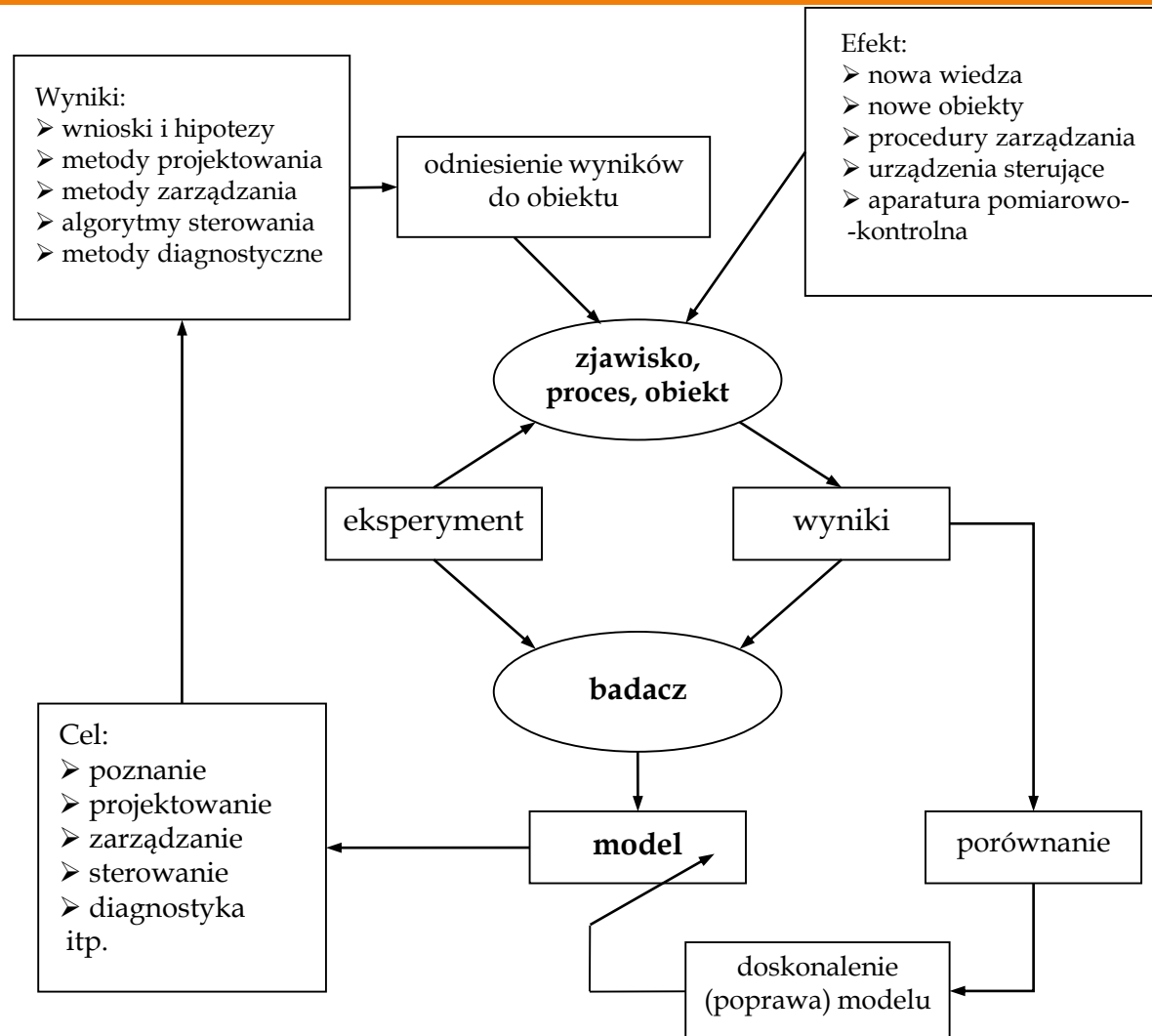


Metody Systemowe i Decyzyjne w Informatyce



Wykład 2b. Podstawowe zadania identyfikacji. Wybór optymalnego modelu

Model w badaniach systemowych

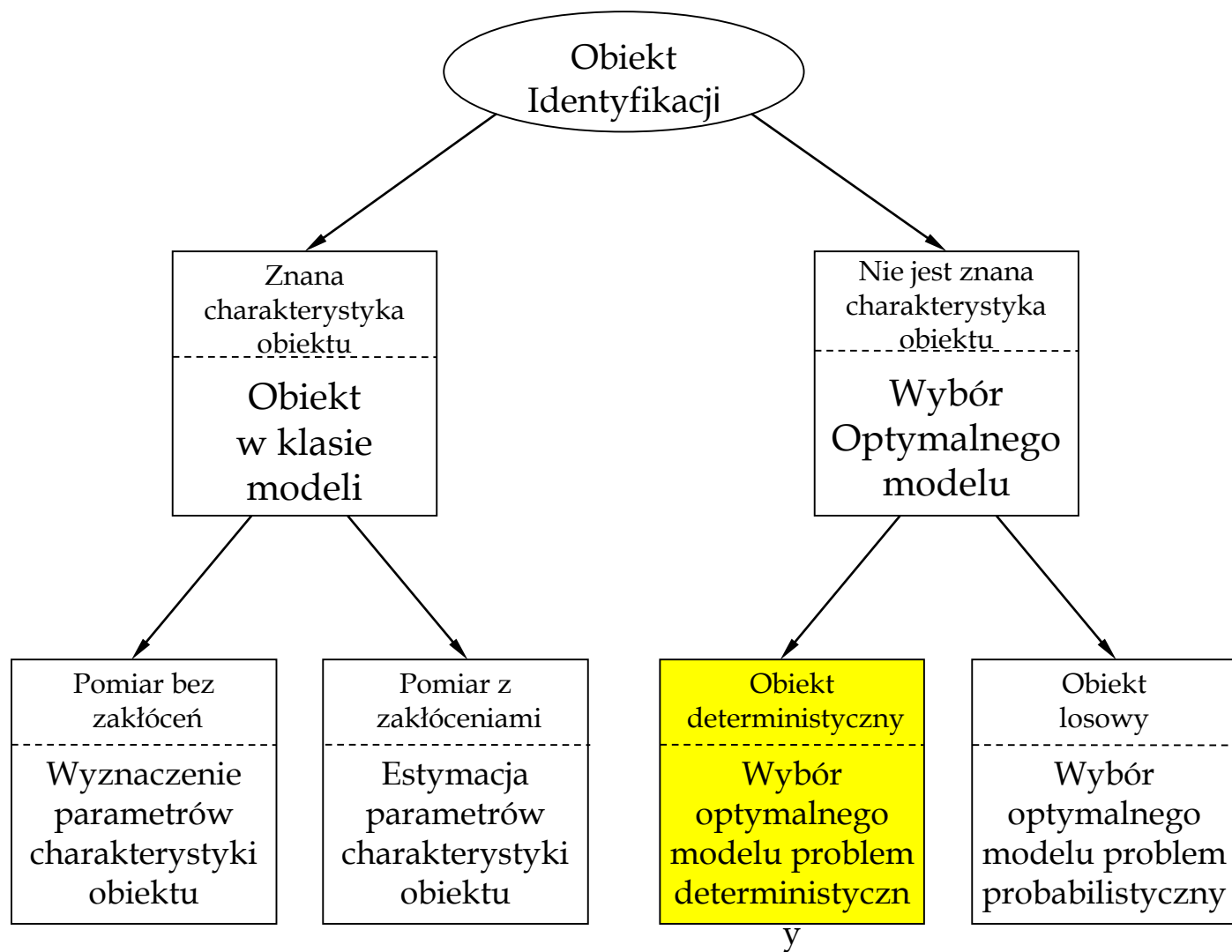


Wybór optymalnego modelu

Przypadek Deterministyczny

∞ 3 ∞

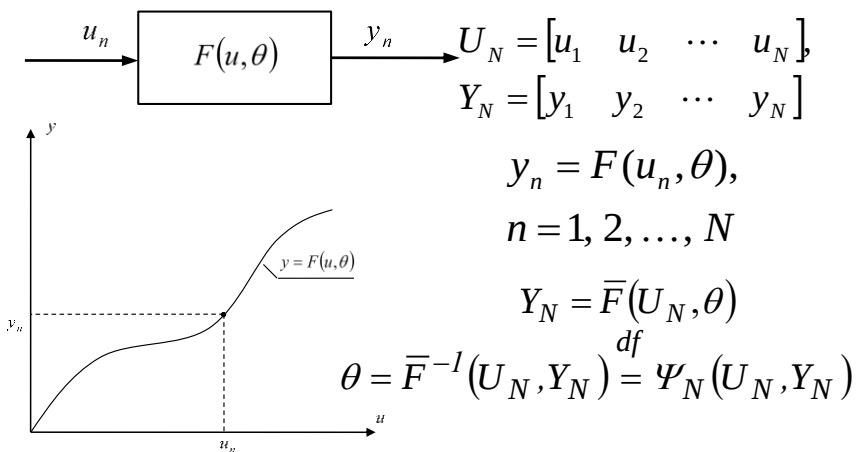
Typowe zadania identyfikacji



Podstawowe zadania identyfikacji podsumowanie

Deterministyczny

Obiekt w klasie modeli

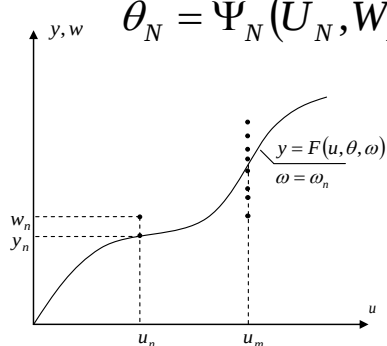
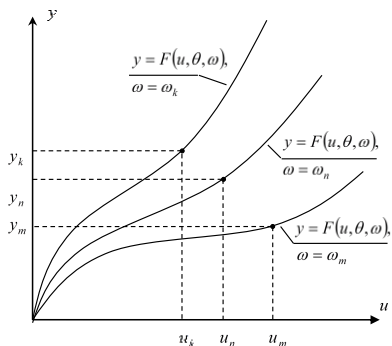


Lowy

Metody:

- Najmniejszych kwadratów
- Maksymalnej wiarygodności
- Bayes'a

$$\theta_N = \Psi_N(U_N, W_N)$$



Wybór optymalnego modelu

$$\bar{y} = \Phi(u, \theta_N^*)$$

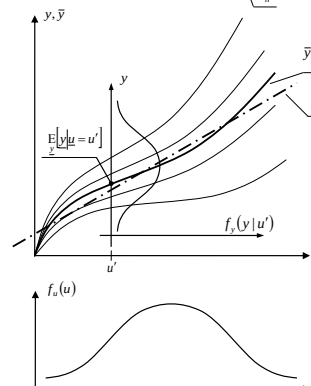
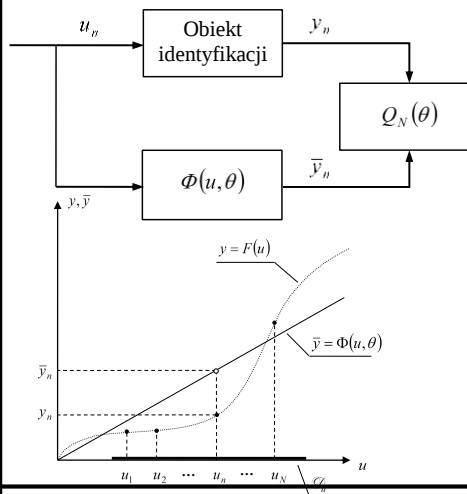
$$\theta_N^* \rightarrow Q_N(\theta_N^*) = \min_{\theta} Q_N(\theta)$$

$$\theta_N^* = \Psi_N(U_N, Y_N)$$

$$U_N = [u_1 \quad u_2 \quad \cdots \quad u_N]$$

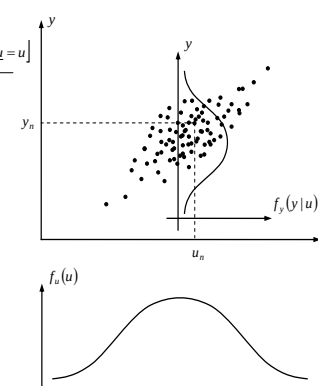
Model jest optymalny:

- dla zadanej serii pomiarowej
- przyjętego modelu
- przyjętego wskaźnika jakości identyfikacji



Pełna informacja

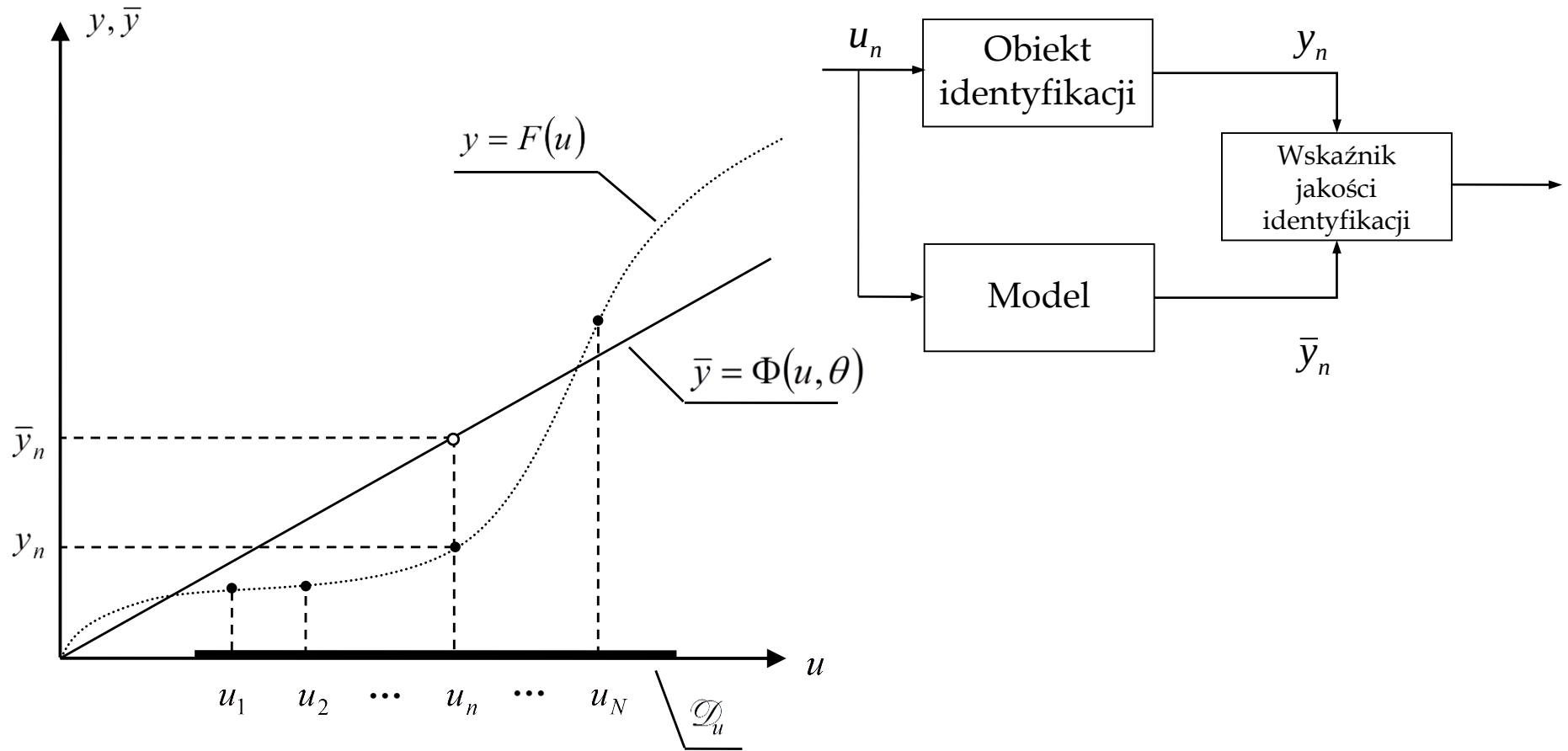
- Regresja I rodzaju
- Regresja II rodzaju



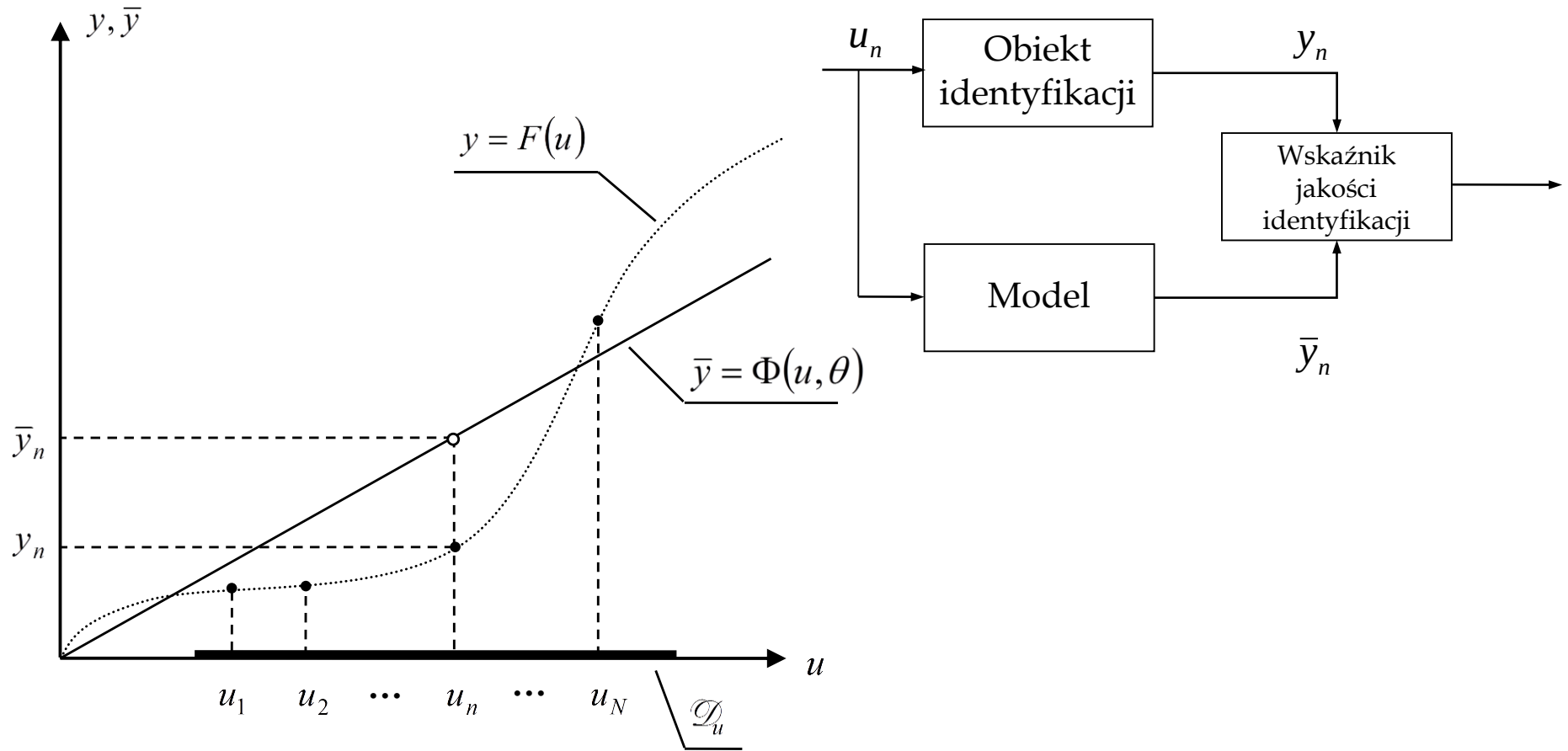
Niepełna informacja

- Estymacja wskaźnika jakości
- Estymacja parametrów rozkładu
- Estymacja rozkładu

Wybór optymalnego modelu - Przypadek deterministyczny

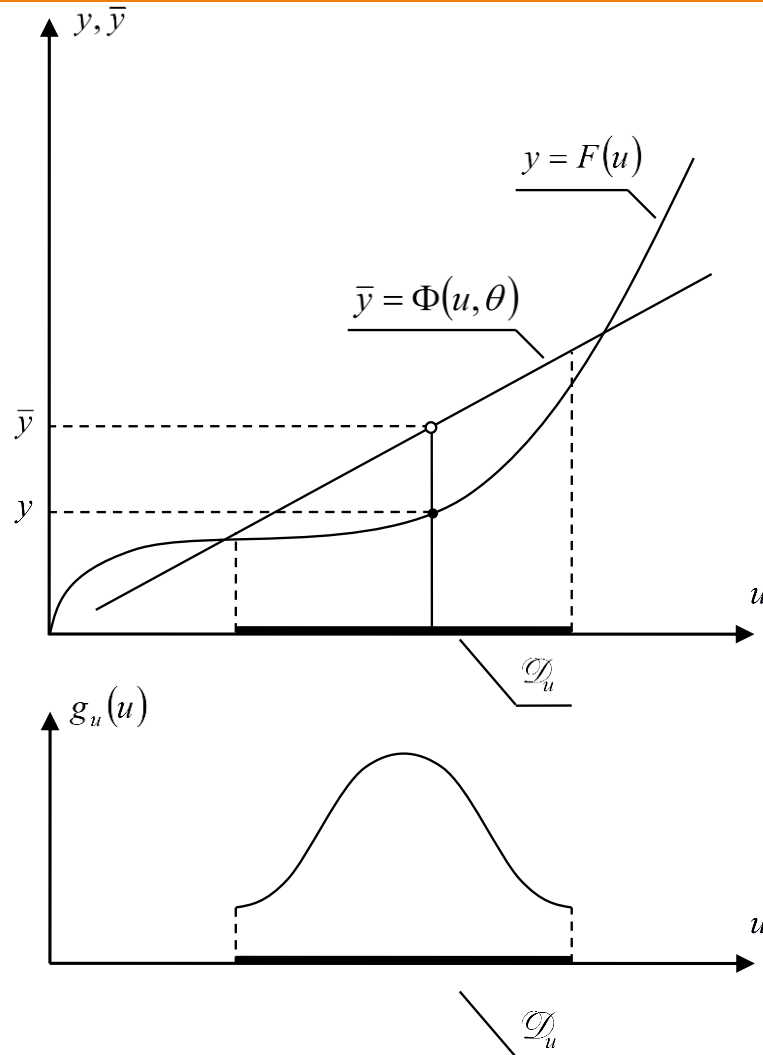


Wybór optymalnego modelu - Przypadek deterministyczny



Zadanie aproksymacji charakterystyki statycznej obiektu

☞ Sformułowanie problemu



Zadanie aproksymacji charakterystyki statycznej obiektu

∞ Sformułowanie problemu

Charakterystyka statyczna obiektu: $y = F(u)$

F - znana funkcja

u - wektor wejść $u \in \mathcal{U} \subseteq \mathbb{R}^S$ $\mathcal{U}$ - przestrzeń wejść

y - wektor wyjść $y \in \mathcal{Y} \subseteq \mathbb{R}^L$ $\mathcal{Y}$ - przestrzeń wyjść

Funkcja aproksymująca (model): $\bar{y} = \Phi(u, \theta)$

Φ - arbitralnie zadana funkcja

$\bar{y}$ - wektor wyjść modelu $\bar{y} \in \overline{\mathcal{Y}} \subseteq \mathbb{R}^L$

θ - wektor parametrów modelu $\theta \in \Theta \subseteq \mathbb{R}^R$

$\mathcal{U}_u \subseteq \mathcal{U}$ - podzbiór przestrzeni wejść $g_u(u)$ - funkcja wagi

Zadanie aproksymacji charakterystyki statycznej obiektu

∞ Sformułowanie problemu

Miara różnicy pomiędzy wartością funkcji aproksymowanej i aproksymującej w punkcie u : $\forall u \in \mathcal{D}_u \quad q(y, \bar{y}) = q(F(u), \Phi(u, \theta))$

gdzie np.: $q(y, \bar{y}) = \sum_{l=1}^L (y^{(l)} - \bar{y}^{(l)})^2$ $q(y, \bar{y}) = \sum_{l=1}^L |y^{(l)} - \bar{y}^{(l)}|$ L – wymiarowe wyjście

$$q(y, \bar{y}) = (y - \bar{y})^2 = (F(u) - \Phi(u, \theta))^2$$

$$q(y, \bar{y}) = |y - \bar{y}| = |F(u) - \Phi(u, \theta)|$$

jednowymiarowe wyjście

$(L = 1)$

Zadanie aproksymacji charakterystyki statycznej obiektu

∞ Sformułowanie problemu

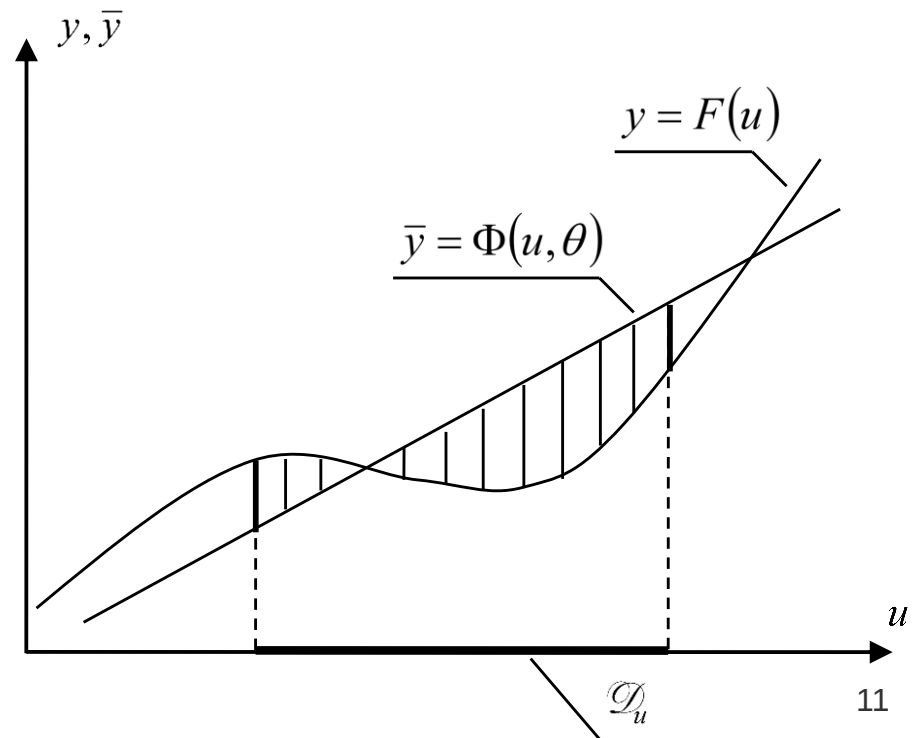
Wskaźnik jakości aproksymacji – miara różnicy pomiędzy funkcją Aproksymowaną $F(u)$ i aproksymującą $\Phi(u, \theta)$ w obszarze $\mathcal{D}_u$

$$Q(\theta) = \|F(u) - \Phi(u, \theta)\|_{\mathcal{D}_u}$$

$$Q(\theta) = \int_{\mathcal{D}_u} q(F(u), \Phi(u, \theta)) g_u(u) du$$

np. :

$$Q(\theta) = \int_{\mathcal{D}_u} |F(u) - \Phi(u, \theta)| g_u(u) du$$



Zadanie aproksymacji charakterystyki statycznej obiektu

∞ Sformułowanie problemu

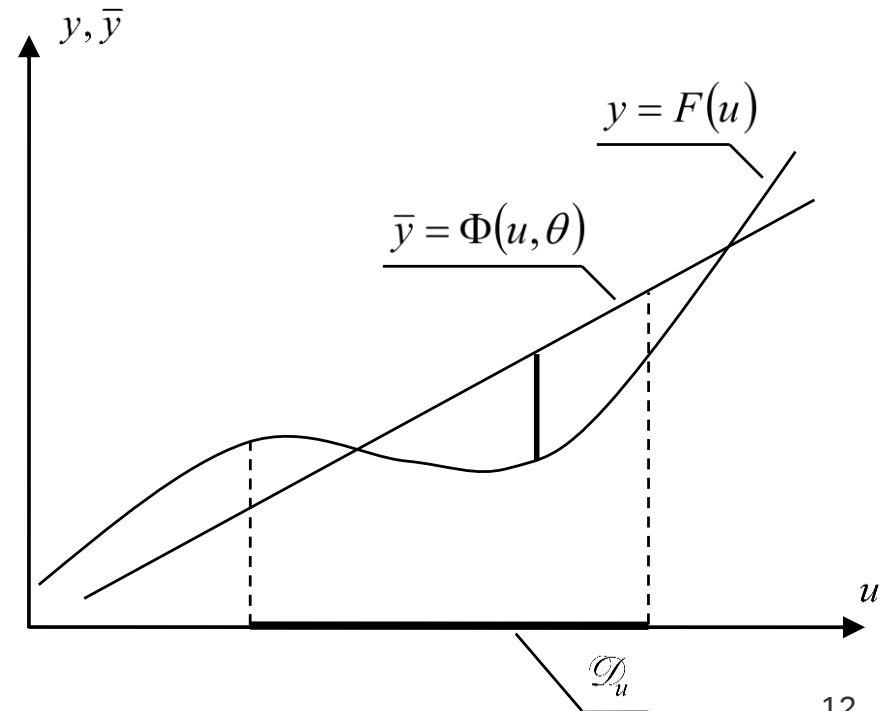
Wskaźnik jakości aproksymacji – miara różnicy pomiędzy funkcją Aproksymowaną $F(u)$ i aproksymującą $\Phi(u, \theta)$ w obszarze $\mathcal{D}_u$

$$Q(\theta) = \|F(u) - \Phi(u, \theta)\|_{\mathcal{D}_u}$$

$$Q(\theta) = \max_{u \in \mathcal{D}_u} \{q(F(u), \Phi(u, \theta)) g_u(u)\}$$

np.:

$$Q(\theta) = \max_{u \in \mathcal{D}_u} \{|F(u) - \Phi(u, \theta)| g_u(u)\}$$

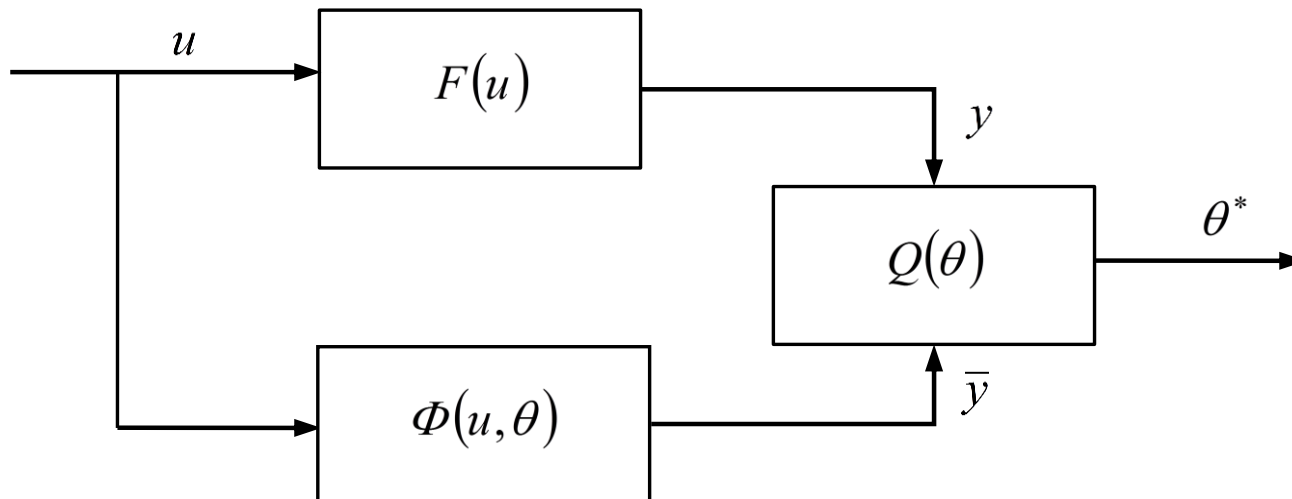


Zadanie aproksymacji charakterystyki statycznej obiektu

∞ Sformułowanie problemu

Optymalny model: $\bar{y} = \Phi(u, \theta^*)$

gdzie: θ^* - optymalny wektor parametrów modelu: $\theta^* \rightarrow Q(\theta^*) = \min_{\theta \in \Theta} Q(\theta)$



Przypadek liniowo - kwadratowy

Model: $\bar{y} = \Phi(u, \theta) = \theta^T \varphi(u)$

gdzie $\theta \stackrel{\text{df}}{=} \begin{bmatrix} \theta^{(1)} \\ \theta^{(2)} \\ \vdots \\ \theta^{(R)} \end{bmatrix}$ $\varphi(u) \stackrel{\text{df}}{=} \begin{bmatrix} \varphi_1(u) \\ \varphi_2(u) \\ \vdots \\ \varphi_R(u) \end{bmatrix}$ np. : $\varphi(u) = \begin{bmatrix} 1 \\ u \\ \vdots \\ u^{R-1} \end{bmatrix}$

Wskaźnik jakości aproksymacji:

$$Q(\theta) = \int_{\mathcal{U}_u} (F(u) - \Phi(u, \theta))^2 g_u(u) du$$

czyli:

$$Q(\theta) = \int_{\mathcal{U}_u} (F(u) - \theta^T \varphi(u))^2 g_u(u) du$$

Przypadek liniowo - kwadratowy

Warunek konieczny optymalności wektora θ :

$$\text{grad}_{\theta} Q(\theta) = -2 \left[\int_{\mathcal{D}_u} (F(u) - \theta^{*T} \varphi(u)) \varphi(u) g_u(u) du \right] = 0_R, \quad 0_R = \left. \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\} R$$

Rozwiązanie powyższego równania jest następujące:

$$\theta^* = \left[\int_{\mathcal{D}_u} \varphi(u) \varphi^T(u) g_u(u) du \right]^{-1} \int_{\mathcal{D}_u} F(u) \varphi(u) g_u(u) du$$

Przypadek liniowo - kwadratowy

Zauważmy że w rozwiązaniu występuje macierz:

$$\int_{\mathcal{U}} \varphi(u) \varphi^T(u) g_u(u) du =$$
$$\begin{bmatrix} \int_{\mathcal{U}} \varphi_1(u) \varphi_1(u) g_u(u) du & \int_{\mathcal{U}} \varphi_1(u) \varphi_2(u) g_u(u) du & \cdots & \int_{\mathcal{U}} \varphi_1(u) \varphi_R(u) g_u(u) du \\ \int_{\mathcal{U}} \varphi_2(u) \varphi_1(u) g_u(u) du & \int_{\mathcal{U}} \varphi_2(u) \varphi_2(u) g_u(u) du & \cdots & \int_{\mathcal{U}} \varphi_2(u) \varphi_R(u) g_u(u) du \\ \vdots & \vdots & \ddots & \vdots \\ \int_{\mathcal{U}} \varphi_R(u) \varphi_1(u) g_u(u) du & \int_{\mathcal{U}} \varphi_R(u) \varphi_2(u) g_u(u) du & \cdots & \int_{\mathcal{U}} \varphi_R(u) \varphi_R(u) g_u(u) du \end{bmatrix}$$

Przypadek liniowo - kwadratowy

Jeżeli wybierzemy funkcje $\varphi_i(u)$ ortonormalne czyli takie że:

$$\int_{\mathcal{U}} \varphi_i(u) \varphi_j(u) g_u(u) du = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}, \quad i, j = 1, 2, \dots, R$$

Wówczas macierz : $\int_{\mathcal{U}} \varphi(u) \varphi^T(u) g_u(u) du = \mathbf{I}$ jest macierzą jednostkową

i algorytm aproksymacji upraszcza się do postaci:

$$\theta^* = \int_{\mathcal{U}} F(u) \varphi(u) g_u(u) du$$

Przypadek liniowo – kwadratowy

Przykład

Charakterystyka statyczna obiektu (funkcja aproksymowana):

$$y = F(u) = u^2$$

Model (funkcja aproksymująca):

$$\bar{y} = \Phi(u, \theta) = \theta^T \varphi(u) = \theta^{(1)} + \theta^{(2)}u \quad \text{gdzie:} \quad \varphi(u) = \begin{bmatrix} 1 \\ u \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta^{(1)} \\ \theta^{(2)} \end{bmatrix}$$

Podzbiór przestrzeni wejść (odcinek):

$$\mathcal{D}_u = \{u \in \mathcal{R} : 0 \leq u \leq 1\}$$

Funkcja wagi:

$$g_u(u) = \begin{cases} u & \text{for } u \in [0, 1] \\ 0 & \text{for } u \notin [0, 1] \end{cases}$$

Wskaźnik jakości aproksymacji:

$$Q(\theta) = \int_{\mathcal{D}_u} (F(u) - \theta^T \varphi(u))^2 g_u(u) du$$

Przypadek liniowo – kwadratowy

Przykład

Wskaźnik jakości aproksymacji przyjmuje postać:

$$Q(\theta) = \int_{\mathcal{D}_u} (F(u) - \theta^T \varphi(u))^2 g_u(u) du = \int_0^1 \left(u^2 - [\theta^{(1)} \ \theta^{(2)}] \begin{bmatrix} 1 \\ u \end{bmatrix} \right)^2 u du$$

Warunek konieczny optymalności jest następujący:

$$\text{grad}_{\theta} Q(\theta) = -2 \left[\int_0^1 \left(u^2 - [\theta^{(1)*} \ \theta^{(2)*}] \begin{bmatrix} 1 \\ u \end{bmatrix} \right) \begin{bmatrix} 1 \\ u \end{bmatrix} u du \right] = 0_2$$

a rozwiązanie ma postać:

$$\theta^* = \left[\int_0^1 \begin{bmatrix} 1 \\ u \end{bmatrix} [1 \ u] u du \right]^{-1} \int_0^1 u^2 \begin{bmatrix} 1 \\ u \end{bmatrix} u du = \begin{bmatrix} \int_0^1 u du & \int_0^1 u^2 du \\ \int_0^1 u^2 du & \int_0^1 u^3 du \end{bmatrix}^{-1} \begin{bmatrix} \int_0^1 u^3 du \\ \int_0^1 u^4 du \end{bmatrix}$$

Przypadek liniowo – kwadratowy

Przykład

Po prostych przekształceniach otrzymujemy:

$$\theta^* = \begin{bmatrix} \theta^{(1)*} \\ \theta^{(2)*} \end{bmatrix} = \begin{bmatrix} \frac{\int_0^1 u^3 du \times \int_0^1 u^3 du - \int_0^1 u^4 du \times \int_0^1 u^2 du}{\int_0^1 u du \times \int_0^1 u^3 du - \left(\int_0^1 u^2 du\right)^2} \\ \frac{\int_0^1 u^4 du \times \int_0^1 u du - \int_0^1 u^3 du \times \int_0^1 u^2 du}{\int_0^1 u du \times \int_0^1 u^3 du - \left(\int_0^1 u^2 du\right)^2} \end{bmatrix} = \begin{bmatrix} \frac{\frac{1}{4}u^4 \Big|_0^1 \times \frac{1}{4}u^4 \Big|_0^1 - \frac{1}{5}u^5 \Big|_0^1 \times \frac{1}{3}u^3 \Big|_0^1}{\frac{1}{2}u^2 \Big|_0^1 \times \frac{1}{4}u^4 \Big|_0^1 - \left(\frac{1}{3}u^3 \Big|_0^1\right)^2} \\ \frac{\frac{1}{5}u^5 \Big|_0^1 \times \frac{1}{2}u^2 \Big|_0^1 - \frac{1}{4}u^4 \Big|_0^1 \times \frac{1}{3}u^3 \Big|_0^1}{\frac{1}{2}u^2 \Big|_0^1 \times \frac{1}{4}u^4 \Big|_0^1 - \left(\frac{1}{3}u^3 \Big|_0^1\right)^2} \end{bmatrix} = \begin{bmatrix} \frac{\frac{1}{4} \times \frac{1}{4} - \frac{1}{5} \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{4} - \left(\frac{1}{3}\right)^2} \\ \frac{\frac{1}{5} \times \frac{1}{2} - \frac{1}{4} \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{4} - \left(\frac{1}{3}\right)^2} \end{bmatrix} = \begin{bmatrix} -0.3 \\ 1.2 \end{bmatrix}$$

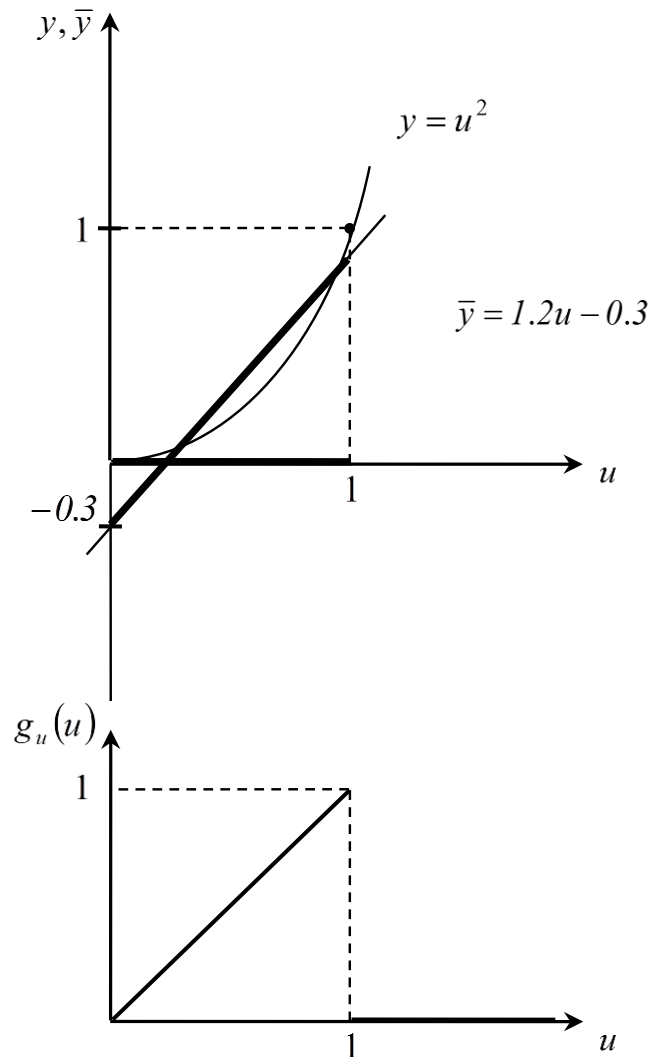
Przypadek liniowo – kwadratowy

Przykład

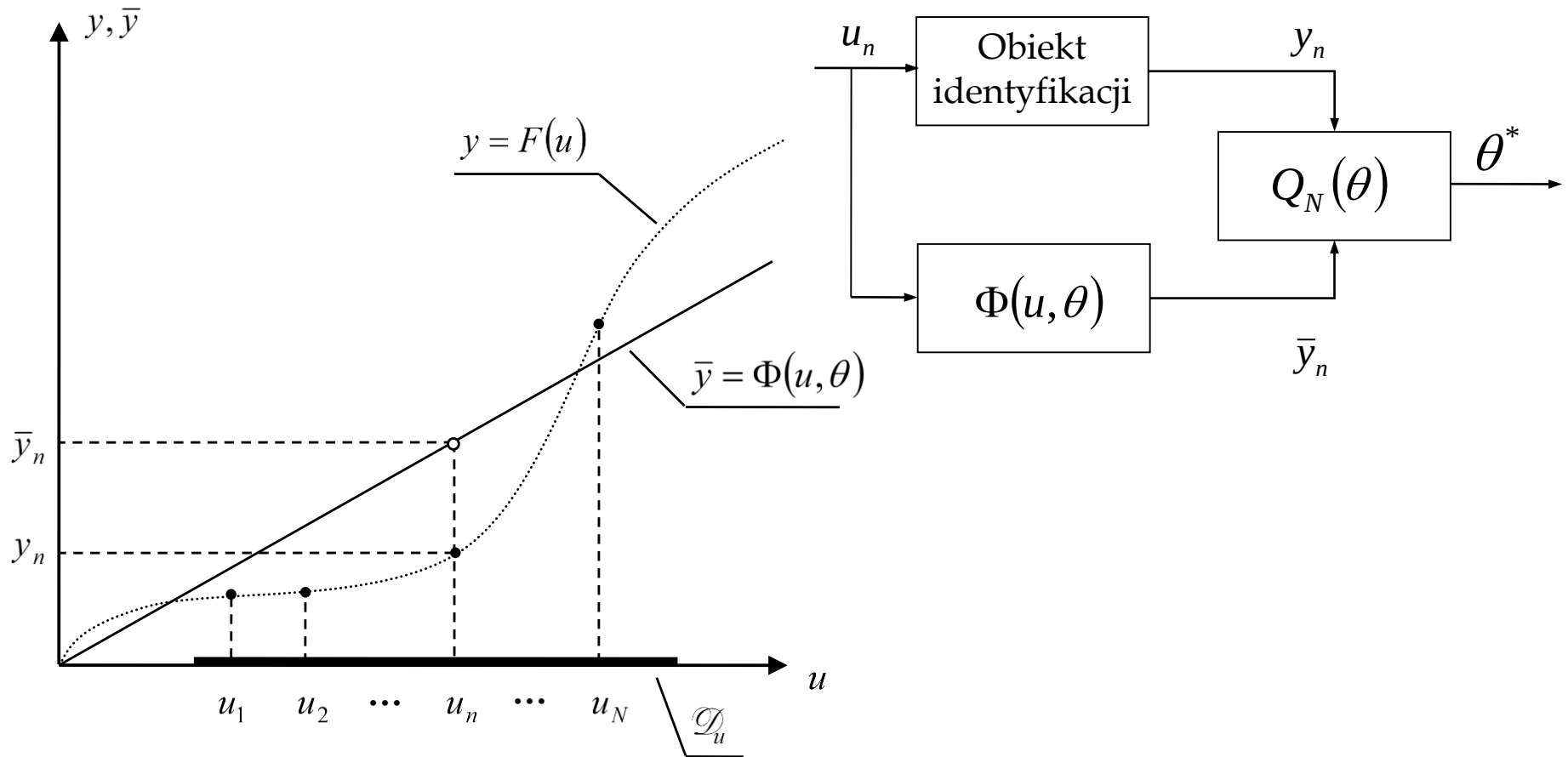
Model optymalny:

$$\bar{y} = 1.2u - 0.3$$

Interpretacja graficzna:



Wybór optymalnego modelu - Przypadek deterministyczny



Wybór optymalnego modelu - Przypadek deterministyczny

∞ Sformułowanie problemu

Eksperyment: $U_N = [u_1 \quad u_2 \quad \cdots \quad u_N]$, $Y_N = [y_1 \quad y_2 \quad \cdots \quad y_N]$

Funkcja aproksymująca (model): $\bar{y} = \Phi(u, \theta)$

Miara różnicy pomiędzy wartością zmierzoną wyjścia obiektu i wartością wyznaczoną z modelu dla tej samej wartości wejścia, tj. dla każdego pomiaru:

$$\forall n = 1, 2, \dots, N \quad q(y_n, \bar{y}_n) = q(y_n, \Phi(u_n, \theta))$$

Wybór optymalnego modelu - Przypadek deterministyczny

∞ Sformułowanie problemu

Wskaźnik jakości identyfikacji $Q_N(\theta) = \|Y_N - \bar{Y}_N(\theta)\|_{U_N}$

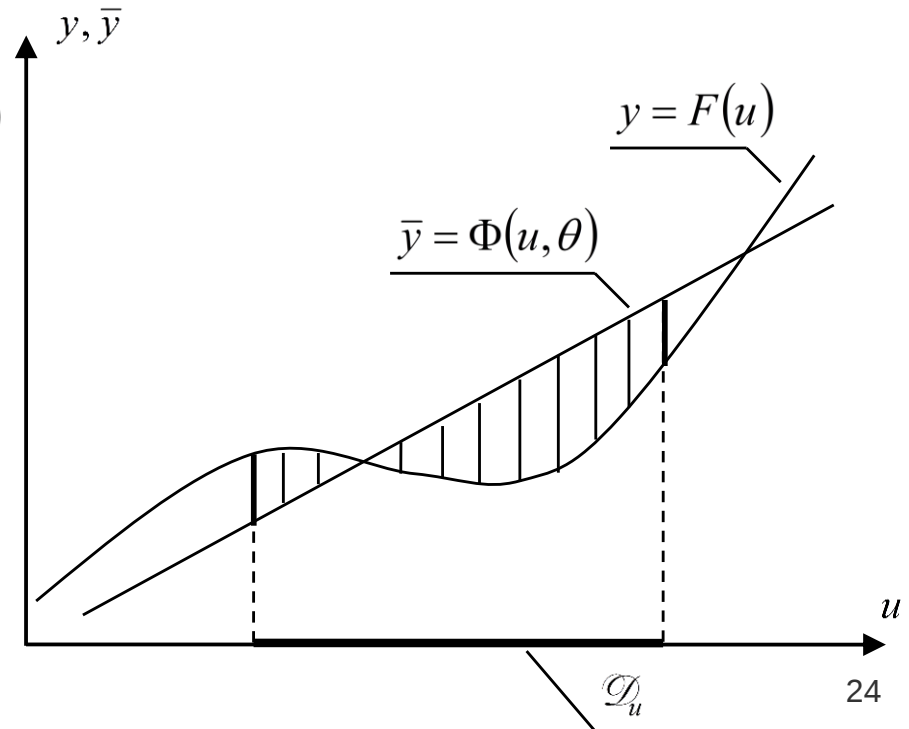
gdzie: $\bar{Y}_N(\theta) \stackrel{df}{=} [\Phi(u_1, \theta) \quad \Phi(u_2, \theta) \quad \dots \quad \Phi(u_N, \theta)]$

$$Q_N(\theta) = \sum_{n=1}^N \alpha_n q(y_n, \bar{y}_n) = \sum_{n=1}^N \alpha_n q(y_n, \Phi(u_n, \theta))$$

np. :

$$Q_N(\theta) = \sum_{n=1}^N |y_n - \bar{y}_n| = \sum_{n=1}^N |y_n - \Phi(u_n, \theta)|$$

$$Q_N(\theta) = \sum_{n=1}^N (y_n - \bar{y}_n)^2 = \sum_{n=1}^N (y_n - \Phi(u_n, \theta))^2$$



Wybór optymalnego modelu - Przypadek deterministyczny

∞ Sformułowanie problemu

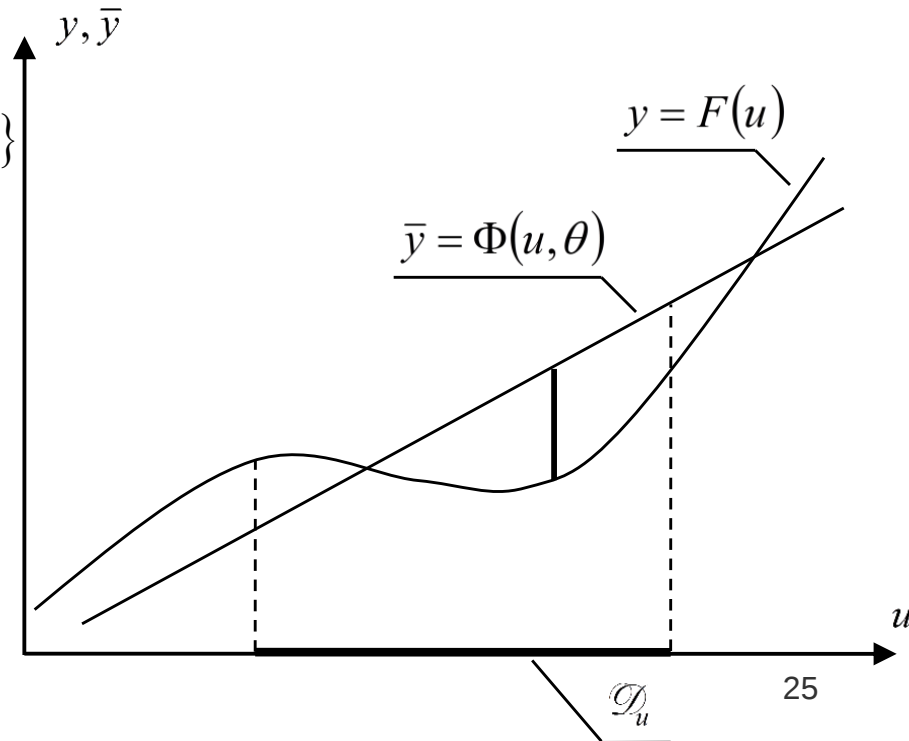
Wskaźnik jakości identyfikacji $Q_N(\theta) = \|Y_N - \bar{Y}_N(\theta)\|_{U_N}$

gdzie: $\bar{Y}_N(\theta) \stackrel{df}{=} [\Phi(u_1, \theta) \quad \Phi(u_2, \theta) \quad \dots \quad \Phi(u_N, \theta)]$

$$Q_N(\theta) = \max_{1 \leq n \leq N} \{ \alpha_n q(y_n, \bar{y}_n) \} = \max_{1 \leq n \leq N} \{ \alpha_n q(y_n, \Phi(u_n, \theta)) \}$$

np. :

$$Q_N(\theta) = \max_{1 \leq n \leq N} \{ |y_n - \bar{y}_n| \} = \max_{1 \leq n \leq N} \{ |y_n - \Phi(u_n, \theta)| \}$$



Wybór optymalnego modelu - Przypadek deterministyczny

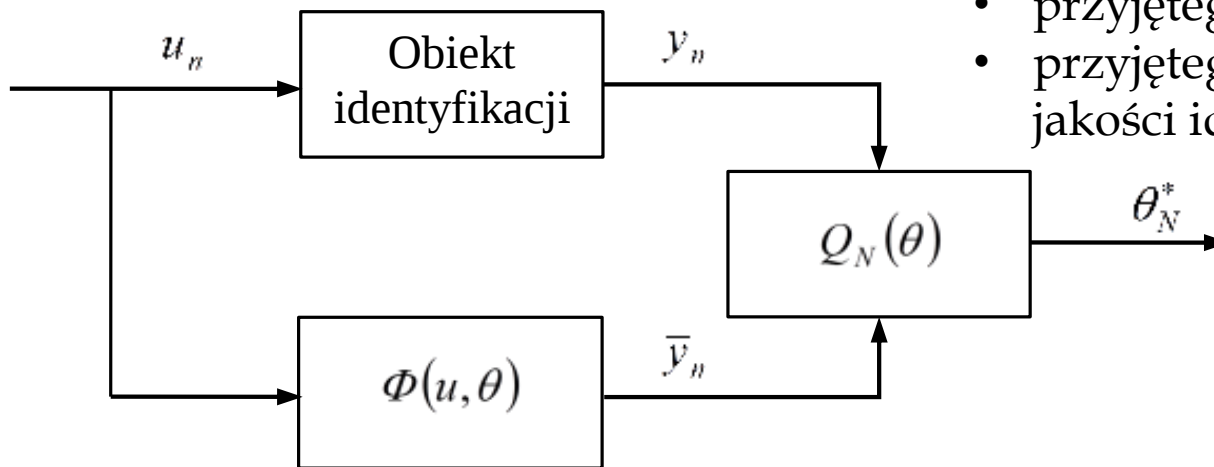
∞ Sformułowanie problemu

Model optymalny: $\bar{y} = \Phi(u, \theta_N^*)$

$$\theta_N^* \rightarrow Q_N(\theta_N^*) = \min_{\theta} Q_N(\theta)$$

Model jest optymalny:

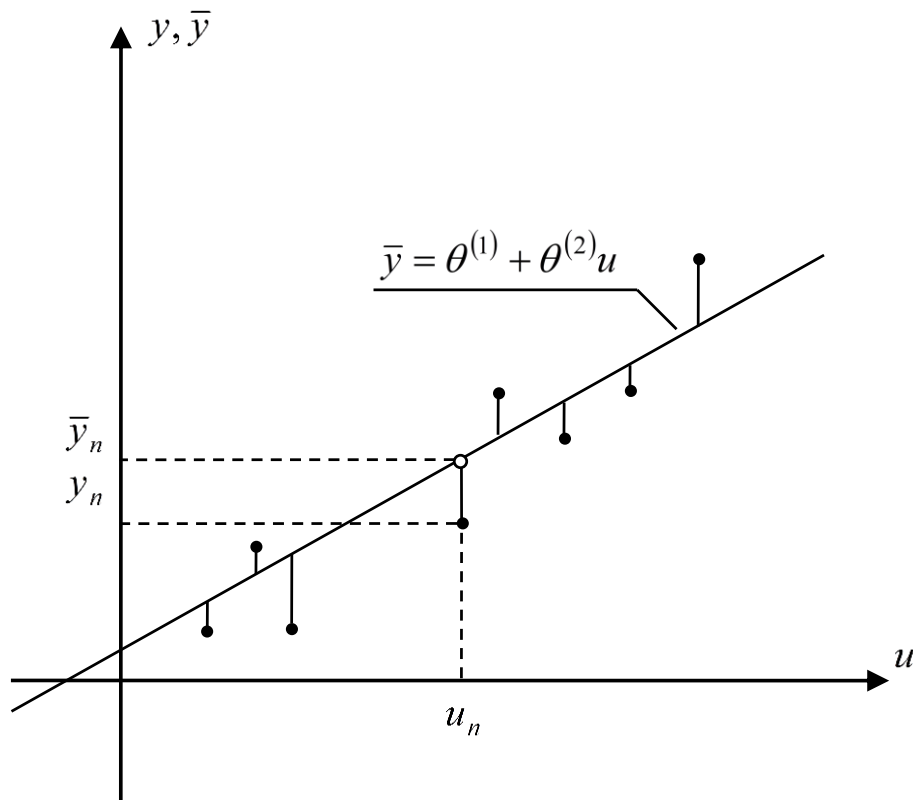
- dla zadanej serii pomiarowej
- przyjętego modelu
- przyjętego wskaźnika jakości identyfikacji



Przypadek liniowo – kwadratowy

Przykład

Eksperyment: $U_N = [u_1 \ u_2 \ \cdots \ u_N]$, $Y_N = [y_1 \ y_2 \ \cdots \ y_N]$



Model: $\bar{y} = \Phi(u, \theta) = \theta^T \varphi(u) = \theta^{(1)} + \theta^{(2)}u$

gdzie $\varphi(u) = \begin{bmatrix} 1 \\ u \end{bmatrix}$, $\theta = \begin{bmatrix} \theta^{(1)} \\ \theta^{(2)} \end{bmatrix}$

Wskaźnik jakości identyfikacji:

$$Q_N(\theta) = \sum_{n=1}^N (y_n - \bar{y}_n)^2 = \sum_{n=1}^N (y_n - \theta^T \varphi(u_n))^2$$

Przypadek liniowo – kwadratowy

Przykład

Warunki konieczne optymalności mają postać

$$\operatorname{grad}_{\theta} Q_N(\theta) = -2 \sum_{n=1}^N (y_n - \theta_N^{*T} \varphi(u_n)) \varphi(u_n) = 0_2$$

Rozwiązanie powyższego równania przyjmuje postać:

$$\theta_N^* = \left[\sum_{n=1}^N \varphi(u_n) \varphi^T(u_n) \right]^{-1} \times \sum_{n=1}^N y_n \varphi(u_n)$$

a po przekształceniach:

$$\theta_N^* = \begin{bmatrix} \theta_N^{*(1)} \\ \theta_N^{*(2)} \end{bmatrix} = \begin{bmatrix} \frac{\left(\sum_{n=1}^N u_n^2 \right) \times \left(\sum_{n=1}^N y_n \right) - \left(\sum_{n=1}^N u_n \right) \times \left(\sum_{n=1}^N y_n u_n \right)}{N \sum_{n=1}^N u_n^2 - \left(\sum_{n=1}^N u_n \right) \times \left(\sum_{n=1}^N y_n u_n \right)} \\ \frac{N \sum_{n=1}^N y_n u_n - \left(\sum_{n=1}^N u_n \right) \times \left(\sum_{n=1}^N y_n \right)}{N \sum_{n=1}^N u_n^2 - \left(\sum_{n=1}^N u_n \right) \times \left(\sum_{n=1}^N y_n u_n \right)} \end{bmatrix}$$

Przypadek liniowo – kwadratowy

Algorytm rekurencyjny

$$Q_N(\theta) = \sum_{n=1}^N (y_n - \bar{y}_n)^2 = \sum_{n=1}^N (y_n - \theta^T \varphi(u_n))^2$$

Wektor parametrów θ_N^* wyznaczamy na podstawie N pomiarów

$$\theta_N^* = \left[\sum_{n=1}^N \varphi(u_n) \varphi^T(u_n) \right]^{-1} \times \sum_{n=1}^N y_n \varphi(u_n)$$

Pojawia się nowy $N+1$ pomiar u_{N+1} , y_{N+1}

Jak wyznaczyć (poprawić) parametr korzystając z nowego pomiaru?

$$\theta_{N+1}^* = A(\theta_N^*, u_{N+1}, y_{N+1})$$
$$Q_{N+1}(\theta) = \sum_{n=1}^{N+1} (y_n - \bar{y}_n)^2 = \sum_{n=1}^{N+1} (y_n - \theta^T \varphi(u_n))^2$$

$$\theta_{N+1}^* = \left[\sum_{n=1}^{N+1} \varphi(u_n) \varphi^T(u_n) \right]^{-1} \times \sum_{n=1}^{N+1} y_n \varphi(u_n) =$$

$$\left[\sum_{n=1}^N \varphi(u_n) \varphi^T(u_n) + \varphi(u_{N+1}) \varphi^T(u_{N+1}) \right]^{-1} \times \left[\sum_{n=1}^N y_n \varphi(u_n) + y_{N+1} \varphi(u_{N+1}) \right]$$

Przypadek liniowo – kwadratowy

Algorytm rekurencyjny

Przydatne przekształcenia algebraiczne:

$$(\mathbf{A} + \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}$$

$\mathbf{B} = \varphi$ – wektor kolumnowy

$$\mathbf{D}^{-1} = 1$$

$$\mathbf{C} = \varphi^T$$

$$(\mathbf{A} + \varphi\varphi^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\varphi(1 + \varphi^T\mathbf{A}^{-1}\varphi)^{-1}\varphi^T\mathbf{A}^{-1}$$

$$\mathbf{D}^{-1} = -1$$

$$(\mathbf{A} - \varphi\varphi^T)^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\varphi(1 - \varphi^T\mathbf{A}^{-1}\varphi)^{-1}\varphi^T\mathbf{A}^{-1}$$

Przypadek liniowo – kwadratowy

Algorytm rekurencyjny

Oznaczmy przez: $\varphi_n = \varphi(u_n)$, $n = 1, 2, \dots, N + 1$

$$\begin{aligned} \left(\sum_{n=1}^{N+1} \varphi_n \varphi_n^T \right)^{-1} &= \left(\sum_{n=1}^N \varphi_n \varphi_n^T + \varphi_{N+1} \varphi_{N+1}^T \right)^{-1} = \\ &= \left(\sum_{n=1}^N \varphi_n \varphi_n^T \right)^{-1} - \left(\sum_{n=1}^N \varphi_n \varphi_n^T \right)^{-1} \frac{\varphi_{N+1} \varphi_{N+1}^T}{1 + \varphi_{N+1}^T \left(\sum_{n=1}^N \varphi_n \varphi_n^T \right)^{-1} \varphi_{N+1}} \left(\sum_{n=1}^N \varphi_n \varphi_n^T \right)^{-1} \end{aligned}$$

Niech: $P_N = \left(\sum_{n=1}^N \varphi_n \varphi_n^T \right)^{-1}$

Powyższe przyjmuje postać:

$$P_{N+1} = P_N - P_N \frac{\varphi_{N+1} \varphi_{N+1}^T}{1 + \varphi_{N+1}^T P_N \varphi_{N+1}} P_N$$

Przypadek liniowo – kwadratowy

Algorytm rekurencyjny

$$\begin{aligned}\theta_{N+1}^* &= \left(P_N - P_N \frac{\varphi_{N+1} \varphi_{N+1}^T}{1 + \varphi_{N+1}^T P_N \varphi_{N+1}} P_N \right) \left(\sum_{n=1}^N \varphi_n y_n + \varphi_{N+1} y_{N+1} \right) = \\&= P_N \sum_{n=1}^N \varphi_n y_n + P_N \varphi_{N+1} y_{N+1} - \frac{P_N \varphi_{N+1} \varphi_{N+1}^T P_N}{1 + \varphi_{N+1}^T P_N \varphi_{N+1}} \sum_{n=1}^N \varphi_n y_n - \frac{P_N \varphi_{N+1} \varphi_{N+1}^T P_N \varphi_{N+1} y_{N+1}}{1 + \varphi_{N+1}^T P_N \varphi_{N+1}} = \\&= P_N \sum_{n=1}^N \varphi_n y_n + \frac{P_N \varphi_{N+1} y_{N+1} + P_N \varphi_{N+1} y_{N+1} \varphi_{N+1}^T P_N \varphi_{N+1} - P_N \varphi_{N+1} y_{N+1} \varphi_{N+1}^T P_N \varphi_{N+1} - P_N \varphi_{N+1} \varphi_{N+1}^T P_N \sum_{n=1}^N \varphi_n y_n}{1 + \varphi_{N+1}^T P_N \varphi_{N+1}} = \\&= P_N \sum_{n=1}^N \varphi_n y_n + \frac{P_N \varphi_{N+1} y_{N+1} - P_N \varphi_{N+1} \varphi_{N+1}^T P_N \sum_{n=1}^N \varphi_n y_n}{1 + \varphi_{N+1}^T P_N \varphi_{N+1}} y = \\&= P_N \sum_{n=1}^N \varphi_n y_n + \frac{P_N \varphi_{N+1} y_{N+1}}{1 + \varphi_{N+1}^T P_N \varphi_{N+1}} \left(y_{N+1} + \varphi_{N+1}^T P_N \sum_{n=1}^N \varphi_n y_n \right) \\&= \theta_N^* + K_{N+1} (y_{N+1} - \varphi_{N+1}^T \theta_N)\end{aligned}$$

gdzie: $K_{N+1} = \frac{P_N \varphi_{N+1}}{1 + \varphi_{N+1}^T P_N \varphi_{N+1}}$

Przypadek liniowo – kwadratowy

Algorytm rekurencyjny

Podsumowując, otrzymujemy następujący algorytm

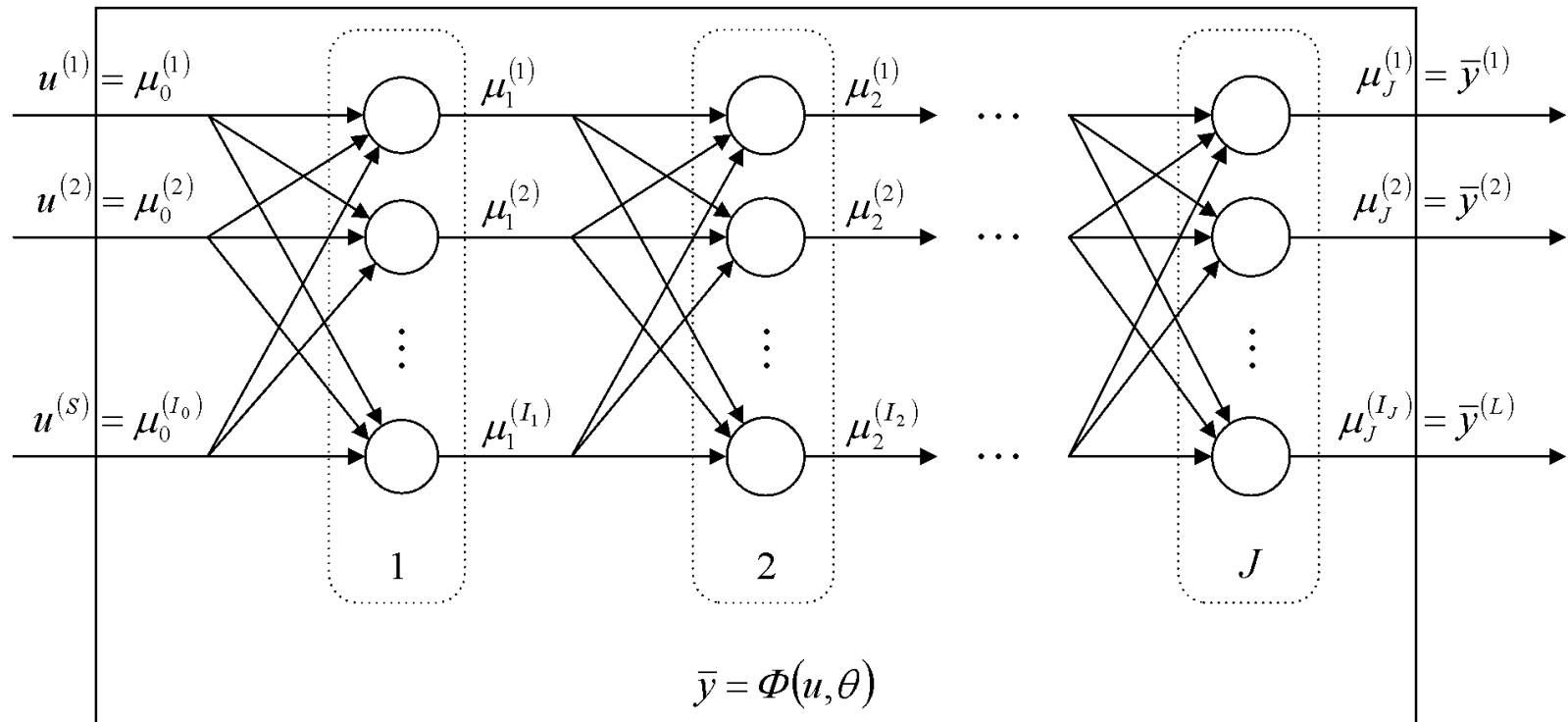
$$\theta_N^* = \left(\sum_{n=1}^N \varphi_n \varphi_n^T \right)^{-1} \sum_{n=1}^N \varphi_n y_n = P_N \sum_{n=1}^N \varphi_n y_n, \text{ gdzie: } P_N = \left(\sum_{n=1}^N \varphi_n \varphi_n^T \right)^{-1}$$

$$\theta_{N+1}^* = A(\theta_N, u_{N+1}, y_{N+1}) = \theta_N^* + K_{N+1} [y_{N+1} - \varphi(u_{N+1})^T \theta_N^*]$$

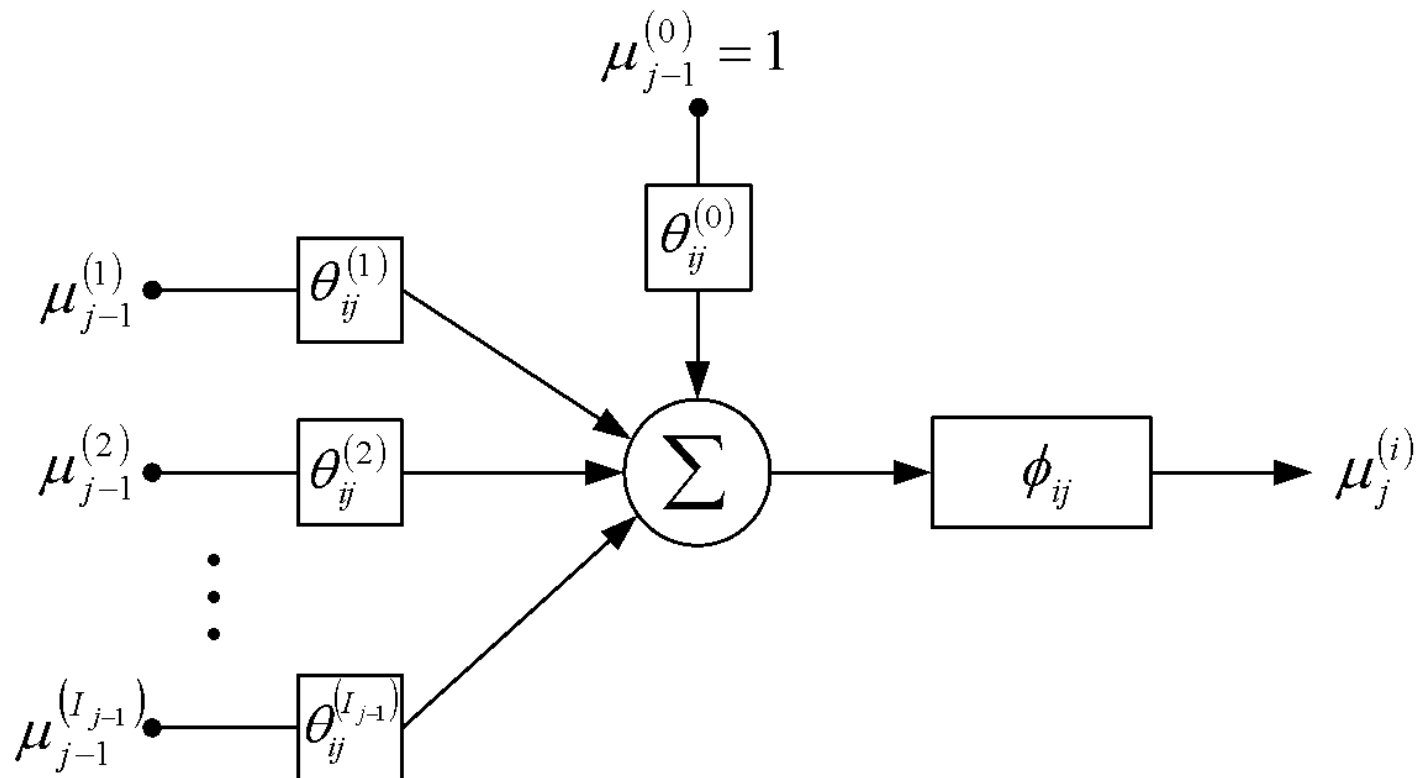
$$K_{N+1} = \frac{P_N \varphi(u_{N+1})}{1 + \varphi(u_{N+1})^T P_N \varphi(u_{N+1})}$$

$$P_{N+1} = P_N - \frac{P_N \varphi(u_{N+1}) \varphi(u_{N+1})^T P_N}{1 + \varphi(u_{N+1})^T P_N \varphi(u_{N+1})}$$

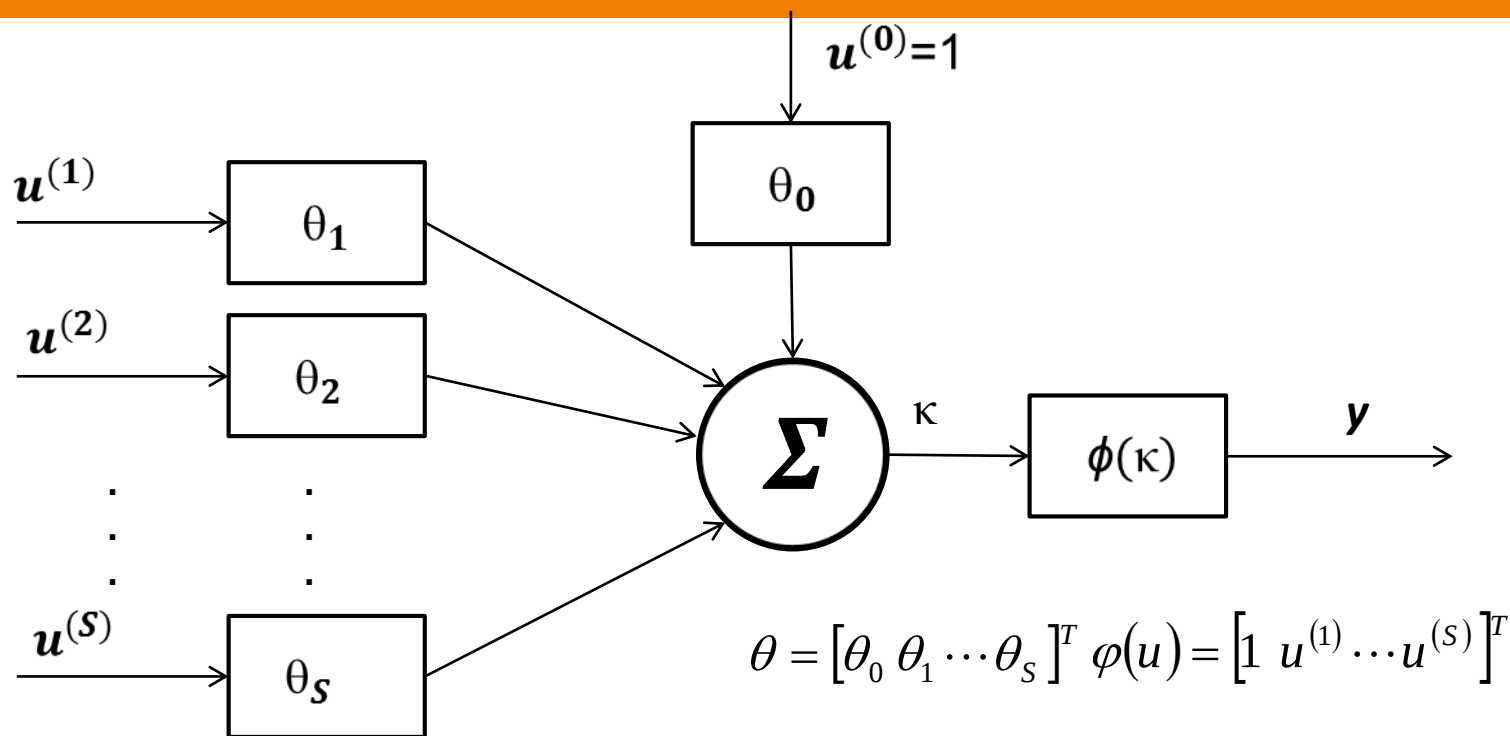
Wybór optymalnego modelu, a sieci neuronowe – szczególny przypadek



Model neuronu (wyrwany z sieci)



Model neuronu (uproszczenie)

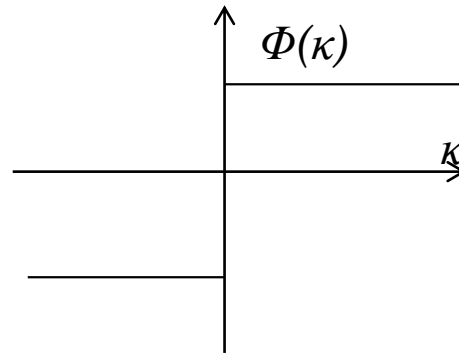


$$y = \phi\left(\sum_{s=1}^s \theta_s u^{(s)} + \theta_0\right) = \phi(\theta^T \varphi(u)) \quad \Phi - \text{funkcja aktywacji}$$

Funkcja aktywacji

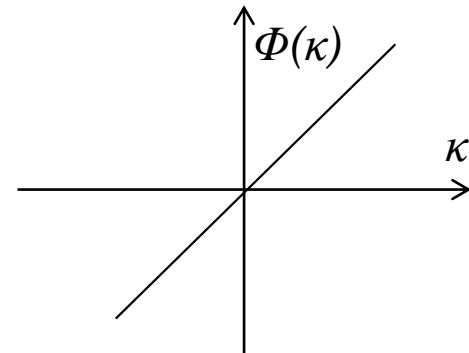
$$\phi(\kappa) = \begin{cases} 1 & \text{dla } \kappa > 0 \\ -1 & \text{dla } \kappa \leq 0 \end{cases}$$

Perceptron



$$\phi(\kappa) = \kappa$$

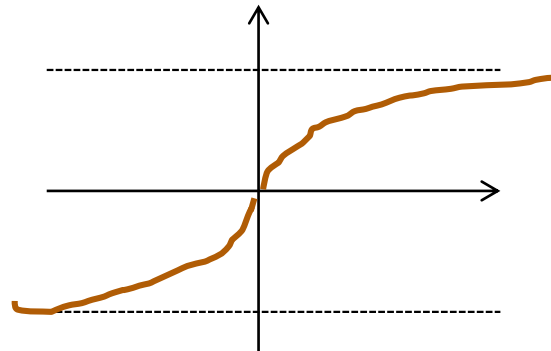
Adaline
Adaptive Linear Neuron



$$\phi(\kappa) = \frac{1}{1 + e^{-\beta\kappa}}$$

$$\phi(\kappa) = \tanh(\beta\kappa) = \frac{1 - e^{-\beta\kappa}}{1 + e^{-\beta\kappa}}$$

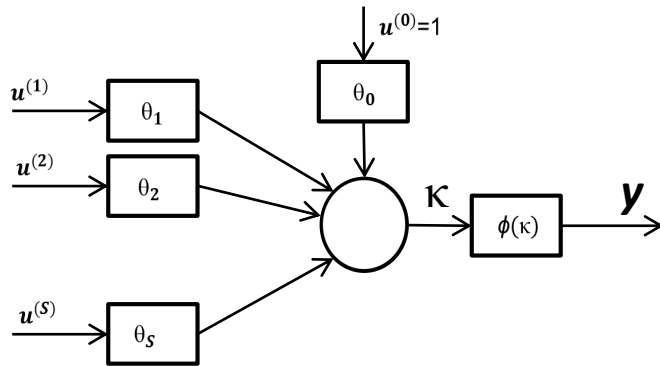
Neuron sigmoidalny



Uczenie neuronu

Neuron Adaline (Adaptive Linear Neuron)

$$\phi(\kappa) = \kappa$$



$$y = \phi\left(\sum_{s=1}^S \theta_s u^{(s)} + \theta_0\right) = \phi(\theta^T \varphi(u)) = \theta^T \varphi(u)$$

$$\theta = [\theta_0 \ \theta_1 \ \dots \ \theta_S]^T \quad \varphi(u) = [1 \ u^{(1)} \ \dots \ u^{(S)}]^T$$

Uczenie: $[u_1 \ u_2 \ \dots \ u_N] = U_N \quad [y_1 \ y_2 \ \dots \ y_N] = Y_N$

$$Q_N(\theta) = \sum_{n=1}^N (y_n - \bar{y}_n)^2 = \sum_{n=1}^N (y_n - \theta^T \varphi(u_n))^2 \quad \text{To już było}$$

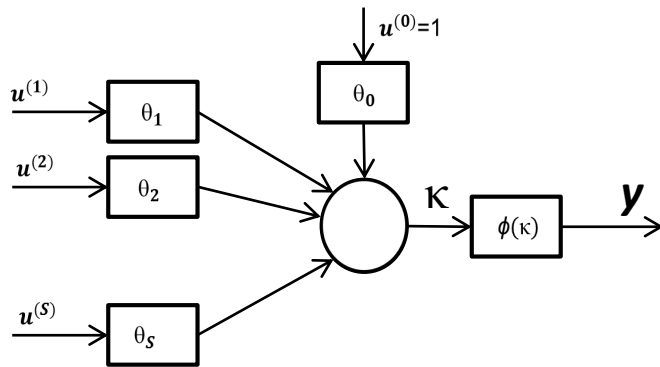
$$\theta_N^* = \left[\sum_{n=1}^N \varphi(u_n) \varphi^T(u_n) \right]^{-1} \times \sum_{n=1}^N y_n \varphi(u_n)$$

lub rekurencyjnie

Uczenie neuronu

Neuron Adaline (Adaptive Linear Neuron)

$$\phi(\kappa) = \kappa$$



$$y = \phi\left(\sum_{s=1}^S \theta_s u^{(s)} + \theta_0\right) = \phi(\theta^T \varphi(u)) = \theta^T \varphi(u)$$

$$\theta = [\theta_0 \ \theta_1 \ \dots \ \theta_S]^T \quad \varphi(u) = [1 \ u^{(1)} \ \dots \ u^{(S)}]^T$$

Uczenie: $[u_1 \ u_2 \ \dots \ u_N] = U_N \quad [y_1 \ y_2 \ \dots \ y_N] = Y_N$

$$Q_N(\theta) = \sum_{n=1}^N (y_n - \bar{y}_n)^2 = \sum_{n=1}^N (y_n - \theta^T \varphi(u_n))^2 \quad \text{To już było}$$

$$\theta_N^* = \left[\sum_{n=1}^N \varphi(u_n) \varphi^T(u_n) \right]^{-1} \times \sum_{n=1}^N y_n \varphi(u_n)$$

lub rekurencyjnie

Algortm rekurencyjny

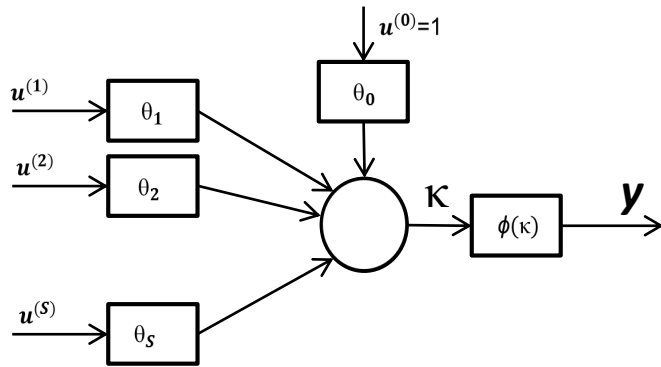
$$\theta_{N+1}^* = A(\theta_N^*, u_{N+1}, y_{N+1}) = \theta_N^* + K_{N+1} [y_{N+1} - \varphi(u_{N+1})^T \theta_N^*]$$

$$K_{N+1} = \frac{P_N \varphi(u_{N+1})}{1 + \varphi(u_{N+1})^T P_N \varphi(u_{N+1})}$$

$$P_{N+1} = P_N - \frac{P_N \varphi(u_{N+1}) \varphi(u_{N+1})^T P_N}{1 + \varphi(u_{N+1})^T P_N \varphi(u_{N+1})}$$

Uczenie neuronu

Podobnie jak neuron Adaline (Adaptive Linear Neuron) $\phi(\kappa)$ - wzajemnie jednoznaczna



$$y = \phi\left(\sum_{s=1}^S \theta_s u^{(s)} + \theta_0\right) = \phi(\theta^T \varphi(u))$$

$$\theta = [\theta_0 \ \theta_1 \ \dots \ \theta_s]^T \quad \varphi(u) = [1 \ u^{(1)} \ \dots \ u^{(s)}]^T$$

Uczenie: $[u_1 \ u_2 \ \dots \ u_N] = U_N \quad [y_1 \ y_2 \ \dots \ y_N] = Y_N$

$$Q_N(\theta) = \sum_{n=1}^N (\kappa_n - \bar{\kappa}_n)^2 = \sum_{n=1}^N (\phi^{-1}(y_n) - \theta^T \varphi(u_n))^2 \quad \text{To już było}$$

$$\theta_N^* = \left[\sum_{n=1}^N \varphi(u_n) \varphi^T(u_n) \right]^{-1} \times \sum_{n=1}^N \phi^{-1}(y_n) \varphi(u_n)$$

lub rekurencyjnie

Algorytm rekurencyjny

$$\theta_{N+1}^* = A(\theta_N^*, u_{N+1}, y_{N+1}) = \theta_N^* + K_{N+1} [\phi^{-1}(y_{N+1}) - \phi(u_{N+1})^T \theta_N^*]$$

$$K_{N+1} = \frac{P_N \phi(u_{N+1})}{1 + \phi(u_{N+1})^T P_N \phi(u_{N+1})}$$

$$P_{N+1} = P_N - \frac{P_N \phi(u_{N+1}) \phi(u_{N+1})^T P_N}{1 + \phi(u_{N+1})^T P_N \phi(u_{N+1})}$$

Uczenie neuronu

reguła delta

$$Q(\theta) = \frac{1}{2} (y - \theta^T \varphi(u))^2, \quad \theta_0 \quad \text{Neuron adaline}$$

Numeryczna metoda optymalizacji (to będzie)

$$\theta^* \rightarrow Q(\theta^*) = \min_{\theta} Q(\theta)$$

$$\theta_{n+1}^* = \theta_n^* - \eta \nabla_{\theta} Q(\theta_n^*), \quad \theta_0^* \quad \eta - \text{współczynnik uczenia}$$

$$\nabla Q(\theta) = -(y - \theta^T \varphi(u)) \varphi(u), \quad \theta_0$$

$$\theta_{n+1}^* = \theta_n^* + \eta (y_{n+1} - \theta_n^{*T} \varphi(u_{n+1})) \varphi(u_{n+1}), \quad \theta_0^*$$

$$\Delta_n = (y_{n+1} - \theta_n^{*T} \varphi(u_{n+1}))$$

$$\theta_{n+1}^* = \theta_n^* + \eta \Delta_n \varphi(u_{n+1})$$

Uczenie neuronu

reguła delta

$$Q(\theta) = \frac{1}{2} (y - \phi(\theta^T \varphi(u)))^2, \quad \theta_0 \quad \phi - \text{funkcja aktywacji, różniczkowalna}$$

Numeryczna metoda optymalizacji (to będzie)

$$\theta^* \rightarrow Q(\theta^*) = \min_{\theta} Q(\theta)$$

$$\theta_{n+1}^* = \theta_n^* - \eta \nabla_{\theta} Q(\theta_n^*), \quad \theta_0^* \quad \eta - \text{współczynnik uczenia}$$

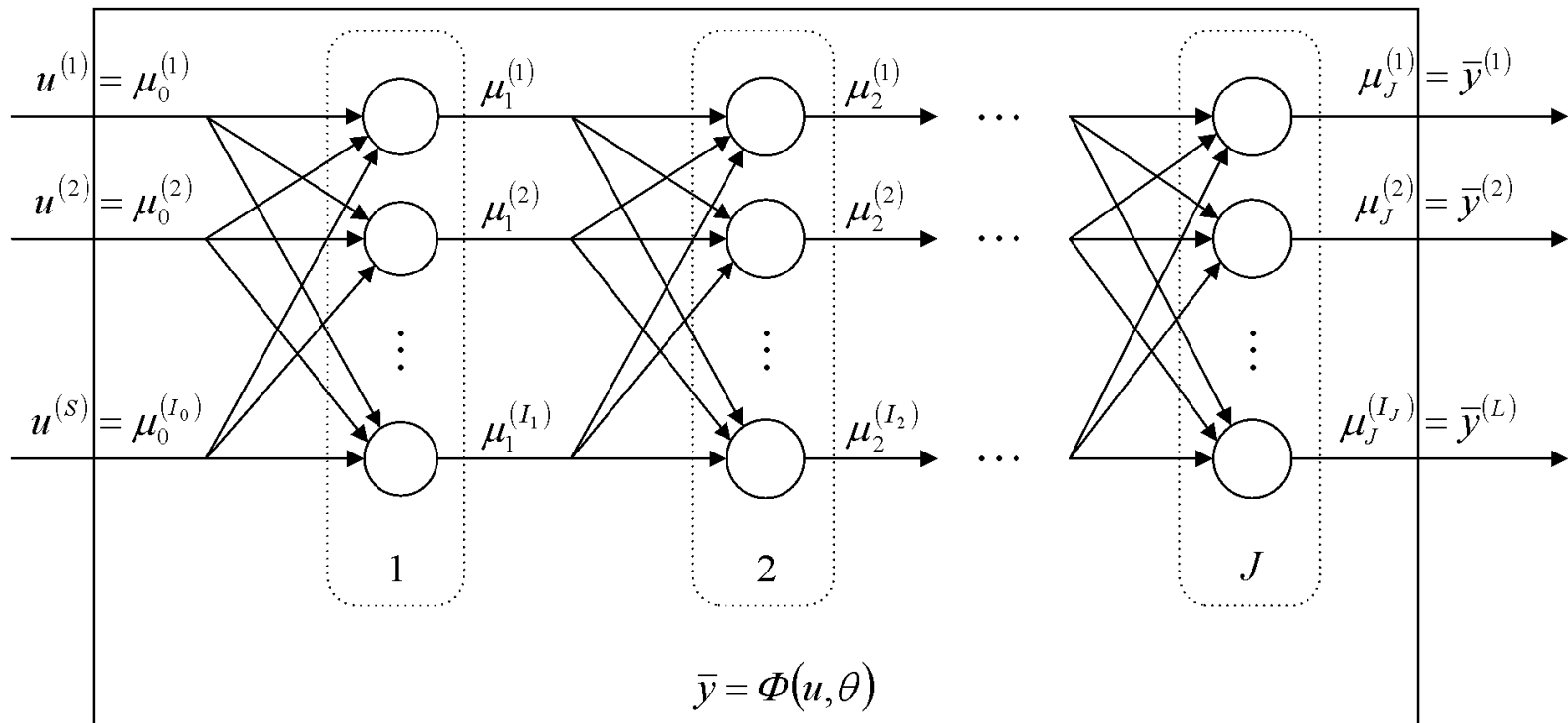
$$\nabla Q(\theta) = -(y - \phi(\theta^T \varphi(u))) \frac{\partial \phi(\kappa)}{\partial \kappa} \varphi(u), \quad \theta_0$$

$$\theta_{n+1}^* = \theta_n^* + \eta (y_{n+1} - \phi(\theta_n^{*T} \varphi(u_{n+1}))) \frac{\partial \phi(\kappa)}{\partial \kappa} \varphi(u_{n+1}), \quad \theta_0^*$$

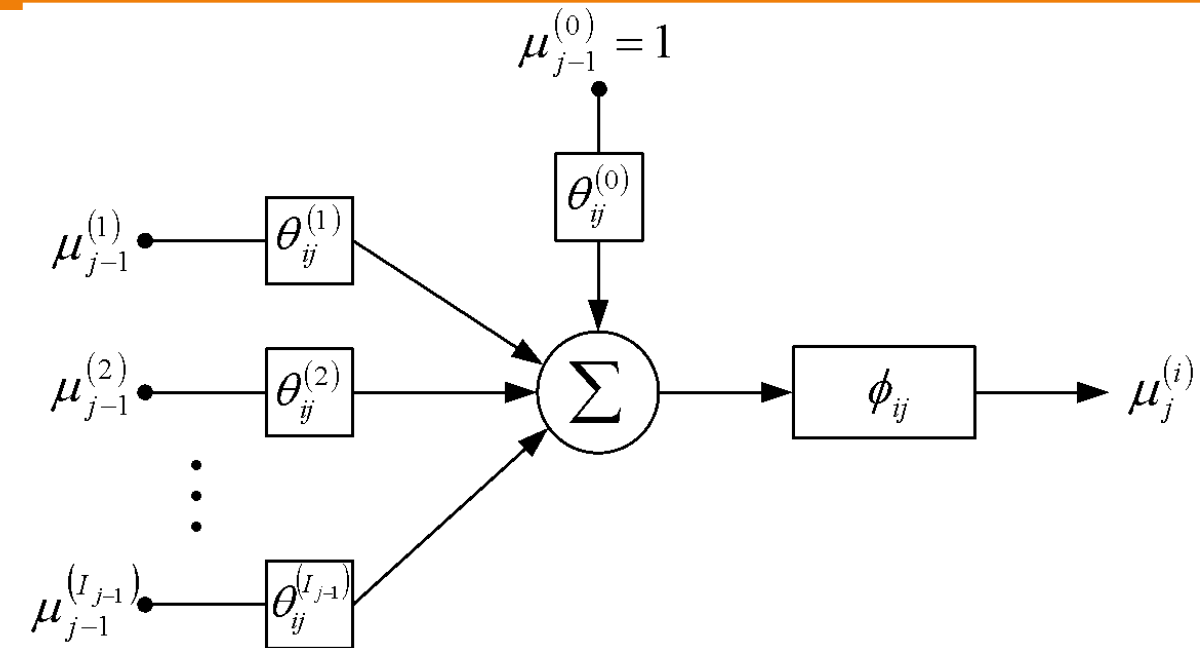
$$\Delta_n = (y_{n+1} - \phi(\theta_n^{*T} \varphi(u_{n+1}))) \frac{\partial \phi(\kappa)}{\partial \kappa}$$

$$\theta_{n+1}^* = \theta_n^* + \eta \Delta_n \varphi(u_{n+1})$$

Sieć wielowarstwowa



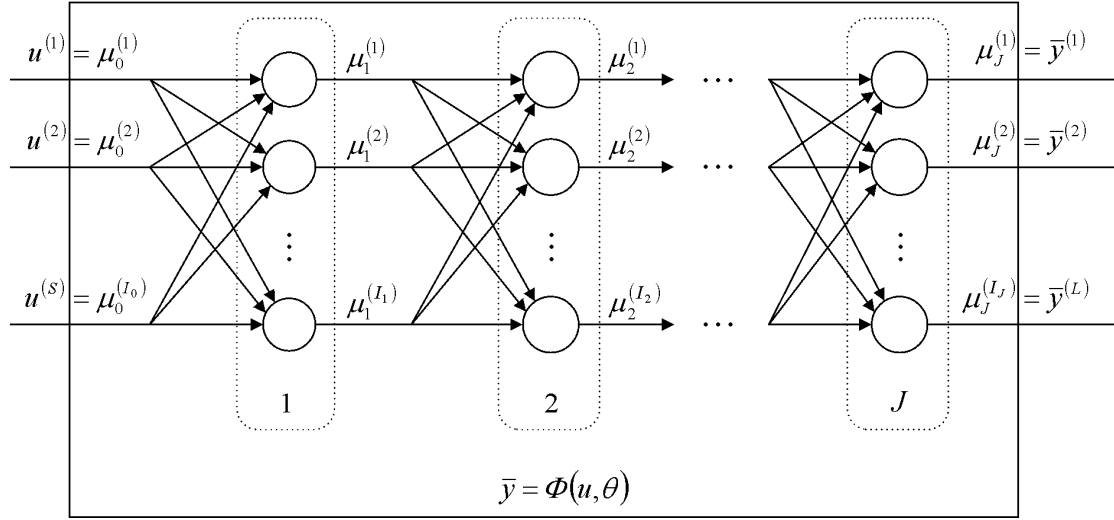
Model neuronu (wyrwany z sieci)



$$\mu_j^{(i)} = \phi_{ij}(\bar{\mu}_{j-1}^T \theta_{ij}).$$

$$\mu_j^{(i)} = \phi_{ij} \left(\sum_{s=1}^{I_{j-1}} \mu_{j-1}^{(s)} \theta_{ij}^{(s)} + \theta_{ij}^{(0)} \right), \quad \mu_{j-1} = \begin{bmatrix} \mu_{j-1}^{(1)} \\ \mu_{j-1}^{(2)} \\ \vdots \\ \mu_{j-1}^{(I_{j-1})} \end{bmatrix}, \quad \bar{\mu}_{j-1} \stackrel{\text{df}}{=} \begin{bmatrix} \mu_{j-1}^{(0)} \\ \mu_{j-1} \end{bmatrix} = \begin{bmatrix} \mu_{j-1}^{(0)} \\ \mu_{j-1}^{(1)} \\ \mu_{j-1}^{(2)} \\ \vdots \\ \mu_{j-1}^{(I_{j-1})} \end{bmatrix}, \quad \theta_{ij} \stackrel{\text{df}}{=} \begin{bmatrix} \theta_{ij}^{(0)} \\ \theta_{ij}^{(1)} \\ \vdots \\ \theta_{ij}^{(I_{j-1})} \end{bmatrix},$$

Sieć wielowarstwowa



$$\mu_j = \begin{bmatrix} \mu_j^{(1)} \\ \mu_j^{(2)} \\ \vdots \\ \mu_j^{(I_j)} \end{bmatrix} = \begin{bmatrix} \phi_{1j}(\bar{\mu}_{j-1}^T \theta_{1j}) \\ \phi_{2j}(\bar{\mu}_{j-1}^T \theta_{2j}) \\ \vdots \\ \phi_{I_{jj}}(\bar{\mu}_{j-1}^T \theta_{I_{jj}}) \end{bmatrix} \stackrel{\text{df}}{=} \phi_j(\bar{\mu}_{j-1}, \theta_j),$$

$$\theta_j \stackrel{\text{df}}{=} \begin{bmatrix} \theta_{1j} & \theta_{2j} & \dots & \theta_{I_{jj}} \end{bmatrix} = \begin{bmatrix} \theta_{1j}^{(0)} & \theta_{2j}^{(0)} & \dots & \theta_{I_{jj}}^{(0)} \\ \theta_{1j}^{(1)} & \theta_{2j}^{(1)} & \dots & \theta_{I_{jj}}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{1j}^{(I_{j-1})} & \theta_{2j}^{(I_{j-1})} & \dots & \theta_{I_{jj}}^{(I_{j-1})} \end{bmatrix}.$$

$$\mu_j = \phi_j(\bar{\mu}_{j-1}, \theta_j), \quad j = 1, 2, \dots, J.$$

$$\mu_1 = \phi_1(\bar{\mu}_0, \theta_1) = \phi_1(u, \theta_1).$$

$$\mu_j = \phi_j(\bar{\mu}_{j-1}, \theta_j), \quad j = 2, 3, \dots, J-1.$$

$$\bar{y} = \mu_J = \phi_J(\bar{\mu}_{J-1}, \theta_J).$$

$$\bar{y} = \begin{bmatrix} \bar{y}^{(1)} \\ \bar{y}^{(2)} \\ \vdots \\ \bar{y}^{(L)} \end{bmatrix} = \begin{bmatrix} \mu_J^{(1)} \\ \mu_J^{(2)} \\ \vdots \\ \mu_J^{(I_J)} \end{bmatrix} = \phi_J(\phi_{J-1}(\dots \phi_1(u, \theta_1), \dots, \theta_{J-1}), \theta_J) \stackrel{\text{df}}{=} \begin{bmatrix} \Phi^{(1)}(u, \theta) \\ \Phi^{(2)}(u, \theta) \\ \vdots \\ \Phi^{(L)}(u, \theta) \end{bmatrix} \stackrel{\text{df}}{=} \Phi(u, \theta),$$

$$\theta \stackrel{\text{df}}{=} \{\theta_1, \theta_2, \dots, \theta_J\}.$$

Uczenie sieci wielowarstwowej

$$Q_n(\theta) = [y_n - \Phi(u_n, \theta)]^T [y_n - \Phi(u_n, \theta)] = \sum_{l=1}^L (y_n^{(l)} - \Phi^{(l)}(u_n, \theta))^2 = \sum_{l=1}^L \varepsilon_n^{(l)2} \quad \varepsilon_n^{(l)} \stackrel{\text{df}}{=} y_n^{(l)} - \bar{y}_n^{(l)} = y_n^{(l)} - \Phi^{(l)}(u_n, \theta), \quad l=1, 2, \dots, L$$

$$\left. \frac{\partial Q_n(\theta)}{\partial \theta_{ij}^{(s)}} \right|_{\theta_{ij}^{(s)} = \tilde{\theta}_{ij,n}^{(s)}} = -2\delta_{jn}^{(i)} \mu_{j-1,n}^{(s)}, \quad \tilde{\theta}_{ij,n+1}^{(s)} = \tilde{\theta}_{ij,n}^{(s)} - \eta_n \left. \frac{\partial Q_n(\theta)}{\partial \theta_{ij}^{(s)}} \right|_{\theta_{ij}^{(s)} = \tilde{\theta}_{ij,n}^{(s)}}, \quad s=0, 1, \dots, I_{J-1}, \quad i=1, 2, \dots, I_J, \quad j=1, 2, \dots, J,$$

$$\varepsilon_{jn}^{(i)} \stackrel{\text{df}}{=} \begin{cases} \varepsilon_n^{(i)} & \text{dla } j = J \\ \sum_{p=1}^{I_{j+1}} \delta_{j+1,n}^{(p)} \tilde{\theta}_{pj+1,n}^{(i)} & \text{dla } j = J-1, J-2, \dots, 1 \end{cases}$$

$$\delta_{jn}^{(i)} \stackrel{\text{df}}{=} \varepsilon_{jn}^{(i)} \left. \frac{d\phi_{ij}(\kappa_j^{(i)})}{d\kappa_j^{(i)}} \right|_{\theta_{ij}^{(s)} = \tilde{\theta}_{ij,n}^{(s)}}$$

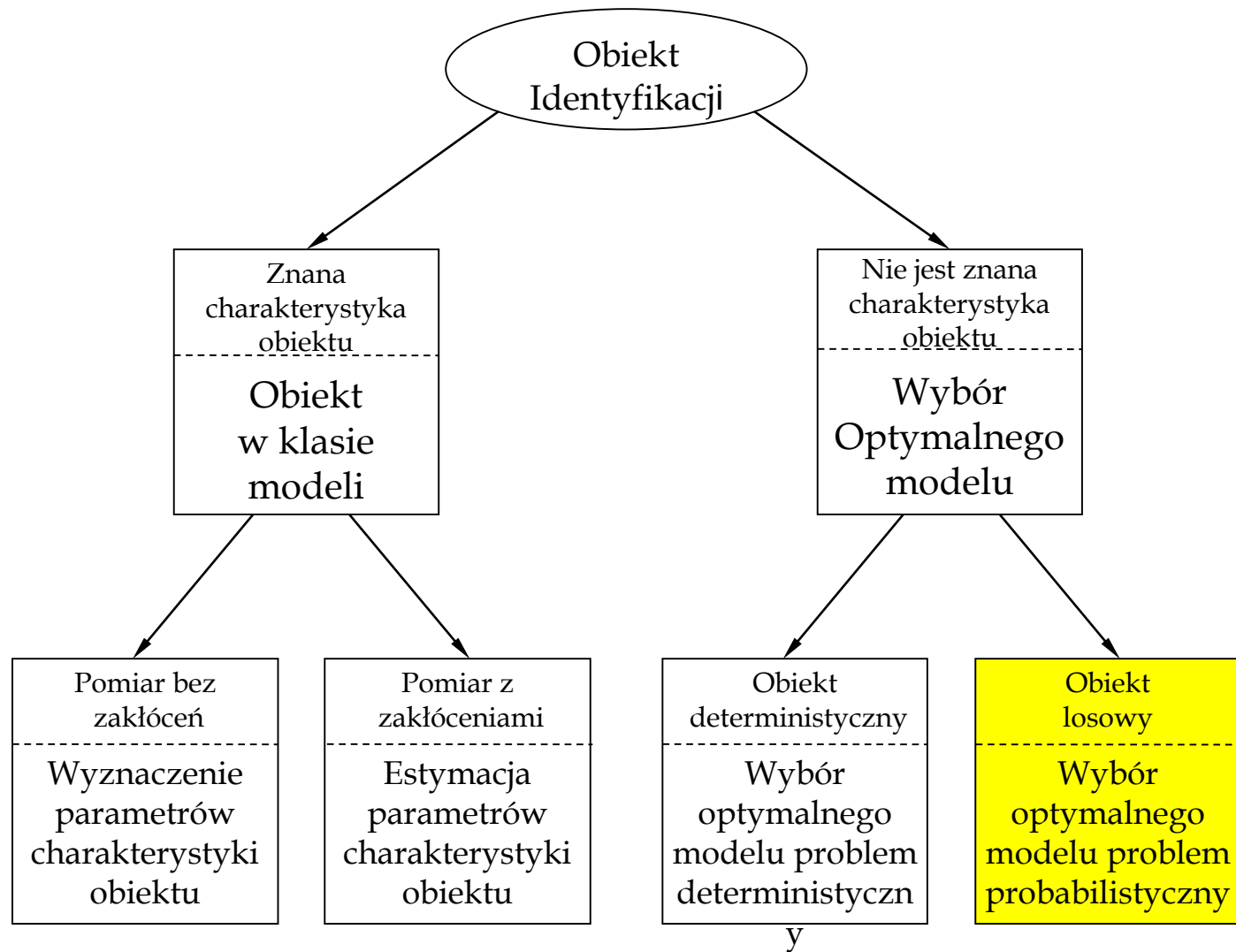
$$\kappa_j^{(i)} \stackrel{\text{df}}{=} \sum_{s=1}^{I_{j-1}} \mu_{j-1}^{(s)} \theta_{ij}^{(s)} + \theta_{ij}^{(0)}.$$

$$\tilde{\theta}_{ij,n+1}^{(s)} = \tilde{\theta}_{ij,n}^{(s)} + 2\eta_n \delta_{jn}^{(i)} \mu_{j-1,n}^{(s)}, \quad s=0, 1, \dots, I_{J-1}, \quad i=1, 2, \dots, I_J, \quad j=1, 2, \dots, J.$$

Wybór Optymalnego Modelu Przypadek Probabilistyczny

∞ 49 ∞

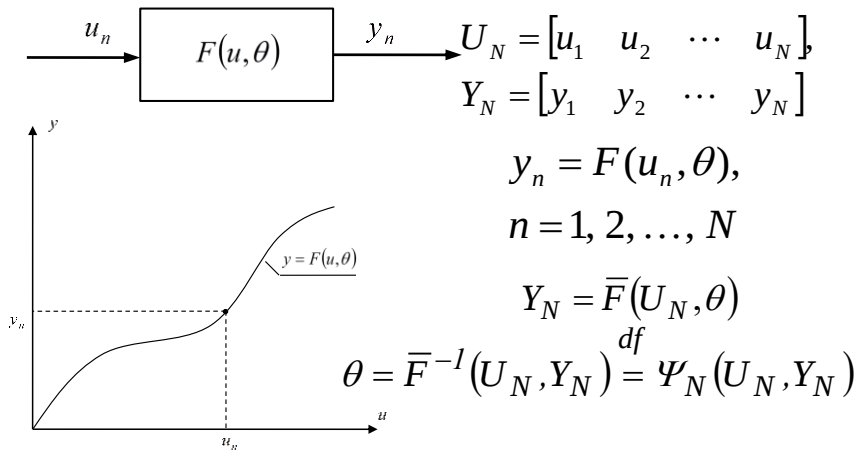
Typowe zadania identyfikacji



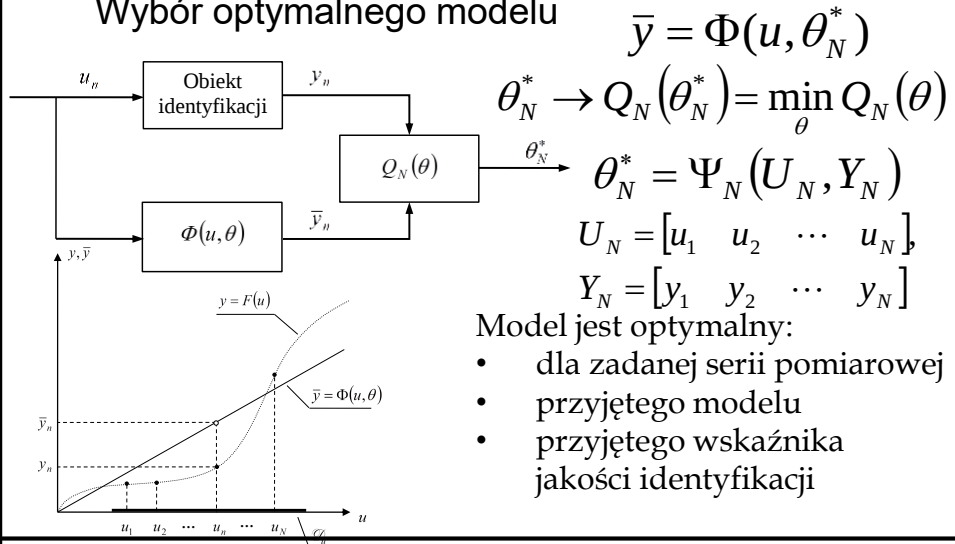
Podstawowe zadania identyfikacji podsumowanie

Obiekt w klasie modeli

D
e
t
e
r
m
i
n
i
s
t
y
c
z
n
y



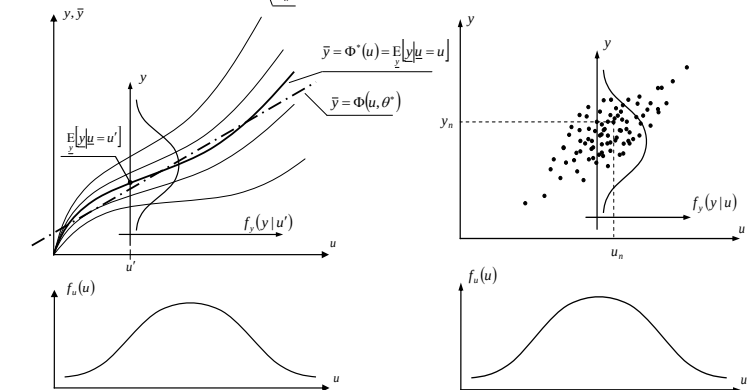
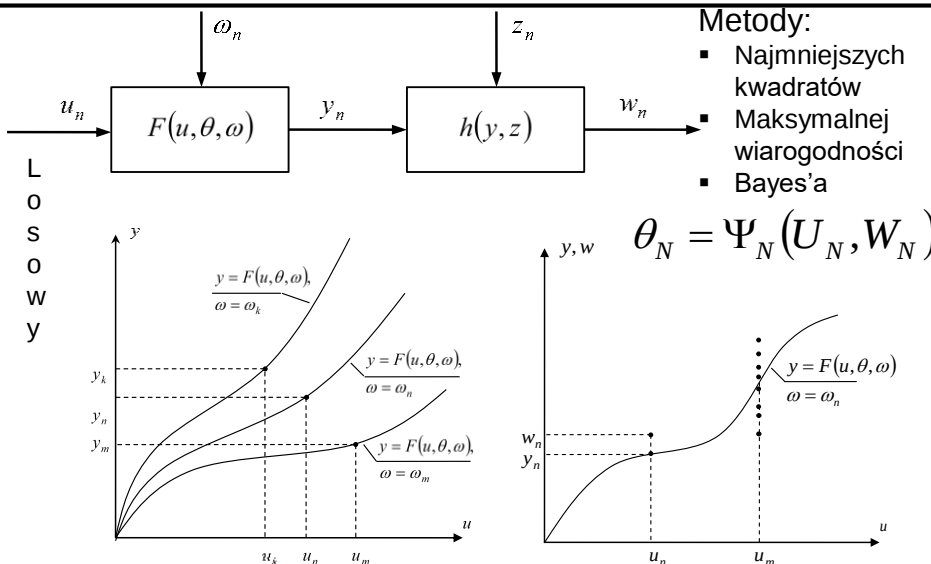
Wybór optymalnego modelu



Metody:

- Najmniejszych kwadratów
- Maksymalnej wiarygodności
- Bayes'a

$$\theta_N = \Psi_N(U_N, W_N)$$



Pełna informacja

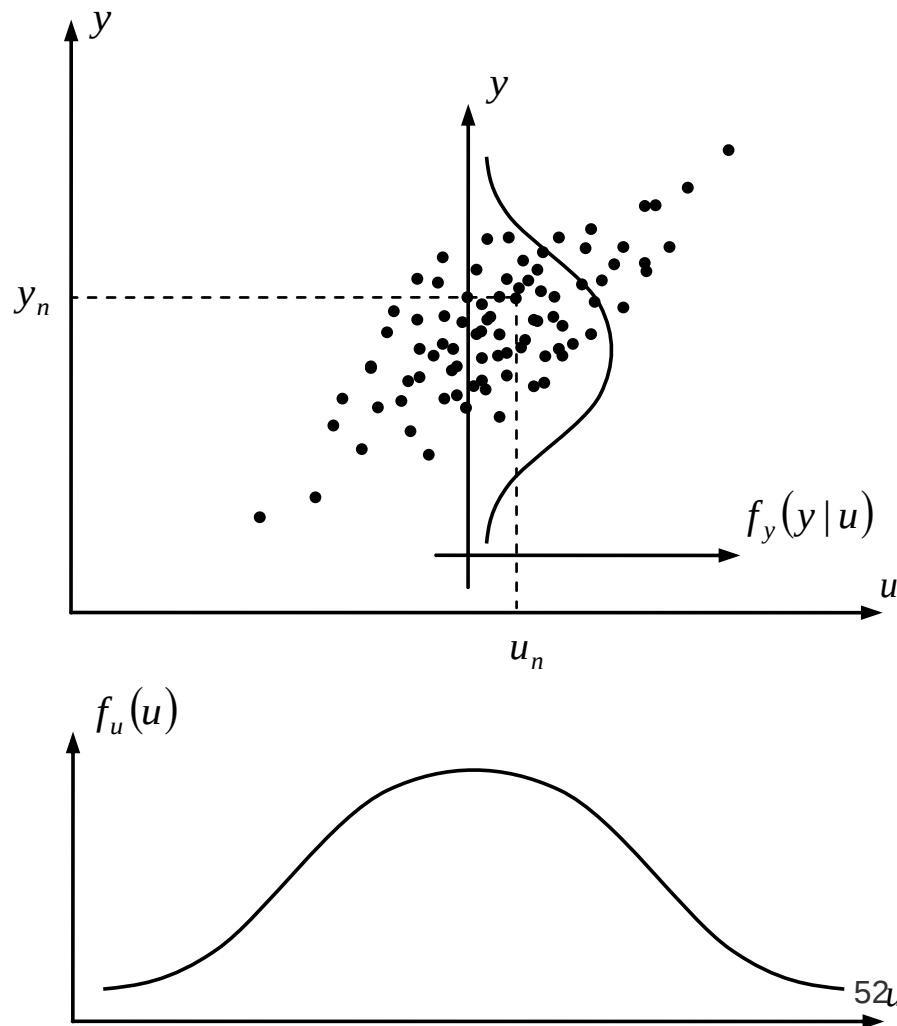
- Regresja I rodzaju
- Regresja II rodzaju

Niepełna informacja

- Estymacja wskaźnika jakości
- Estymacja parametrów rozkładu
- Estymacja rozkładu

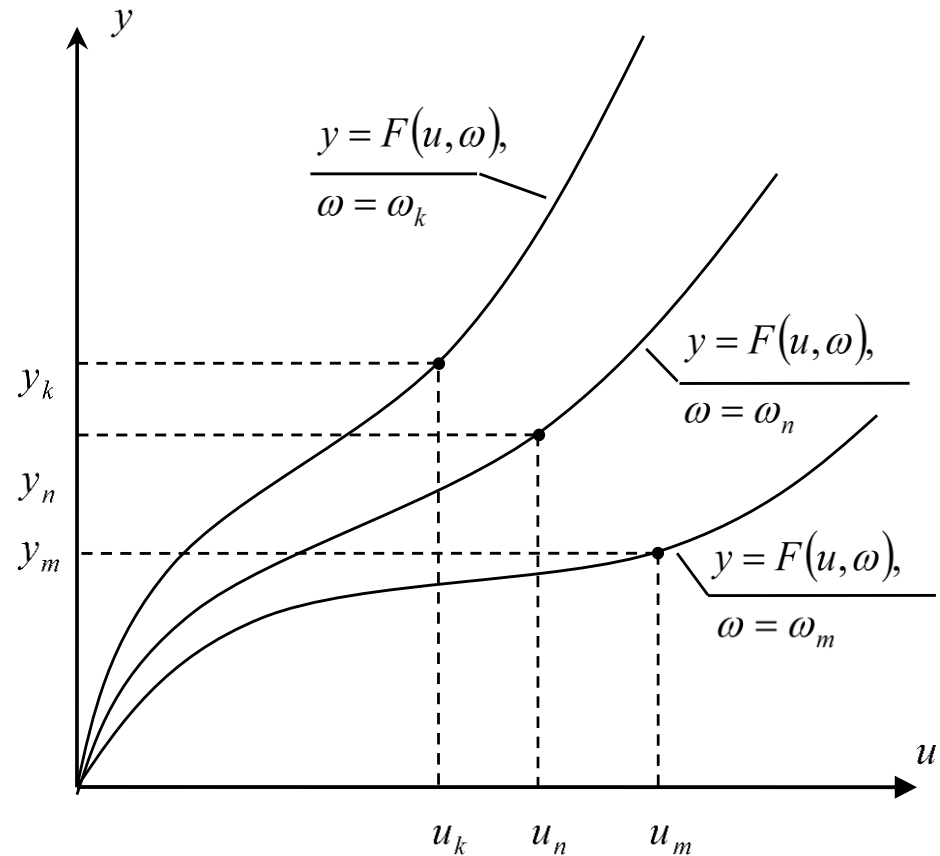
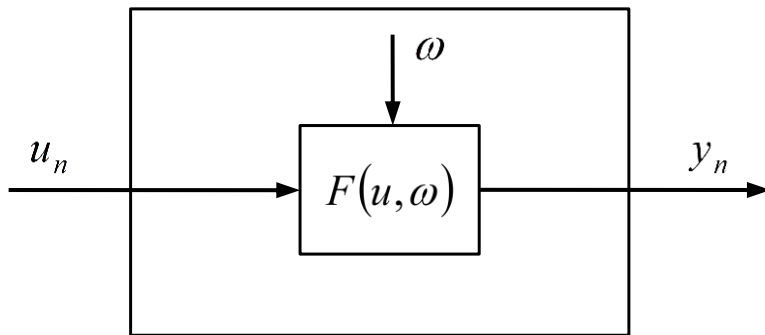
$$(u_n, y_n), n = 1, 2, \dots, N$$

Są wartościami pary
zmiennych losowych $(\underline{u}, \underline{y})$



Wybór optymalnego modelu w warunkach losowych

∞ Obiekt identyfikacji z losowo zmienną wielkością ω



Wybór optymalnego modelu w warunkach losowych

∞ Sformułowanie problemu

Charakterystyka statyczna obiektu identyfikacji : $\underline{y} = F(u, \underline{\omega})$

ω_n - wartość zmiennej losowej $\underline{\omega}$ (losowo zmienna wielkość w obiekcie)

$f_{\omega}(\omega)$ - funkcja gęstości rozkładu prawdopodobieństwa zmiennej losowej $\underline{\omega}$ ($\omega \in \Omega \subseteq \mathcal{R}^L$),

y_n - wartość zmiennej losowej $\underline{y}$ - przekształcenie zmiennej losowej $\underline{\omega}$

F - funkcja wzajemnie jednoznaczna względem ω ,
a zatem istnieje funkcja odwrotna : $\underline{\omega} = F_{\omega}^{-1}(u, \underline{y})$

Wybór optymalnego modelu w warunkach losowych

Warunkowa gęstość rozkładu prawdopodobieństwa zmiennej losowej $\underline{y}$, pod warunkiem, że zmienna losowa $\underline{u} = u$:

$$f_y(y | u) = f_\omega(F_\omega^{-1}(u, y)) \cdot |J_F|$$

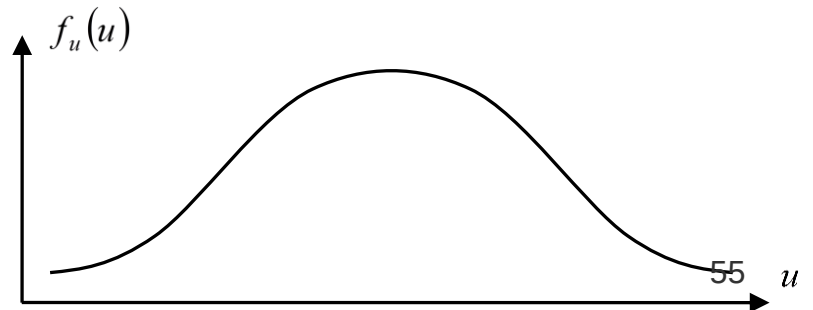
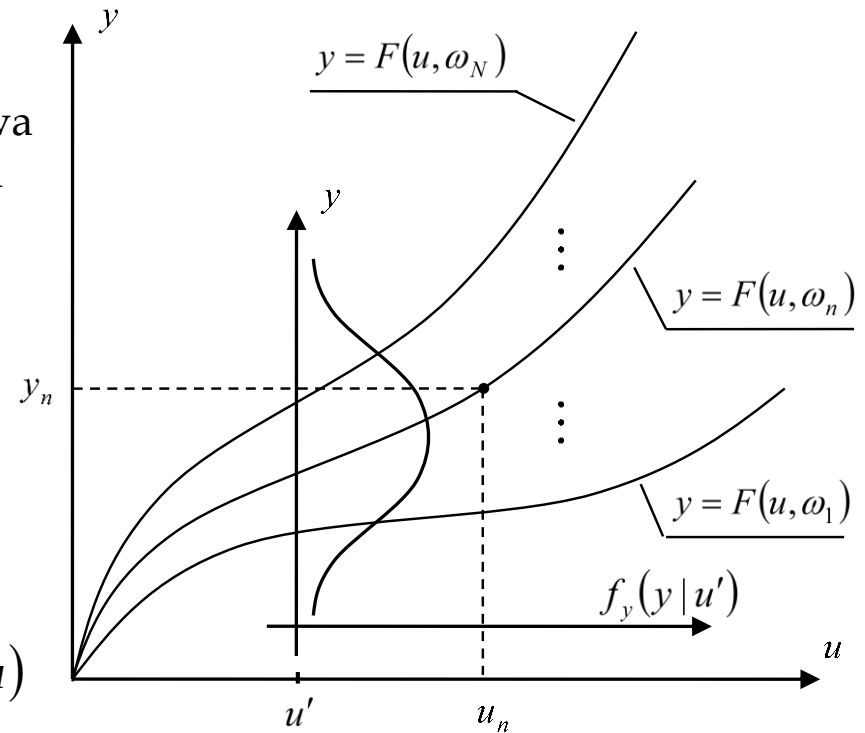
gdzie

$$J_F = \frac{\partial F_\omega^{-1}(u, y)}{\partial \underline{y}}$$

u_n - wartość zmiennej losowej $\underline{u}$,
o gęstości rozkładu prawdopodobieństwa $f_u(u)$

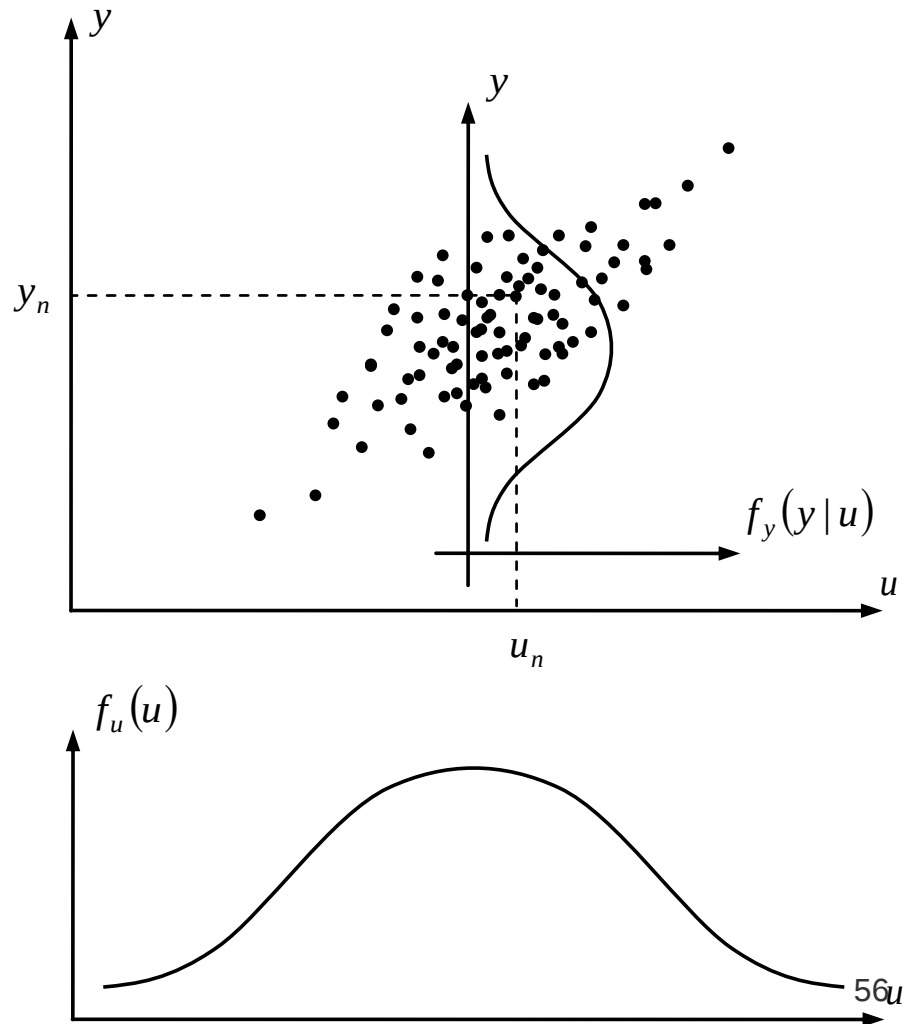
Wejście u_n i wyjście y_n są wartościami pary zmiennych losowych $(\underline{u}, \underline{y})$

$$u \in \mathcal{U} \subseteq \mathbb{R}^S, \quad y \in \mathcal{Y} \subseteq \mathbb{R}^L$$



$$(u_n, y_n), n = 1, 2, \dots, N$$

Są wartościami pary
zmiennych losowych $(\underline{u}, \underline{y})$



Wybór optymalnego modelu w warunkach losowych

Dwa możliwe przypadki

Pełna informacja probabilistyczna

łączna gęstość rozkładu prawdopodobieństwa
pary zmiennych losowych $(\underline{u}, \underline{y})$
 $f(u, y)$

lub

warunkowa gęstość rozkładu
prawdopodobieństwa $f_y(y|u)$

i brzegowa $f_u(u)$

są znane

Niepełna informacja probabilistyczna

łączna gęstość rozkładu prawdopodobieństwa
pary zmiennych losowych $(\underline{u}, \underline{y})$ istnieje
ale nie jest znana

W zamian dysponujemy pomiarami:

$$(u_n, y_n), n = 1, 2, \dots, N$$

które są wartościami pary zmiennych
losowych $(\underline{u}, \underline{y})$

Pełna informacja probabilistyczna

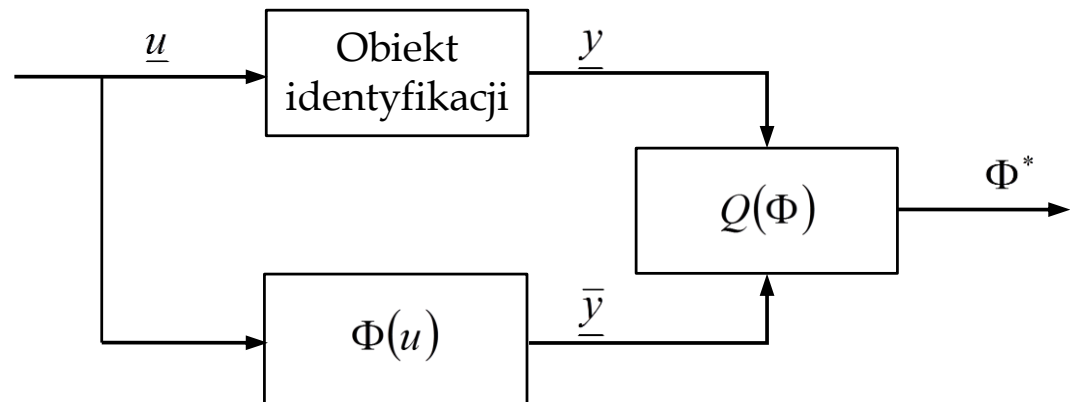
$$f(u, y) = f_y(y|u) \times f_u(u) \quad - \text{jest znane}$$

Model: $\bar{y} = \Phi(u)$ - szukamy funkcji Φ

Wskaźnik jakości:
$$Q(\Phi) = E_{\underline{u}, \underline{y}}[q(\underline{y}, \Phi(\underline{u}))] = \int \int_{\mathcal{U} \mathcal{Y}} q(y, \Phi(u)) f(u, y) dy du$$

Model optymalny: $\Phi^* \rightarrow Q(\Phi^*) = \min_{\Phi} Q(\Phi)$

$$\bar{y} = \Phi^*(u)$$



Pełna informacja probabilistyczna

$$Q(\Phi) = E_{\underline{u}} \left[E_{\underline{y}} [q(\underline{y}, \Phi(\underline{u})) | \underline{u} = u] \right] = \int_{\mathcal{U}} \left[\int_{\mathcal{Y}} q(y, \Phi(u)) f_y(y|u) dy \right] f_u(u) du$$

Dla ustalonego u warunkowa wartość oczekiwana ma postać:

$$Q_u(\Phi(u)) = E_{\underline{y}} [q(\underline{y}, \Phi(\underline{u})) | \underline{u} = u] = \int_{\mathcal{Y}} q(y, \Phi(u)) f_y(y|u) dy$$

Optymalny model otrzymujemy:

$$\Phi^*(u) \rightarrow Q_u(\Phi^*(u)) = \min_{\Phi(u)} Q_u(\Phi(u))$$

Dla :

$$q(y, \bar{y}) = [y - \bar{y}]^T [y - \bar{y}]$$

Wskaźnik jakości ma postać:

$$Q_u(\Phi(u)) = E_{\underline{y}} [[y - \Phi(u)]^T [y - \Phi(u)] | \underline{u} = u] = \int_{\mathcal{Y}} [y - \Phi(u)]^T [y - \Phi(u)] f_y(y|u) dy$$

$$Q_u(\Phi(u)) = E_{\underline{y}} \left[[y - \Phi(u)]^T [y - \Phi(u)] | \underline{u} = u \right] = \int_{\mathcal{Y}} [y - \Phi(u)]^T [y - \Phi(u)] f_y(y|u) dy$$

$$Q(\Phi(u)) = E_{\underline{y}} \left[\underline{y}^T \underline{y} | u \right] - 2\Phi(u) E_{\underline{y}} \left[\underline{y} | u \right] + \Phi(u)^T \Phi(u)$$

$$\nabla_{\Phi(u)} Q(\Phi(u)) = -2E_{\underline{y}} \left[\underline{y} | u \right] + 2\Phi^*(u) = 0_L$$

$$\Phi^*(u) = E_{\underline{y}} \left[\underline{y} | u \right] = \int_{\mathcal{Y}} y f_y(y|u) dy$$

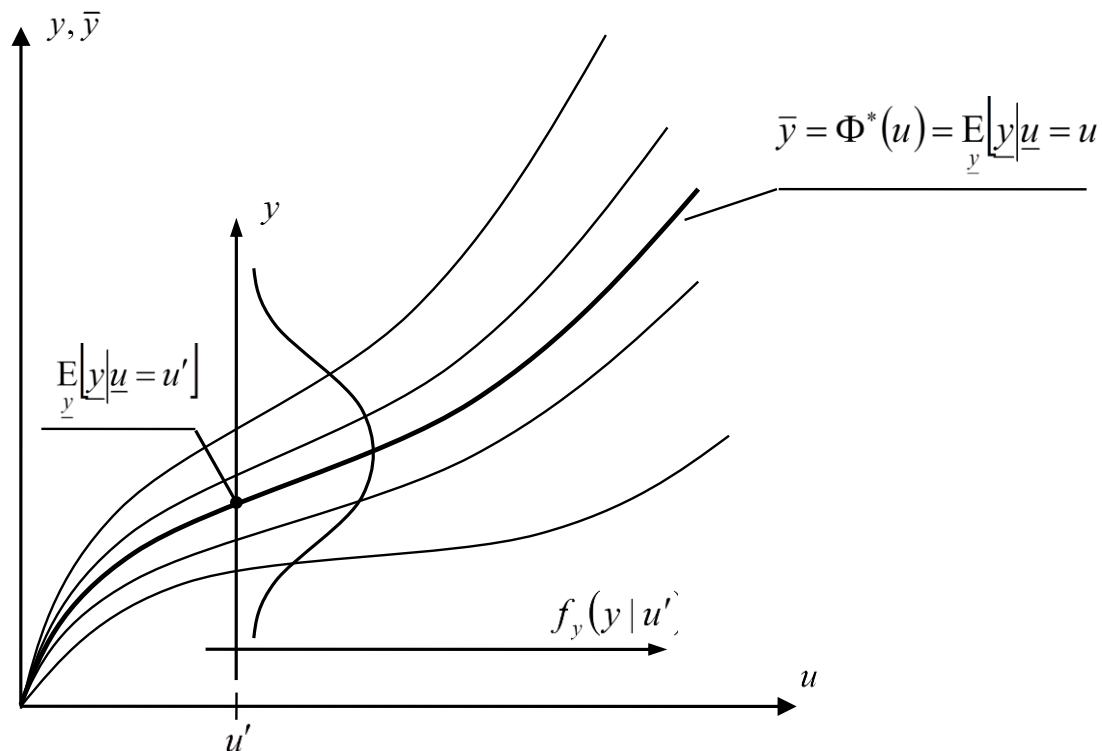
Pełna informacja probabilistyczna

Regresja I rodzaju

$$Q_u(\Phi(u)) = E_{\underline{y}}[[y - \Phi(u)]^T [y - \Phi(u)] | \underline{u} = u] = \int_{\mathcal{Y}} [y - \Phi(u)]^T [y - \Phi(u)] f_y(y|u) dy$$

W wyniku minimalizacji powyższego wskaźnika względem $\Phi(u)$ otrzymujemy optymalny model:

$$\bar{y} = \Phi^*(u) = E_{\underline{y}}[y | \underline{u} = u] = \int_{\mathcal{Y}} y f(y|u) dy$$



Pełna informacja probabilistyczna

∞ Przykład regresja I rodzaju dla rozkładu normalnego

Łączna gęstość rozkładu pary zmiennych losowych $(\underline{u}, \underline{y})$ ma postać:

$$f(u, y) = (2\pi)^{-\frac{L+S}{2}} |\Sigma^{-1}|^{\frac{1}{2}} \exp \left[-\frac{1}{2} \left(\begin{bmatrix} u \\ y \end{bmatrix} - \begin{bmatrix} m_u \\ m_y \end{bmatrix} \right)^T \Sigma^{-1} \left(\begin{bmatrix} u \\ y \end{bmatrix} - \begin{bmatrix} m_u \\ m_y \end{bmatrix} \right) \right]$$

gdzie: Σ – macierz kowariancji, $\Sigma = \begin{bmatrix} \Sigma_{uu} & \Sigma_{uy} \\ \Sigma_{yu} & \Sigma_{yy} \end{bmatrix}$

m_u, m_y – wartości oczekiwane zmiennych losowych odpowiednio: $\underline{u}$ i $\underline{y}$

Warunkowa funkcja gęstości prawdopodobieństwa ma postać:

$$f_y(y|u) = (2\pi)^{-\frac{L}{2}} |\Sigma_{y|u}^{-1}|^{\frac{1}{2}} \exp \left[-\frac{1}{2} (y - m_{y|u})^T \Sigma_{y|u}^{-1} (y - m_{y|u}) \right]$$

Pełna informacja probabilistyczna

∞ Przykład regresja I rodzaju dla rozkładu normalnego

$$f(y|u) = (2\pi)^{-\frac{L}{2}} |\Sigma_{y|u}^{-1}|^{\frac{1}{2}} \exp \left[-\frac{1}{2} (y - m_{y|u})^T \Sigma_{y|u}^{-1} (y - m_{y|u}) \right]$$

Macierz kowariancji: $\Sigma_{y|u} = \Sigma_{yy} - \Sigma_{yu} \Sigma_{uu}^{-1} \Sigma_{uy}$

Wartość oczekiwana: $m_{y|u} = m_y + \Sigma_{yu} \Sigma_{uu}^{-1} (u - m_u)$

Optymalny model: $\bar{y} = \Phi^*(u) = E_{\underline{y}}[y | \underline{u} = u] = m_{y|u} = m_y + \Sigma_{yu} \Sigma_{uu}^{-1} (u - m_u)$

$$\bar{y} = \Phi^*(u) = Au + b$$

gdzie: $A = \Sigma_{yu} \Sigma_{uu}^{-1}$

$$b = m_y - \Sigma_{yu} \Sigma_{uu}^{-1} m_u$$

Pełna informacja probabilistyczna

Model: $\bar{y} = \Phi(u, \theta)$ - szukamy wektora parametrów θ

Φ zadana funkcja, θ - wektora parametrów modelu

$$u \in \mathcal{U} \subseteq \mathcal{R}^S, \quad \bar{y} \in \overline{\mathcal{Y}} \subseteq \mathcal{R}^L, \quad \theta \in \Theta \subseteq \mathcal{R}^R$$

Wskaźnik jakości: $Q(\theta) = E_{\underline{u}, \underline{y}}[q(\underline{y}, \Phi(\underline{u}, \theta))] = \int \int_{\mathcal{U} \mathcal{Y}} q(y, \Phi(u, \theta)) f(u, y) dy du$

$$\theta^* \rightarrow Q(\theta^*) = \min_{\theta \in \Theta} Q(\theta)$$

Optimalny model: $\bar{y} = \Phi(u, \theta^*)$

Pełna informacja probabilistyczna

☞ Regresja II rodzaju

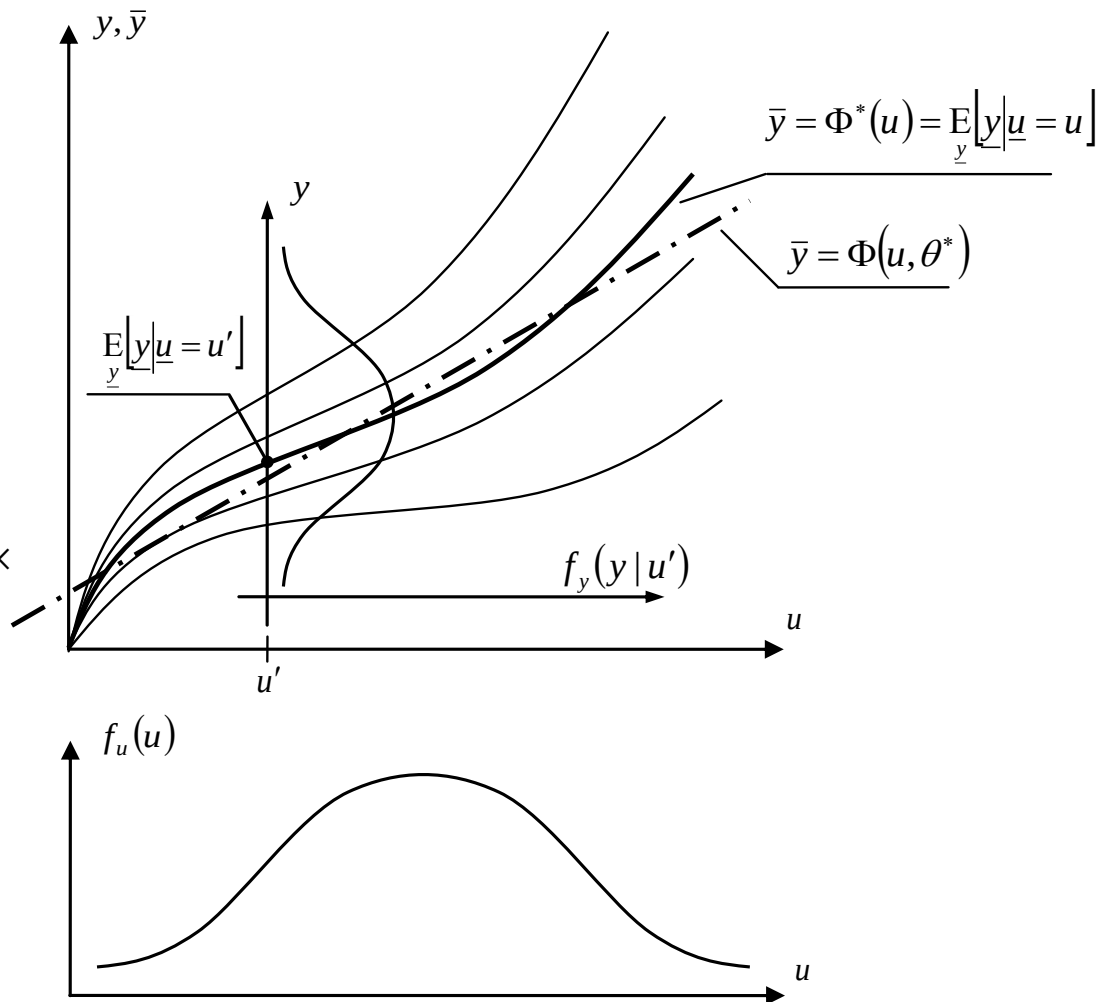
Dla:

$$q(y, \bar{y}) = [y - \bar{y}]^T [y - \bar{y}]$$

$$Q(\theta) = \int \int_{\mathcal{U} \times \mathcal{Y}} [y - \Phi(u, \theta)]^T [y - \Phi(u, \theta)] \times \\ \times f(u, y) dy du$$

$$\theta^* \rightarrow Q(\theta^*) = \min_{\theta \in \Theta} Q(\theta)$$

$$\bar{y} = \Phi(u, \theta^*)$$



Pełna informacja probabilistyczna

Czy jest jakaś relacja pomiędzy regresją I rodzaju i II rodzaju ?

$$\Phi^*(u) \sim ? \sim \Phi(u, \theta^*)$$

Wskaźnik jakości:

$$Q(\theta) = E_{\underline{u}, \underline{y}} \left[\left(\underline{y} - \Phi(\underline{u}, \theta) \right)^2 \right] = \int \int_{\mathcal{U} \mathcal{Y}} (y - \Phi(u, \theta))^2 f(u, y) dy du$$

Przyjmując:

$(\Phi^*(u) - \Phi^*(u) = 0)$ oraz $f(u, y) = f_y(y|u) f_u(u)$ otrzymujemy:

$$Q(\theta) = \int \int_{\mathcal{U} \mathcal{Y}} \left((y - \Phi^*(u)) + (\Phi^*(u) - \Phi(u, \theta)) \right)^2 f_y(y|u) f_u(u) dy du$$

Pełna informacja probabilistyczna

$$\begin{aligned} Q(\theta) = & \int \int_{\mathcal{U} \mathcal{Y}} (y - \Phi^*(u))^2 f_y(y|u) f_u(u) dy du + \\ & + 2 \int \left[\int_{\mathcal{U}} \int_{\mathcal{Y}} (y - \Phi^*(u)) (\Phi^*(u) - \Phi(u, \theta)) f_y(y|u) dy \right] f_u(u) du + \\ & + \int \int_{\mathcal{U} \mathcal{Y}} (\Phi^*(u) - \Phi(u, \theta))^2 f_y(y|u) f_u(u) dy du = \end{aligned}$$

$$\begin{aligned} = & \int_{\mathcal{U}} Q_u(\Phi^*(u)) f_u(u) du + \\ & + 2 \int \int_{\mathcal{U} \mathcal{Y}} (y - \Phi^*(u)) f_y(y|u) dy (\Phi^*(u) - \Phi(u, \theta)) f_u(u) du + \\ & + \int_{\mathcal{U}} (\Phi^*(u) - \Phi(u, \theta))^2 \int_{\mathcal{Y}} f_y(y|u) dy f_u(u) du \end{aligned}$$

Pełna informacja probabilistyczna

$$\begin{aligned} Q(\theta) &= \sigma_y^2 + \\ &+ 2 \int_{\mathcal{U}} \left[\int_{\mathcal{Y}} y f_y(y|u) dy - \Phi^*(u) \int_{\mathcal{Y}} f_y(y|u) dy \right] (\Phi^*(u) - \Phi(u, \theta)) f_u(u) du + \\ &+ \int_{\mathcal{U}} (\Phi^*(u) - \Phi(u, \theta))^2 f_u(u) du = \\ &= \sigma_y^2 + 2 \int_{\mathcal{U}} [\Phi^*(u) - \Phi^*(u) \cdot 1] (\Phi^*(u) - \Phi(u, \theta)) f_u(u) du + \\ &+ \int_{\mathcal{U}} (\Phi^*(u) - \Phi(u, \theta))^2 f_u(u) du = \\ &= \sigma_y^2 + \int_{\mathcal{U}} (\Phi^*(u) - \Phi(u, \theta))^2 f_u(u) du \end{aligned}$$

Pełna informacja probabilistyczna

$$Q(\theta) = \int \int_{\mathcal{U} \times \mathcal{Y}} (y - \Phi^*(u))^2 f(u, y) dy du + \int_{\mathcal{U}} (\Phi^*(u) - \Phi(u, \theta))^2 f_u(u) du$$

$$\theta^* \rightarrow \min_{\theta} Q(\theta) = \min_{\theta} \int_{\mathcal{U}} (\Phi^*(u) - \Phi(u, \theta))^2 f_u(u) du$$


Regresja I rodzaju

Funkcja wagi

Regresja II rodzaju jest najlepszym przybliżeniem regresji I rodzaju z funkcją wagi będącą gęstością rozkładu prawdopodobieństwa wejścia

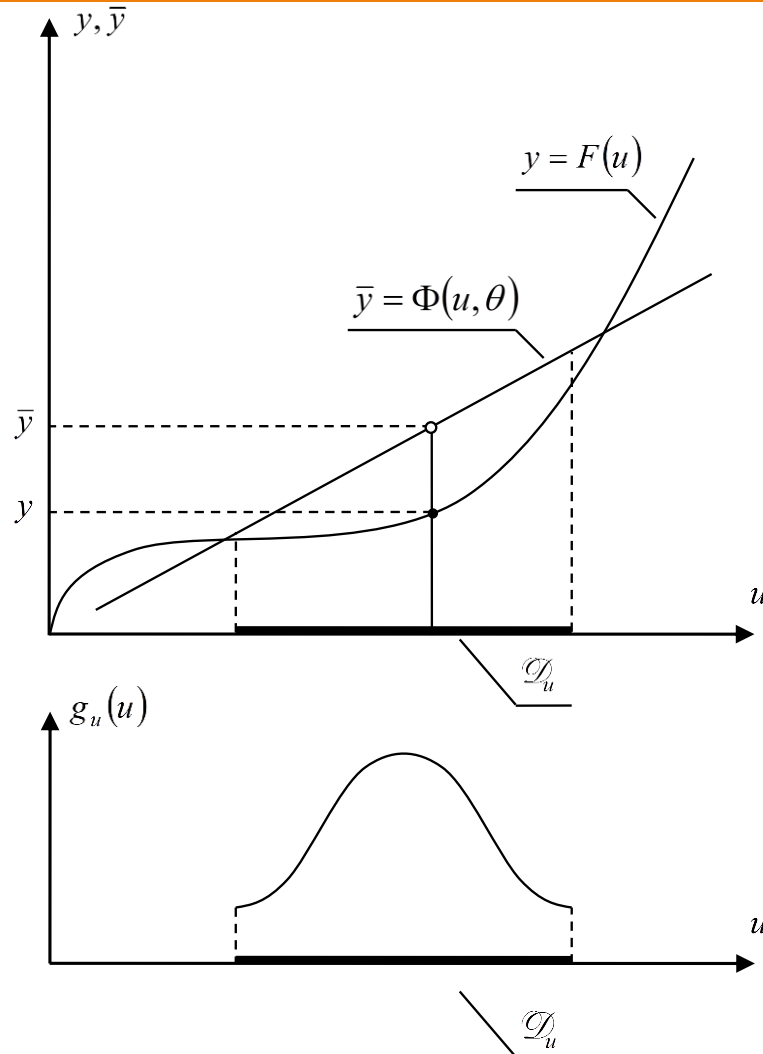
Zadanie aproksymacji charakterystyki statycznej obiektu

☞ Sformułowanie problemu

$$Q(\theta) = \int_{\mathcal{D}_u} q(F(u), \Phi(u, \theta)) g_u(u) du$$

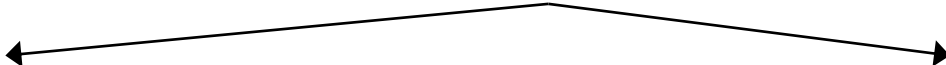
$$\theta_N^* \rightarrow Q_N(\theta_N^*) = \min_{\theta} Q_N(\theta)$$

$$\bar{y} = \Phi(u, \theta_N^*)$$



Wybór optymalnego modelu w warunkach losowych

Dwa możliwe przypadki



Pełna informacja probabilistyczna

łączna gęstość rozkładu prawdopodobieństwa
pary zmiennych losowych $(\underline{u}, \underline{y})$
 $f(u, y)$

lub

warunkowa gęstość rozkładu
prawdopodobieństwa $f_y(y|u)$

i brzegowa $f_u(u)$

są znane

Niepełna informacja probabilistyczna

łączna gęstość rozkładu prawdopodobieństwa
pary zmiennych losowych $(\underline{u}, \underline{y})$ istnieje
ale nie jest znana

W zamian dysponujemy pomiarami:

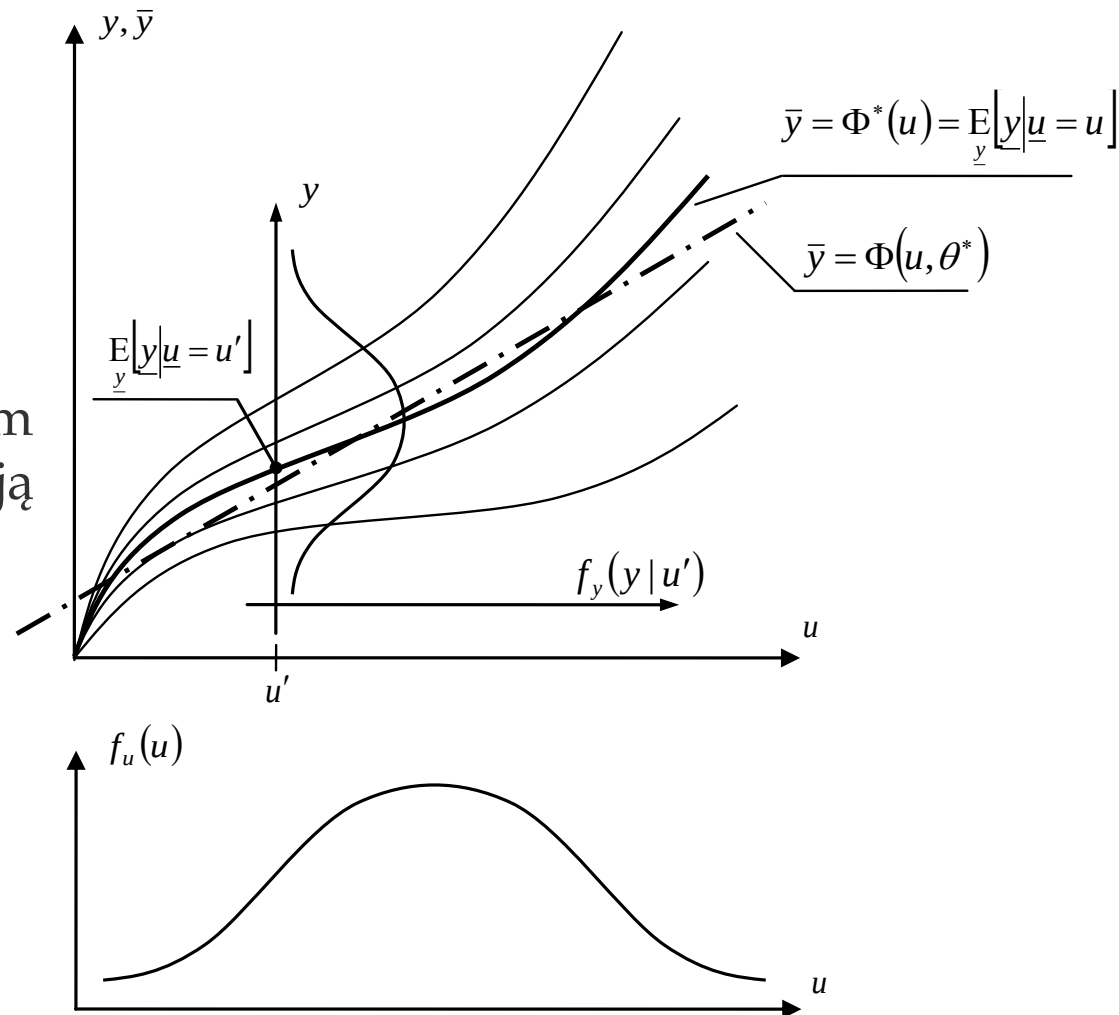
$$(u_n, y_n), n = 1, 2, \dots, N$$

które są wartościami pary zmiennych
losowych $(\underline{u}, \underline{y})$

Pełna informacja probabilistyczna

Relacja pomiędzy regresją I rodzaju a regresją II rodzaju

- Regresja II rodzaju jest najlepszym przybliżeniem regresji I rodzaju z funkcją wagi będącą gęstością rozkładu prawdopodobieństwa wejścia



Niepełna informacja probabilistyczna

Łączna gęstość rozkładu
prawdopodobieństwa
pary zmiennych losowych
 $(\underline{u}, \underline{y})$ istnieje

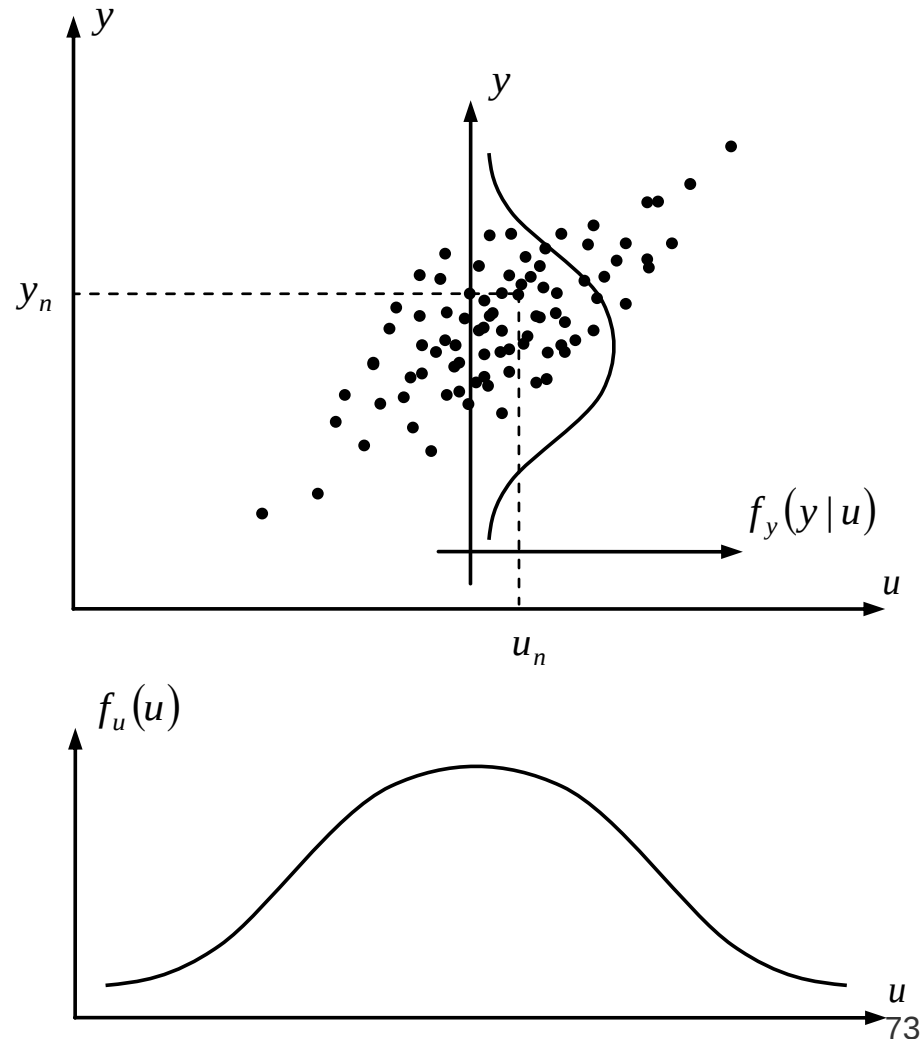
ale nie jest znana

W zamian dysponujemy
pomiarami:

$(u_n, y_n), n = 1, 2, \dots, N$

które są wartościami

pary zmiennych
losowych $(\underline{u}, \underline{y})$



Typowe podejścia

- ⌘ Estymacja wskaźników jakości
- ⌘ Estymacja parametrów rozkładu prawdopodobieństwa
- ⌘ Nieparametryczna estymacja gęstości rozkładów prawdopodobieństwa

Niepełna informacja probabilistyczna

Empiryczne oszacowanie wskaźników jakości

$$Q_u(\Phi(u)) = E_{\underline{y}}[q(\underline{y}, \Phi(\underline{u})) | \underline{u} = u] = \int_{\mathcal{Y}} q(y, \Phi(u)) f_y(y | u) dy$$

Eksperyment:

Dla ustalonego wejścia: u_n
dokonujemy wielokrotnego pomiaru
wyjścia y_{nm} .

W wyniku otrzymujemy:

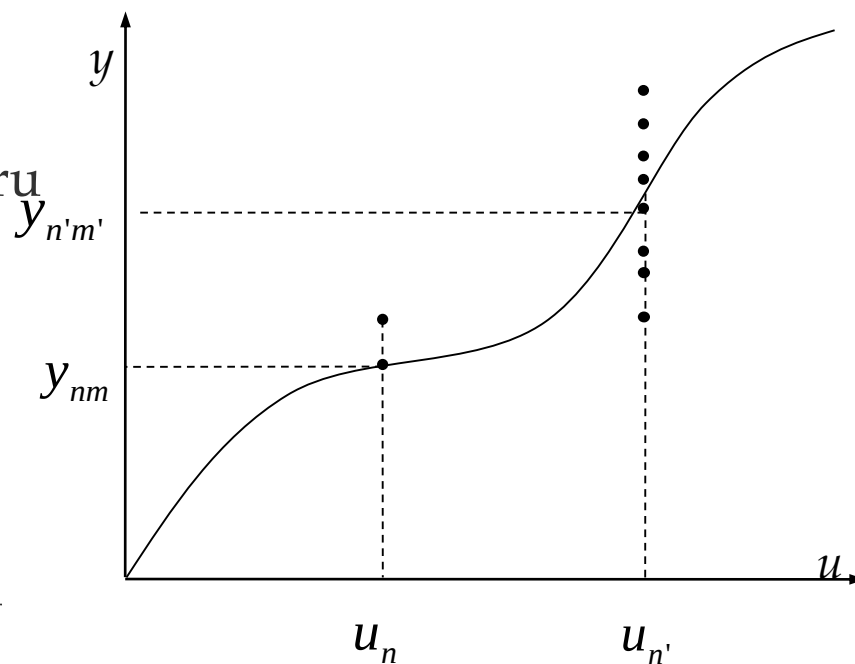
$(u_n, y_{nm}), m = 1, 2, \dots, M_n, n = 1, 2, \dots, \bar{N}$

gdzie:

M_n – liczba pomiarów wyjścia dla
ustalonego wejścia u_n .

$\bar{N}$ – liczba różnych wartości wejść

$N = \sum_{n=1}^{\bar{N}} M_n$ – liczba wszystkich pomiarów



Niepełna informacja probabilistyczna

Empiryczne oszacowanie wskaźników jakości

Oszacowanie wskaźnika jakości:

$$Q_u(\Phi(u)) = E_{\underline{y}}[q(\underline{y}, \Phi(\underline{u})) | \underline{u} = u] = \int_{\mathcal{Y}} q(y, \Phi(u)) f_y(y | u) dy$$

dla $u = u_n, \quad n = 1, 2, \dots, \bar{N}$

$$Q_{uM_n}(\Phi(u_n)) = \frac{1}{M_n} \sum_{m=1}^{M_n} q(y_{nm}, \Phi(u_n))$$

$$\Phi_{M_n}^*(u_n) \rightarrow Q_{uM_n}(\Phi_{M_n}^*(u_n)) = \min_{\Phi(u_n)} Q_{uM_n}(\Phi(u_n))$$

$$\Phi_{M_n}^*(u_n) \quad n = 1, 2, \dots, \bar{N}$$

$$N = \sum_{n=1}^{\bar{N}} M_n$$

Niepełna informacja probabilistyczna

Empiryczne oszacowanie wskaźników jakości

Dla

$$q(y, \bar{y}) = [y - \bar{y}]^T [y - \bar{y}]$$

Wskaźnik jakości:

$$Q_u(\Phi(u)) = E_{\underline{y}} \left[[y - \Phi(u)]^T [y - \Phi(u)] \mid \underline{u} = u \right] = \int_{\mathcal{Y}} [y - \Phi(u)]^T [y - \Phi(u)] f_y(y|u) dy$$

dla ustalonego $u = u_n$

$$Q_{uM_n}(\Phi(u_n)) = \frac{1}{M_n} \sum_{m=1}^{M_n} [y_{nm} - \Phi(u_n)]^T [y_{nm} - \Phi(u_n)]$$

Estymacja funkcji regresji I rodzaju:

$$\bar{y} = \Phi^*(u) = E_{\underline{y}} [y \mid \underline{u} = u] = \int_{\mathcal{Y}} y f(y|u) dy$$

$$\bar{y}_n = \frac{1}{M_n} \sum_{m=1}^{M_n} y_{nm}, \quad n = 1, 2, \dots, \bar{N}$$

Niepełna informacja probabilistyczna

Empiryczne oszacowanie wskaźników jakości

Oszacowanie wskaźnika jakości:

$$Q(\theta) = E_{\underline{u}, \underline{y}} \left[(y - \Phi(u, \theta))^2 \right] = \iint_{\mathcal{U} \times \mathcal{Y}} (y - \Phi(u, \theta))^2 f(u, y) dy du$$

$$Q_N(\theta) = \frac{1}{N} \sum_{n=1}^N q(y_n, \Phi(u_n, \theta))$$

lub

$$Q_N(\theta) = \frac{1}{N} \sum_{n=1}^{\bar{N}} \sum_{m=1}^{M_n} q(y_{nm}, \Phi(u_n, \theta))$$

$$\theta_N^* \rightarrow Q_N(\theta_N^*) = \min_{\theta \in \Theta} Q_N(\theta)$$

Niepełna informacja probabilistyczna

Empiryczne oszacowanie rozkładów prawdopodobieństwa

Estymacja parametrów gęstości rozkładu prawdopodobieństwa

Parametry gęstości rozkładu prawdopodobieństwa nie są znane

$$f(u, y) = f_{\alpha}(u, y, \alpha)$$

f_{α} - known function, $\alpha \in \mathcal{A}$ - nieznan wektor parametrów

Funkcja wiarygodności

$$L(U_N, Y_N, \alpha) = \prod_{n=1}^N f_{\alpha}(u_n, y_n, \alpha)$$

Niepełna informacja probabilistyczna

Empiryczne oszacowanie rozkładów prawdopodobieństwa

Oszacowanie:

$$\alpha_N = \Psi_N(U_N, Y_N) \rightarrow L(U_N, Y_N, \alpha_N) = \max_{\alpha \in \mathcal{A}} L(U_N, Y_N, \alpha)$$

Wskaźnik jakości: $Q_{u\alpha_N}(\Phi(u), \alpha_N) = \int_{\mathcal{Y}} q(y, \Phi(u)) \frac{f_\alpha(u, y, \alpha_N)}{\int_{\mathcal{Y}} f_\alpha(u, y, \alpha_N) dy} dy$

$$\Phi_{\alpha_N}^*(u) \rightarrow Q_{u\alpha_N}(\Phi_{\alpha_N}^*(u), \alpha_N) = \min_{\Phi(u)} Q_{u\alpha_N}(\Phi(u), \alpha_N)$$

Regresja I rodzaju

$$\bar{y} = \Phi_{\alpha_N}^*(u, \alpha_N) = \int_{\mathcal{Y}} y \frac{f_\alpha(u, y, \alpha_N)}{\int_{\mathcal{Y}} f_\alpha(u, y, \alpha_N) dy} dy$$

Wskaźnik jakości: $Q_{\alpha_N}(\theta, \alpha_N) = \int_{\mathcal{U}} \int_{\mathcal{Y}} q(y, \Phi(u, \theta)) f_\alpha(u, y, \alpha_N) dy du$

$$\theta_{\alpha_N}^* \rightarrow Q_{\alpha_N}(\theta_{\alpha_N}^*, \alpha_N) = \min_{\theta \in \Theta} Q_{\alpha_N}(\theta, \alpha_N)$$

Niepełna informacja probabilistyczna

Empiryczne oszacowanie rozkładów prawdopodobieństwa

$\underline{X}$ – zmienna losowa, funkcja gęstości rozkładu prawdopodobieństwa $f_x(x)$ istnieje lecz nie jest znana. $\dim x = P$

Obserwacje: $[x_1 \ x_2 \ \cdots \ x_N] = X_N$

Estymator Parzena :

$$f_x(x) \approx f_{xN}(x; X_N) = \frac{1}{Nh^P(N)} \sum_{n=1}^N K_x \left(\frac{x - x_n}{h(N)} \right)$$

gdzie:

$$h(N) > 0, \lim_{N \rightarrow \infty} h(N) = 0, \lim_{N \rightarrow \infty} h^P(N) = \infty$$

przykład $h(N) = cN^{-\alpha} \quad c > 0, 0 < \alpha < \frac{1}{P}$

K_x – znana funkcja jądra, taka że:

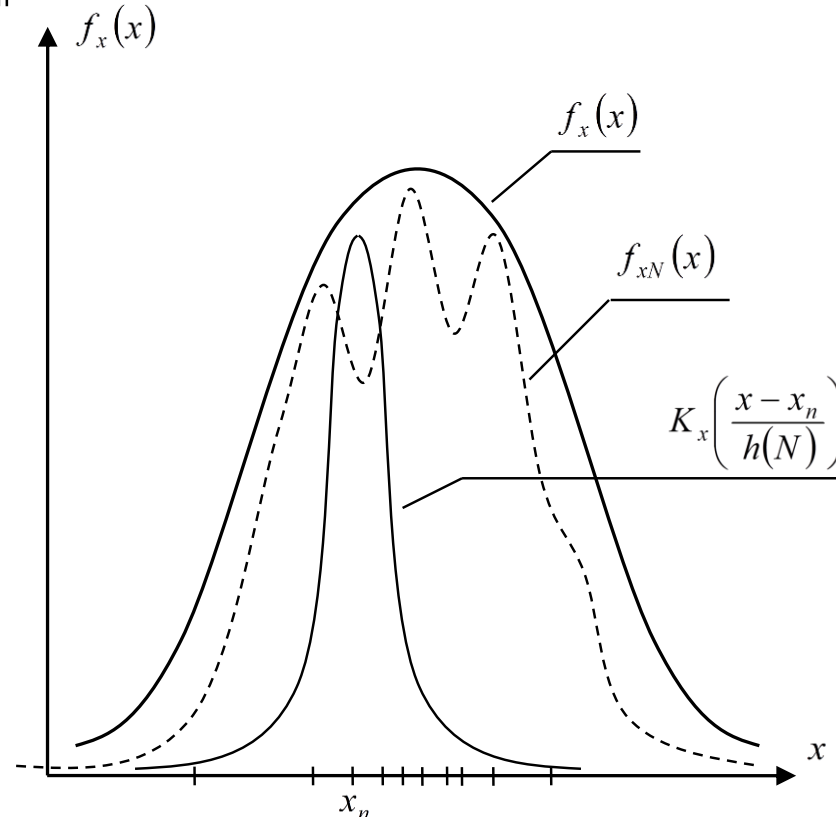
$$\sup_{x \in \mathcal{X}} |K_x(x)| < \infty, \int_{\mathcal{X}} K_x(x) dx = 1, \int_{\mathcal{X}} |K_x(x)| dx < \infty, \lim_{\|x\| \rightarrow \infty} |K_x(x)| = 0$$

Niepełna informacja probabilistyczna

Empiryczne oszacowanie rozkładów prawdopodobieństwa

Przykłady funkcji jądra K_x

$$K_x(x) = \begin{cases} 1 & \text{dla } \|x\| \leq 1 \\ 0 & \text{dla } \|x\| > 1 \end{cases}, \quad K_x(x) = (2\pi)^{-\frac{P}{2}} e^{-\frac{1}{2}\|x\|^2}, \quad K_x(x) = \pi^{-P} \prod_{p=1}^P \sqrt{1 + |x^{(p)}|}$$



Niepełna informacja probabilistyczna

Empiryczne oszacowanie rozkładów prawdopodobieństwa

Estymator Parzena funkcji gęstości rozkładu prawdopodobieństwa pary zmiennych losowych $(\underline{u}, \underline{y})$:

$$f_N(u, y; U_N, Y_N) = \frac{1}{N h^{L+S}(N)} \sum_{n=1}^N K_u\left(\frac{u - u_n}{h(N)}\right) K_y\left(\frac{y - y_n}{h(N)}\right)$$

Estymator Parzena funkcji gęstości rozkładu prawdopodobieństwa zmiennej losowej $\underline{u}$:

$$f_{uN}(u; U_N) = \frac{1}{N h^S(N)} \sum_{n=1}^N K_u\left(\frac{u - u_n}{h(N)}\right)$$

Niepełna informacja probabilistyczna

Empiryczne oszacowanie rozkładów prawdopodobieństwa

Estymator Parzena funkcji gęstości rozkładu warunkowego prawdopodobieństwa $f_y(y|u)$:

$$\begin{aligned} f_{yN}(y|u; U_N, Y_N) &= \frac{f_N(u, y; U_N, Y_N)}{f_{uN}(u; U_N)} = \\ &= \frac{1}{h^L(N)} \frac{\sum_{n=1}^N K_u\left(\frac{u - u_n}{h(N)}\right) K_y\left(\frac{y - y_n}{h(N)}\right)}{\sum_{n=1}^N K_u\left(\frac{u - u_n}{h(N)}\right)} \end{aligned}$$

Niepełna informacja probabilistyczna

Empiryczne oszacowanie rozkładów prawdopodobieństwa

Konsekwentnie:

$$Q_{uN}(\Phi(u); U_N, Y_N) = \int_{\mathcal{Y}} q(y, \Phi(u)) f_{yN}(y|u; U_N, Y_N) dy$$
$$\bar{y} = \Phi_N^*(u; U_N, Y_N) \rightarrow Q_{uN}(\Phi_N^*(u, U_N, Y_N), U_N, Y_N) = \min_{\Phi(u)} Q_{uN}(\Phi(u), U_N, Y_N)$$

Dla: $q(y, \bar{y}) = [y - \bar{y}]^T [y - \bar{y}]$

Regresja I rodzaju

$$\begin{aligned} \bar{y} = \Phi_N^*(u; U_N, Y_N) &= \int_{\mathcal{Y}} y f_{yN}(y|u; U_N, Y_N) dy = \\ &= \frac{1}{h^L(N)} \frac{\sum_{n=1}^N K_u\left(\frac{u - u_n}{h(N)}\right) \int_{\mathcal{Y}} y K_y\left(\frac{y - y_n}{h(N)}\right) dy}{\sum_{n=1}^N K_u\left(\frac{u - u_n}{h(N)}\right)} \end{aligned}$$

Niepełna informacja probabilistyczna

Empiryczne oszacowanie rozkładów prawdopodobieństwa

Dla specjalnej postaci funkcji $K_y(y)$, takiej że $K_y(y) = K_y(-y)$ mamy:

$$\frac{1}{h^L(N)} \int_{\mathcal{Y}} y K_y \left(\frac{y - y_n}{h(N)} \right) dy = y_n$$

Wówczas regresja I rodzaju przyjmuje postać:

$$\bar{y} = \Phi_N^*(u; U_N, Y_N) = \frac{\sum_{n=1}^N y_n K_u \left(\frac{u - u_n}{h(N)} \right)}{\sum_{n=1}^N K_u \left(\frac{u - u_n}{h(N)} \right)}$$

Niepełna informacja probabilistyczna

Empiryczne oszacowanie rozkładów prawdopodobieństwa

Optimalny model:

$$Q_N(\theta; U_N, Y_N) = \int \int_{\mathcal{U} \mathcal{Y}} q(y, \Phi(u, \theta)) f_N(u, y; U_N, Y_N) dy du$$

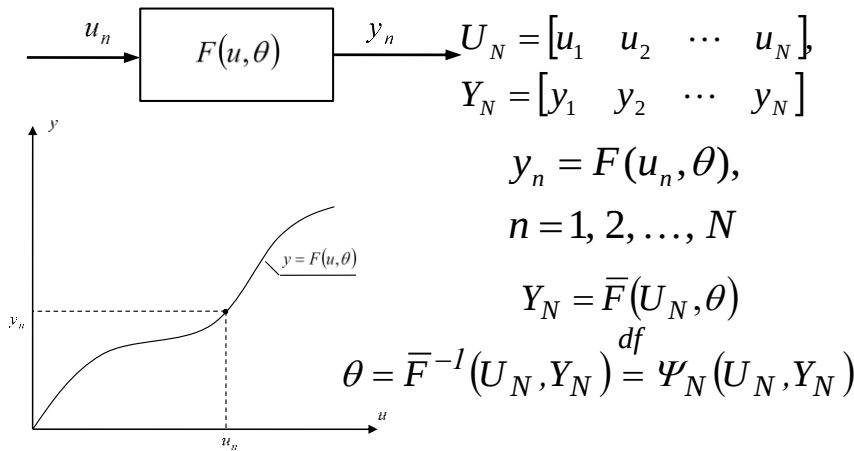
$$f_N(u, y; U_N, Y_N) = \frac{1}{N h^{L+S}(N)} \sum_{n=1}^N K_u\left(\frac{u - u_n}{h(N)}\right) K_y\left(\frac{y - y_n}{h(N)}\right)$$

$$\theta_N^*(U_N, Y_N) \rightarrow Q_N(\theta_N^*; U_N, Y_N) = \min_{\theta \in \Theta} Q_N(\theta; U_N, Y_N)$$

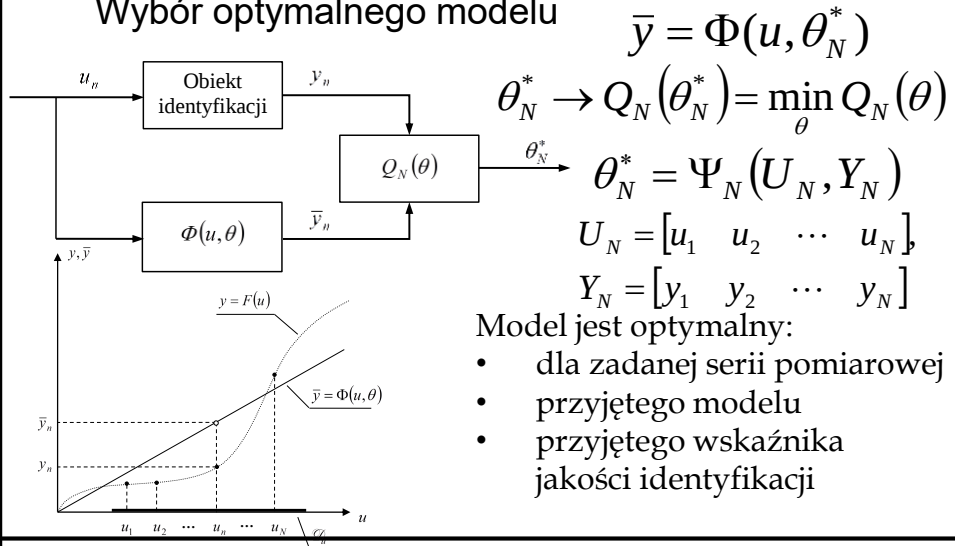
Podstawowe zadania identyfikacji podsumowanie

Obiekt w klasie modeli

D
e
t
e
r
m
i
n
i
s
t
y
c
z
n
y



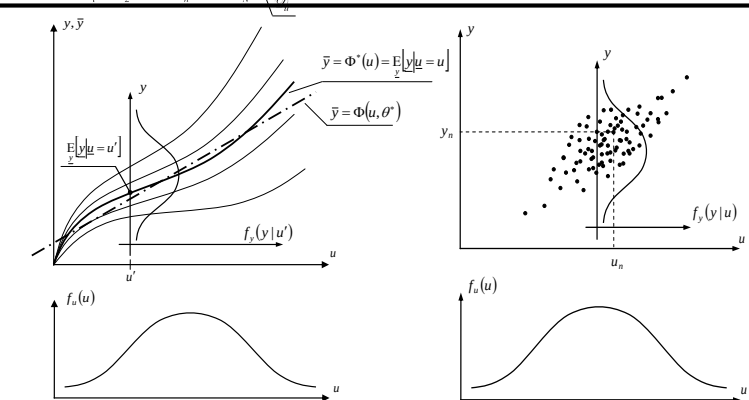
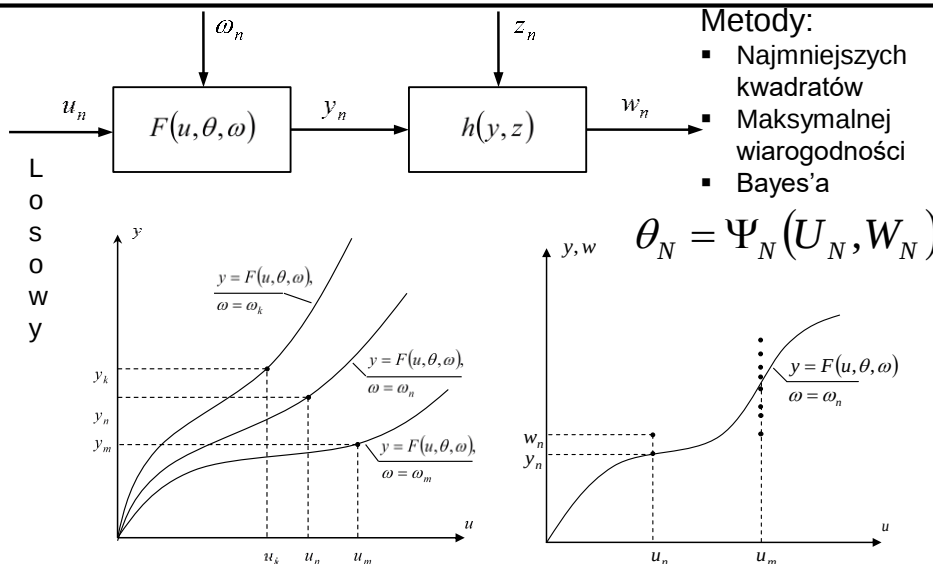
Wybór optymalnego modelu



Metody:

- Najmniejszych kwadratów
- Maksymalnej wiarygodności
- Bayes'a

$$\theta_N = \Psi_N(U_N, W_N)$$



Pełna informacja

- Regresja I rodzaju
- Regresja II rodzaju

Niepełna informacja

- Estymacja wskaźnika jakości
- Estymacja parametrów rozkładu
- Estymacja rozkładu

Dziękuję za uwagę

