

$X_1, X_2, X_3, \dots, X_p$  - predyktory, zmienne niezależne,  
input variables, features, cechy  
variables,

Bill Gates

$$\vec{X} = (X_1, X_2, X_3, \dots, X_p) \in \mathbb{R}^p$$

$Y$  - output, response, dependent variable

metody  
próby  
regulacji  
regulacji

RZECZYWISTOŚĆ

systematyczna  
zależność

? niepoznawalna

$$Y = f(X) + \epsilon$$

? niepoznawalna

losowość, szum  
 $\mathcal{N}(0, \sigma^2)$

MODEL

$$\hat{Y} = \hat{f}(x)$$

przybliżona  
wartość  
zmienną zależną

estymata (przybliżenie)  $f$

$E$  - błąd (avg. error)

$$\begin{aligned} E(Y - \hat{Y})^2 &= E[f(x) + \varepsilon - \hat{f}(x)]^2 \\ &= |f(x) - \hat{f}(x)|^2 + \text{Var}(\varepsilon) \end{aligned}$$

Składowa  
redukowalna

Składowa  
nieusuwalna



# Metody parametryczne

Przyjmujemy, że:

$$\hat{f}(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

b a

$$y = ax + b$$

Na podstawie danych możemy  
przyjąć, że  $X = (x_1)$

gdzie  $x_1$  - powierzchnia mieszkania

$$\hat{Y} = \hat{f}(x) = \underset{?}{\beta_0} + \underset{?}{\beta_1} x_1$$

metraz

TODO: wyznaczyć  $\beta_0$  i  $\beta_1$

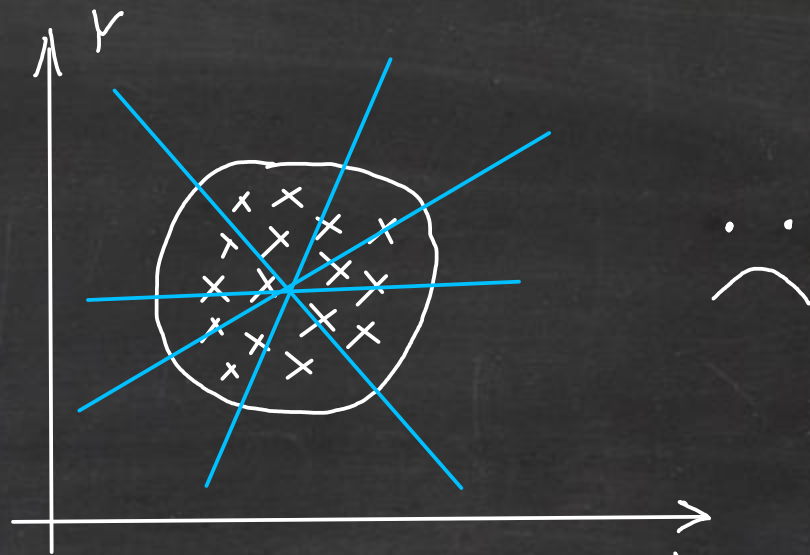






Metoda najmniejszych  
kwadratów

$$\hat{\beta}_1 = \text{tg}(\alpha)$$



brak korelacji pomaga  $x$  i  $y$

Dobrobyt

(x)  
outlier

