X}^T \mathbf{X} \quad (4)$$

Macierz $\mathbf{X}^T \mathbf{X}$ jest symetryczna¹ i dodatnio określona², zatem w punkcie, w którym spełnione są równania (3) i (4), faktycznie S ma minimum.

¹Ponieważ $(\mathbf{X}^T \mathbf{X})^T = \mathbf{X}^T (\mathbf{X}^T)^T = \mathbf{X}^T \mathbf{X}$.

²Ponieważ rząd macierzy \mathbf{X} wynosi k .

$$\frac{\partial S}{\partial \mathbf{b}} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \mathbf{b} = 0 \quad (3)$$

$$\frac{\partial^2 S}{\partial \mathbf{b}^2} = 2\mathbf{X}^T \mathbf{X} \quad (4)$$

Macierz $\mathbf{X}^T \mathbf{X}$ jest symetryczna¹ i dodatnio określona², zatem w punkcie, w którym spełnione są równania (3) i (4), faktycznie S ma minimum. Rozwiązaniem równ. (3) jest

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (5)$$

Jest to estymator wektora β —nieobciążony i zgodny (co udowodnimy), a także najefektywniejszy (tw. Gaussa-Markowa—dowód pomijamy).

¹Ponieważ $(\mathbf{X}^T \mathbf{X})^T = \mathbf{X}^T (\mathbf{X}^T)^T = \mathbf{X}^T \mathbf{X}$.

²Ponieważ rząd macierzy \mathbf{X} wynosi k .

$$\frac{\partial S}{\partial \mathbf{b}} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \mathbf{b} = 0 \quad (3)$$

$$\frac{\partial^2 S}{\partial \mathbf{b}^2} = 2\mathbf{X}^T \mathbf{X} \quad (4)$$

Macierz $\mathbf{X}^T \mathbf{X}$ jest symetryczna¹ i dodatnio określona², zatem w punkcie, w którym spełnione są równania (3) i (4), faktycznie S ma minimum. Rozwiązaniem równ. (3) jest

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (5)$$

Jest to estymator wektora β —nieobciążony i zgodny (co udowodnimy), a także najefektywniejszy (tw. Gaussa-Markowa—dowód pomijamy). Obliczymy również wariancję $V(\mathbf{b})$ oraz σ^2 :

¹Ponieważ $(\mathbf{X}^T \mathbf{X})^T = \mathbf{X}^T (\mathbf{X}^T)^T = \mathbf{X}^T \mathbf{X}$.

²Ponieważ rząd macierzy \mathbf{X} wynosi k .

Przepisując równ. (5) dla estymatorów:

$$\mathbf{B} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, \quad (6)$$

uwzględniając, że $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$ oraz działając obustronnie na równ. (6) operatorem wartości oczekiwanej otrzymujemy

$$\begin{aligned} \mathbb{E}(\mathbf{B}) &= \mathbb{E}\left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi})\right) \\ &= \boldsymbol{\beta} + \mathbb{E}\left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\xi}\right) \\ &= \boldsymbol{\beta} + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbb{E}(\boldsymbol{\xi}) = \boldsymbol{\beta} \end{aligned}$$

co dowodzi nieobciążoności estymatora \mathbf{B} .

Regresja liniowa wielowymiarowa

Następnie, skoro $\mathbb{E}(\mathbf{B}) = \boldsymbol{\beta}$ oraz $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$:

$$V(\mathbf{B}) = \mathbb{E}[(\mathbf{B} - \mathbb{E}(\mathbf{B}))(\mathbf{B} - \mathbb{E}(\mathbf{B}))^T] = \mathbb{E}[(\mathbf{B} - \boldsymbol{\beta})(\mathbf{B} - \boldsymbol{\beta})^T]$$

Regresja liniowa wielowymiarowa

Następnie, skoro $\mathbb{E}(\mathbf{B}) = \boldsymbol{\beta}$ oraz $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$:

$$\begin{aligned} V(\mathbf{B}) &= \mathbb{E}[(\mathbf{B} - \mathbb{E}(\mathbf{B}))(\mathbf{B} - \mathbb{E}(\mathbf{B}))^T] = \mathbb{E}[(\mathbf{B} - \boldsymbol{\beta})(\mathbf{B} - \boldsymbol{\beta})^T] \\ &= \mathbb{E}\left[\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} - \boldsymbol{\beta}\right)\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} - \boldsymbol{\beta}\right)^T\right] \end{aligned}$$

Regresja liniowa wielowymiarowa

Następnie, skoro $\mathbb{E}(\mathbf{B}) = \boldsymbol{\beta}$ oraz $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$:

$$\begin{aligned}V(\mathbf{B}) &= \mathbb{E}[(\mathbf{B} - \mathbb{E}(\mathbf{B}))(\mathbf{B} - \mathbb{E}(\mathbf{B}))^T] = \mathbb{E}[(\mathbf{B} - \boldsymbol{\beta})(\mathbf{B} - \boldsymbol{\beta})^T] \\&= \mathbb{E}\left[\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} - \boldsymbol{\beta}\right)\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} - \boldsymbol{\beta}\right)^T\right] \\&= \mathbb{E}\left[\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\xi}\right)\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\xi}\right)^T\right] \quad \{\text{zob. przypis 1}\}\end{aligned}$$

Regresja liniowa wielowymiarowa

Następnie, skoro $\mathbb{E}(\mathbf{B}) = \boldsymbol{\beta}$ oraz $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$:

$$\begin{aligned}V(\mathbf{B}) &= \mathbb{E}[(\mathbf{B} - \mathbb{E}(\mathbf{B}))(\mathbf{B} - \mathbb{E}(\mathbf{B}))^T] = \mathbb{E}[(\mathbf{B} - \boldsymbol{\beta})(\mathbf{B} - \boldsymbol{\beta})^T] \\&= \mathbb{E}\left[\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} - \boldsymbol{\beta}\right)\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} - \boldsymbol{\beta}\right)^T\right] \\&= \mathbb{E}\left[\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\xi}\right)\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\xi}\right)^T\right] \quad \{\text{zob. przypis 1}\} \\&= \mathbb{E}\left[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\xi}\boldsymbol{\xi}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\right]\end{aligned}$$

Regresja liniowa wielowymiarowa

Następnie, skoro $\mathbb{E}(\mathbf{B}) = \boldsymbol{\beta}$ oraz $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$:

$$\begin{aligned}V(\mathbf{B}) &= \mathbb{E}[(\mathbf{B} - \mathbb{E}(\mathbf{B}))(\mathbf{B} - \mathbb{E}(\mathbf{B}))^T] = \mathbb{E}[(\mathbf{B} - \boldsymbol{\beta})(\mathbf{B} - \boldsymbol{\beta})^T] \\&= \mathbb{E}\left[\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} - \boldsymbol{\beta}\right)\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} - \boldsymbol{\beta}\right)^T\right] \\&= \mathbb{E}\left[\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\xi}\right)\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\xi}\right)^T\right] \quad \{\text{zob. przypis 1}\} \\&= \mathbb{E}\left[\mathbf{X}^T\boldsymbol{\xi}\boldsymbol{\xi}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\right] \\&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}^T)\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\end{aligned}$$

Regresja liniowa wielowymiarowa

Następnie, skoro $\mathbb{E}(\mathbf{B}) = \boldsymbol{\beta}$ oraz $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$:

$$\begin{aligned}V(\mathbf{B}) &= \mathbb{E}[(\mathbf{B} - \mathbb{E}(\mathbf{B}))(\mathbf{B} - \mathbb{E}(\mathbf{B}))^T] = \mathbb{E}[(\mathbf{B} - \boldsymbol{\beta})(\mathbf{B} - \boldsymbol{\beta})^T] \\&= \mathbb{E}\left[\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} - \boldsymbol{\beta}\right)\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} - \boldsymbol{\beta}\right)^T\right] \\&= \mathbb{E}\left[\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\xi}\right)\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\xi}\right)^T\right] \quad \{\text{zob. przypis 1}\} \\&= \mathbb{E}\left[\mathbf{X}^T\boldsymbol{\xi}\boldsymbol{\xi}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\right] \\&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}^T)\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\end{aligned}$$

Macierz $\mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}^T)$ jest macierzą kowariancji zmiennych losowych ξ_i , które są iid o wariancji $V(\xi_i) = \sigma^2$, zatem $\mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}^T) = \sigma^2\mathbf{I}$, czyli otrzymujemy, że

$$V(\mathbf{B}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$$

Regresja liniowa wielowymiarowa

Skoro mamy, że $\mathbb{E}(\mathbf{B}) = \boldsymbol{\beta}$ i $V(\mathbf{B}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$ oraz wiemy, że $\mathbf{X}^T \mathbf{X}$ jest dodatnio określona, czyli $\mathbf{Q}_n \equiv \frac{1}{n} \mathbf{X}^T \mathbf{X}$ też jest dodatnio określona, to żeby wykazać zgodność estymatora \mathbf{B} wystarczy pokazać, że $V(\mathbf{B}) \xrightarrow[n \rightarrow \infty]{} 0$. Zakładamy, że $\mathbf{Q}_n \xrightarrow[n \rightarrow \infty]{} \mathbf{Q}$ też dodatnio określona.

Regresja liniowa wielowymiarowa

Skoro mamy, że $\mathbb{E}(\mathbf{B}) = \boldsymbol{\beta}$ i $V(\mathbf{B}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$ oraz wiemy, że $\mathbf{X}^\top \mathbf{X}$ jest dodatnio określona, czyli $\mathbf{Q}_n \equiv \frac{1}{n} \mathbf{X}^\top \mathbf{X}$ też jest dodatnio określona, to żeby wykazać zgodność estymatora \mathbf{B} wystarczy pokazać, że $V(\mathbf{B}) \xrightarrow[n \rightarrow \infty]{} 0$. Zakładamy, że $\mathbf{Q}_n \xrightarrow[n \rightarrow \infty]{} \mathbf{Q}$ też dodatnio określona.

Zachodzi wtedy

$$V(\mathbf{B}) = \frac{\sigma^2}{n} \left(\frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} = \frac{\sigma^2}{n} \mathbf{Q}_n^{-1}$$

Regresja liniowa wielowymiarowa

Skoro mamy, że $\mathbb{E}(\mathbf{B}) = \boldsymbol{\beta}$ i $V(\mathbf{B}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$ oraz wiemy, że $\mathbf{X}^\top \mathbf{X}$ jest dodatnio określona, czyli $\mathbf{Q}_n \equiv \frac{1}{n} \mathbf{X}^\top \mathbf{X}$ też jest dodatnio określona, to żeby wykazać zgodność estymatora \mathbf{B} wystarczy pokazać, że $V(\mathbf{B}) \xrightarrow{n \rightarrow \infty} 0$. Zakładamy, że $\mathbf{Q}_n \xrightarrow{n \rightarrow \infty} \mathbf{Q}$ też dodatnio określona.

Zachodzi wtedy

$$V(\mathbf{B}) = \frac{\sigma^2}{n} \left(\frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} = \frac{\sigma^2}{n} \mathbf{Q}_n^{-1}$$

Ponieważ \mathbf{Q}_n jest dodatnio określona, to \mathbf{Q}_n^{-1} istnieje i jej elementy są skończone, tj. $\sup_{i,j} |(\mathbf{Q}_n^{-1})_{ij}| = C < \infty$, zatem

$$\forall j : V(B_j) \leq \frac{C\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0$$

Regresja liniowa wielowymiarowa

Skoro mamy, że $\mathbb{E}(\mathbf{B}) = \boldsymbol{\beta}$ i $V(\mathbf{B}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$ oraz wiemy, że $\mathbf{X}^\top \mathbf{X}$ jest dodatnio określona, czyli $\mathbf{Q}_n \equiv \frac{1}{n} \mathbf{X}^\top \mathbf{X}$ też jest dodatnio określona, to żeby wykazać zgodność estymatora \mathbf{B} wystarczy pokazać, że $V(\mathbf{B}) \xrightarrow{n \rightarrow \infty} 0$. Zakładamy, że $\mathbf{Q}_n \xrightarrow{n \rightarrow \infty} \mathbf{Q}$ też dodatnio określona.

Zachodzi wtedy

$$V(\mathbf{B}) = \frac{\sigma^2}{n} \left(\frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} = \frac{\sigma^2}{n} \mathbf{Q}_n^{-1}$$

Ponieważ \mathbf{Q}_n jest dodatnio określona, to \mathbf{Q}_n^{-1} istnieje i jej elementy są skończone, tj. $\sup_{i,j} |(\mathbf{Q}_n^{-1})_{ij}| = C < \infty$, zatem

$$\forall j : V(B_j) \leq \frac{C\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0$$

i z nierówności Czebyszewa (slajd 2, [wykład 8](#))

$$\forall j : P(|B_j - \beta_j| \geq \varepsilon) \leq \frac{V(B_j)}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

zatem faktycznie \mathbf{B} jest zgodny.

Na koniec rozważmy wektor reszt:

$$\mathbf{r} = \mathbf{Y} - \mathbf{XB}$$

Na koniec rozważmy wektor reszt:

$$\mathbf{r} = \mathbf{Y} - \mathbf{XB}$$

Wtedy estymatorem σ^2 jest (dowód pomijamy)

$$\hat{\sigma}^2 = \frac{\mathbf{r}^T \mathbf{r}}{n - k} \equiv \frac{RSS}{n - k}$$

Na koniec rozważmy wektor reszt:

$$\mathbf{r} = \mathbf{Y} - \mathbf{XB}$$

Wtedy estymatorem σ^2 jest (dowód pomijamy)

$$\hat{\sigma}^2 = \frac{\mathbf{r}^T \mathbf{r}}{n - k} \equiv \frac{RSS}{n - k}$$

gdzie RSS to tzw. *residual sum of squares*:

$$RSS = \sum_{i=1}^n r_i^2 = \sum_{i=1}^n \left(Y_i - \sum_{j=1}^k X_{ij} B_j \right)^2$$

Regresja liniowa wielowymiarowa

Przykład. Spodziewamy się, że cena kamery CCD zależy głównie od jej wieku (okresu używalności) oraz wielkości (pola powierzchni, ilości pikseli). Oferty sprzedażowe podsumowuje tabela:

| Cena (\$\$) | Wiek (miesiące) | Powierzchnia (cm ²) |
|-------------|-----------------|---------------------------------|
| 745 | 36 | 66 |
| 895 | 37 | 68 |
| 442 | 47 | 64 |
| 440 | 32 | 53 |
| 1598 | 1 | 101 |

Możliwy model ceny kamery to

$$y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \beta_2 x_{i2} + \xi_i, \quad i = 1, \dots, 5$$

gdzie wyraz wolny β_0 został wprowadzony poprzez przyjęcie $x_{i0} = 1$.

Regresja liniowa wielowymiarowa

$$\mathbf{y} = \begin{pmatrix} 745 \\ 895 \\ 442 \\ 440 \\ 1598 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 36 & 66 \\ 1 & 37 & 68 \\ 1 & 47 & 64 \\ 1 & 32 & 53 \\ 1 & 1 & 101 \end{pmatrix},$$

Regresja liniowa wielowymiarowa

$$\mathbf{y} = \begin{pmatrix} 745 \\ 895 \\ 442 \\ 440 \\ 1598 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 36 & 66 \\ 1 & 37 & 68 \\ 1 & 47 & 64 \\ 1 & 32 & 53 \\ 1 & 1 & 101 \end{pmatrix},$$

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} =$$

$$\begin{pmatrix} 5 & 153 & 352 \\ 153 & 5899 & 9697 \\ 352 & 9697 & 26086 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 36 & 37 & 47 & 32 & 1 \\ 66 & 68 & 64 & 53 & 101 \end{pmatrix} \begin{pmatrix} 745 \\ 895 \\ 442 \\ 440 \\ 1598 \end{pmatrix} = \begin{pmatrix} -281.43 \\ -7.611 \\ 19.01 \end{pmatrix}$$

czyli

$$y = -281.43 - 7.611x_1 + 19.01x_2$$

oraz

$$\hat{\sigma}^2 = 25.344$$

Regresja nieliniowa

Rozważaliśmy dotychczas modele liniowe w parametrach \mathbf{a} , dla których minimum funkcji (zob. slajd 3)

$$S \equiv S(\mathbf{a}) = \sum_{i=1}^n r_i^2(\mathbf{a})$$

można znaleźć analitycznie.

Regresja nieliniowa

Rozważaliśmy dotychczas modele liniowe w parametrach \mathbf{a} , dla których minimum funkcji (zob. slajd 3)

$$S \equiv S(\mathbf{a}) = \sum_{i=1}^n r_i^2(\mathbf{a})$$

można znaleźć analitycznie.

W praktycznych zastosowaniach często (zazwyczaj) dopasowuje się modele faktycznie nieliniowe, np.

$$y(x) = ae^{-bx} + c,$$

$$y(x) = a \sin(bx + c),$$

$$y(x) = \frac{a}{1 + e^{-b(x-c)}}.$$

Regresja nieliniowa

Rozważaliśmy dotychczas modele liniowe w parametrach \mathbf{a} , dla których minimum funkcji (zob. slajd 3)

$$S \equiv S(\mathbf{a}) = \sum_{i=1}^n r_i^2(\mathbf{a})$$

można znaleźć analitycznie.

W praktycznych zastosowaniach często (zazwyczaj) dopasowuje się modele faktycznie nieliniowe, np.

$$y(x) = ae^{-bx} + c,$$

$$y(x) = a \sin(bx + c),$$

$$y(x) = \frac{a}{1 + e^{-b(x-c)}}.$$

W takich przypadkach:

- równania $\partial S / \partial a_k = 0$ są nieliniowe,
- rozwiązanie analityczne powyższych zwykle nie istnieje,
- minimum funkcji celu znajduje się metodami numerycznymi.

Regresja nieliniowa

Rozważaliśmy dotychczas modele liniowe w parametrach \mathbf{a} , dla których minimum funkcji (zob. slajd 3)

$$S \equiv S(\mathbf{a}) = \sum_{i=1}^n r_i^2(\mathbf{a})$$

można znaleźć analitycznie.

W praktycznych zastosowaniach często (zazwyczaj) dopasowuje się modele faktycznie nieliniowe, np.

$$y(x) = ae^{-bx} + c,$$

$$y(x) = a \sin(bx + c),$$

$$y(x) = \frac{a}{1 + e^{-b(x-c)}}.$$

W takich przypadkach:

- równania $\partial S / \partial a_k = 0$ są nieliniowe,
- rozwiązanie analityczne powyższych zwykle nie istnieje,
- minimum funkcji celu znajduje się metodami numerycznymi.

Regresja nieliniowa

Metody iteracyjne

Startując od pewnego przybliżenia (wartości początkowych) parametrów

$$\mathbf{a}_0$$

algorytm generuje kolejne przybliżenia:

$$\mathbf{a}_0 \rightarrow \mathbf{a}_1 \rightarrow \mathbf{a}_2 \rightarrow \dots$$

Regresja nieliniowa

Metody iteracyjne

Startując od pewnego przybliżenia (wartości początkowych) parametrów

$$\mathbf{a}_0$$

algorytm generuje kolejne przybliżenia:

$$\mathbf{a}_0 \rightarrow \mathbf{a}_1 \rightarrow \mathbf{a}_2 \rightarrow \dots$$

Celem jest znalezienie minimum funkcji:

$$S(\mathbf{a}) = \sum_{i=1}^n r_i^2(\mathbf{a}) = \mathbf{r}^T \mathbf{r} = \min.$$

Regresja nieliniowa

Metody iteracyjne

Startując od pewnego przybliżenia (wartości początkowych) parametrów

$$\mathbf{a}_0$$

algorytm generuje kolejne przybliżenia:

$$\mathbf{a}_0 \rightarrow \mathbf{a}_1 \rightarrow \mathbf{a}_2 \rightarrow \dots$$

Celem jest znalezienie minimum funkcji:

$$S(\mathbf{a}) = \sum_{i=1}^n r_i^2(\mathbf{a}) = \mathbf{r}^T \mathbf{r} = \min.$$

Najważniejsze informacje wykorzystywane przez tego typu algorytmy:

- wartość funkcji $S(\mathbf{a})$,
- jej gradient: $\nabla S(\mathbf{a})$,
- macierz \mathbf{H} drugich pochodnych (hesjan): $H_{lk} = \frac{\partial^2 S(\mathbf{a})}{\partial a_l \partial a_k}$.

Regresja nieliniowa

Metody iteracyjne

Startując od pewnego przybliżenia (wartości początkowych) parametrów

$$\mathbf{a}_0$$

algorytm generuje kolejne przybliżenia:

$$\mathbf{a}_0 \rightarrow \mathbf{a}_1 \rightarrow \mathbf{a}_2 \rightarrow \dots$$

Celem jest znalezienie minimum funkcji:

$$S(\mathbf{a}) = \sum_{i=1}^n r_i^2(\mathbf{a}) = \mathbf{r}^T \mathbf{r} = \min.$$

Najważniejsze informacje wykorzystywane przez tego typu algorytmy:

- wartość funkcji $S(\mathbf{a})$,
- jej gradient: $\nabla S(\mathbf{a})$,
- macierz \mathbf{H} drugich pochodnych (hesjan): $H_{lk} = \frac{\partial^2 S(\mathbf{a})}{\partial a_l \partial a_k}$.

Problem sprowadza się więc do optymalizacji funkcji wielu zmiennych.

Regresja nieliniowa

Niektóre metody minimalizacji

- **Metoda Newtona** — wykorzystuje gradient i hesjan:

$$\mathbf{a}_{n+1} = \mathbf{a}_n - \mathbf{H}^{-1} \Big|_{\mathbf{a}_n} \nabla S(\mathbf{a}_n)$$

Bardzo szybka blisko minimum, ale wymaga obliczania drugich pochodnych (numerycznie kosztowne).

- **Metody gradientowe** — poruszają się w kierunku najszybszego spadku:

$$\mathbf{a}_{n+1} = \mathbf{a}_n - \eta \nabla S(\mathbf{a}_n)$$

gdzie η jest krokiem algorytmu.

- **Metody Gaussa–Newtona i Levenberga–Marquardta** — najczęściej stosowane w regresji nieliniowej. Łączą szybkość metody Newtona ze stabilnością metod gradientowych.
- **Metoda Brenta** — jednowymiarowa metoda minimalizacji bez użycia pochodnych. Często stosowana pomocniczo przy wyznaczaniu optymalnego kroku algorytmu np. w metodach gradientowych.

Regresja nieliniowa

Niektóre metody minimalizacji

- **Metoda Newtona** — wykorzystuje gradient i hesjan:

$$\mathbf{a}_{n+1} = \mathbf{a}_n - \mathbf{H}^{-1} \Big|_{\mathbf{a}_n} \nabla S(\mathbf{a}_n)$$

Bardzo szybka blisko minimum, ale wymaga obliczania drugich pochodnych (numerycznie kosztowne).

- **Metody gradientowe** — poruszają się w kierunku najszybszego spadku:

$$\mathbf{a}_{n+1} = \mathbf{a}_n - \eta \nabla S(\mathbf{a}_n)$$

gdzie η jest krokiem algorytmu.

- **Metody Gaussa–Newtona i Levenberga–Marquardta** — najczęściej stosowane w regresji nieliniowej. Łączą szybkość metody Newtona ze stabilnością metod gradientowych.
- **Metoda Brenta** — jednowymiarowa metoda minimalizacji bez użycia pochodnych. Często stosowana pomocniczo przy wyznaczaniu optymalnego kroku algorytmu np. w metodach gradientowych.

Regresja nieliniowa

Niektóre metody minimalizacji

- **Metoda Newtona** — wykorzystuje gradient i hesjan:

$$\mathbf{a}_{n+1} = \mathbf{a}_n - \mathbf{H}^{-1} \Big|_{\mathbf{a}_n} \nabla S(\mathbf{a}_n)$$

Bardzo szybka blisko minimum, ale wymaga obliczania drugich pochodnych (numerycznie kosztowne).

- **Metody gradientowe** — poruszają się w kierunku najszybszego spadku:

$$\mathbf{a}_{n+1} = \mathbf{a}_n - \eta \nabla S(\mathbf{a}_n)$$

gdzie η jest krokiem algorytmu.

- **Metody Gaussa–Newtona i Levenberga–Marquardta** — najczęściej stosowane w regresji nieliniowej. Łączą szybkość metody Newtona ze stabilnością metod gradientowych.
- **Metoda Brenta** — jednowymiarowa metoda minimalizacji bez użycia pochodnych. Często stosowana pomocniczo przy wyznaczaniu optymalnego kroku algorytmu np. w metodach gradientowych.

Regresja nieliniowa

Niektóre metody minimalizacji

- **Metoda Newtona** — wykorzystuje gradient i hesjan:

$$\mathbf{a}_{n+1} = \mathbf{a}_n - \mathbf{H}^{-1} \Big|_{\mathbf{a}_n} \nabla S(\mathbf{a}_n)$$

Bardzo szybka blisko minimum, ale wymaga obliczania drugich pochodnych (numerycznie kosztowne).

- **Metody gradientowe** — poruszają się w kierunku najszybszego spadku:

$$\mathbf{a}_{n+1} = \mathbf{a}_n - \eta \nabla S(\mathbf{a}_n)$$

gdzie η jest krokiem algorytmu.

- **Metody Gaussa–Newtona i Levenberga–Marquardta** — najczęściej stosowane w regresji nieliniowej. Łączą szybkość metody Newtona ze stabilnością metod gradientowych.
- **Metoda Brenta** — jednowymiarowa metoda minimalizacji bez użycia pochodnych. Często stosowana pomocniczo przy wyznaczaniu optymalnego kroku algorytmu np. w metodach gradientowych.

Najczęstsze problemy:

- wiele minimów lokalnych,
- zależność wyniku od wartości początkowych,
- silne korelacje parametrów,
- trudna geometria funkcji $S(\mathbf{a})$ (np. minima jako wąskie doliny, płaskie obszary itp.),
- wolna zbieżność algorytmu.

Najczęstsze problemy:

- wiele minimów lokalnych,
- zależność wyniku od wartości początkowych,
- silne korelacje parametrów,
- trudna geometria funkcji $S(\mathbf{a})$ (np. minima jako wąskie doliny, płaskie obszary itp.),
- wolna zbieżność algorytmu.

W praktyce:

- dobry wybór parametrów początkowych jest kluczowy,
- testuje się wiele algorytmów,
- nakłada się więzy na parametry,
- najczęściej używane pakiety

PYTHON:

- `scipy.optimize.curve_fit`
- `scipy.optimize.minimize`
- `lmfit`

Problemy praktyczne

Najczęstsze problemy:

- wiele minimów lokalnych,
- zależność wyniku od wartości początkowych,
- silne korelacje parametrów,
- trudna geometria funkcji $S(\mathbf{a})$ (np. minima jako wąskie doliny, płaskie obszary itp.),
- wolna zbieżność algorytmu.

Funkcja Rosenbrocka:

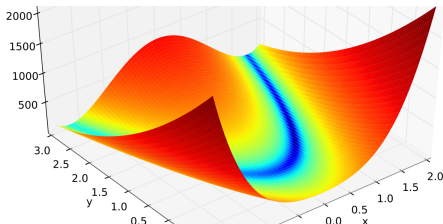
$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2$$

ma minimum w punkcie $(1, 1)$, ale jej kształt jest wyzwaniem dla algorytmów:

W praktyce:

- dobry wybór parametrów początkowych jest kluczowy,
- testuje się wiele algorytmów,
- nakłada się więzy na parametry,
- najczęściej używane pakiety PYTHON:

- `scipy.optimize.curve_fit`
- `scipy.optimize.minimize`
- `lmfit`



Inne zagadnienia statystyczne

- splajny (np. kubiczny splajn naturalny)—interpolacja danych
- filtr Savitzky'ego-Golaya—wygładzanie zaszumionego szeregu czasowego
- kryteria informacyjne, AIC & BIC—wybór modelu
- analiza skupień (*clustering*):
 - model mieszaniny rozkładów Gaussa (*Gaussian Mixture Model, GMM*)—w przestrzeniach o dowolnym wymiarze
 - *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)*
 - k -średnie
- redukcja wymiarów:
 - analiza głównych składowych (*Principal Component Analysis, PCA*)
 - stochastyczna metoda porządkowania sąsiadów w oparciu o rozkład t (*t-distributed Stochastic Neighbour Embedding, t-SNE*)
 - *Uniform Manifold Approximation and Projection (UMAP)*
- uczenie maszynowe:
 - las losowy (*random forest*)
 - maszyna wektorów nośnych (*support vector machines, SVM*)
 - konwolucyjne i rekurencyjne sieci neuronowe (*convolutional/recurrence neural networks, CNN & RNN*)

Niektóre kursy powiązane ze statystyką lub poszerzające poruszane zagadnienia

- [Fizyka statystyczna](#), dr hab. Jacek Jurkowski, prof. UMK
- [Metody numeryczne I](#), prof. dr hab. Ireneusz Grabowski;
[Metody numeryczne II](#), prof. dr hab. Oleksandr Sokolov
- [Analiza sygnałów](#), dr hab. Tomasz Piotrowski, prof. UMK
- [Podstawy teorii sygnałów](#) oraz [Podstawy i algorytmy przetwarzania sygnałów](#), dr hab. inż. Jadwiga Lal-Jadziak, prof. UMK
- [Wprowadzenie do procesów stochastycznych](#), dr Tomasz Wasak, prof. UMK
- [Analiza szeregów czasowych w fizyce i astronomii, w języku Python](#), dr Grzegorz Nowak
- [Uczenie maszynowe, algorytmy i systemy datamining](#), dr hab. Norbert Jankowski, prof. UMK
- [Sztuczne sieci neuronowe](#), dr Marek Grochowski; [Sieci neuronowe w modelowaniu i sterowaniu](#), prof. dr hab. Oleksandr Sokolov
- [Sztuczna inteligencja](#), prof. dr hab. Włodzisław Duch