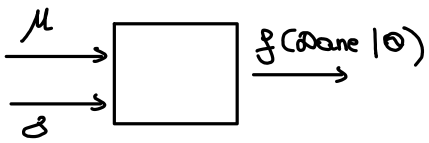
DOPASOWANIE ROZKŁADU DO DANYCH

① dane z jednego rozkładu



Mamy dane z rozkładu normalnego i chcemy znaleźć parametry μ i σ, aby model był najlepiej dopasowany.

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

$$\Theta = \begin{bmatrix} \mu \\ \sigma \end{bmatrix}$$

metoda max wiarygodności

$$L(\mu, \sigma) = \prod_{i=1}^{N} f(x_i, \mu, \sigma)$$

$$L(\mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] = \frac{1}{\sigma^N(\sqrt{2\pi})^N} \exp \sum_{i=1}^{N}\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

↓ przejście na skalę log, min i max fun zostają w tym samym miejscu

$$\ln L(\mu, \sigma) = -N\ln\sigma - N\ln\sqrt{2\pi} - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x-\mu)^2$$

$$\boxed{\hat{\Theta}_{ML} = \underset{\Theta}{\operatorname{argmax}}\, L(\Theta)}$$

$$\begin{cases} \frac{\partial L}{\partial \mu} = 0 \\ \frac{\partial L}{\partial \sigma} = 0 \end{cases} \Rightarrow \begin{matrix} \frac{\partial \ln L}{\partial \mu} = 0 \\ \frac{\partial \ln L}{\partial \sigma} = 0 \end{matrix}$$

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{2\sigma^2}\sum_{i=1}^{N} -2(x-\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{N}(x-\mu) = 0$$

$$\sum_{i=1}^{N} x_i - n\mu = 0$$

$$\boxed{\mu = \frac{\sum_{i=1}^{N} x_i}{n}}$$

$$\frac{\partial \ln L}{\partial \sigma} = -N\cdot\frac{1}{\sigma} + 2\cdot\frac{1}{2\sigma^3}\sum_{i=1}^{N}(x-\mu)^2 = -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{N}(x-\mu)^2 = 0$$

$$-N + \frac{1}{\sigma^2}\sum_{i=1}^{N}(x-\mu)^2 = 0$$

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x-\mu)^2}{N}$$

$$\boxed{\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x-\mu)^2}{N}}}$$

METODA BAYESA

Tym razem oprócz pomiarów, rozkładu zakłóceń mamy jeszcze wiedzę, że ten dany parametr ma jakiś rozkład.

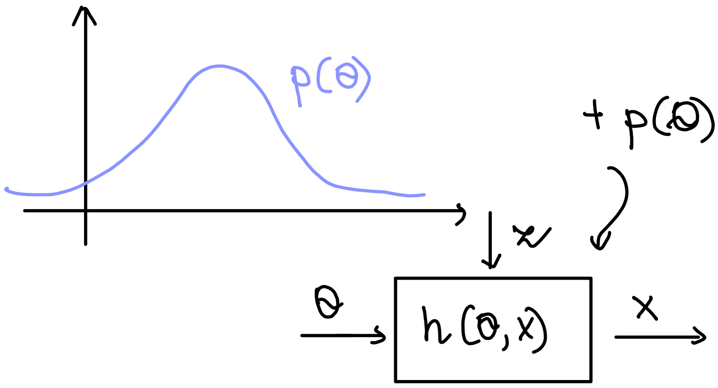
$$\boxed{\hat{\Theta}_{MAP} = \arg\max p(X|\Theta)\cdot p(\Theta)}$$

znane: $\mu,\ \sigma_z,\ \sigma\ ,\ \mu_x = 0$

$$p_x(x) = \frac{1}{\sigma_z\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x}{\sigma_z}\right)^2\right]$$

$X = \Theta + x \qquad \det J = 1$

$$p_\Theta(\Theta) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\Theta-\mu}{\sigma}\right)^2\right]$$



$$\hat{\Theta}_{MAP} = L(\Theta)\cdot p(\Theta) = p(X|\Theta)\cdot p(\Theta) = f(\Theta)$$

$$f(\Theta) = \prod_{i=1}^{N}\overbrace{p_x(x_i)\,|\det J|}^{p_x(x_i)}\cdot p(\Theta) = \frac{1}{\sigma_z^N(\sqrt{2\pi})^N}\exp\sum_{i=1}^{N}\left(-\frac{1}{2}\frac{(x_i-\Theta)^2}{\sigma_z^2}\right)\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\frac{(\Theta-\mu)^2}{\sigma^2}\right] =$$

$$= \frac{1}{\sigma_z^N\sigma(\sqrt{2\pi})^{N+1}}\exp\sum_{i=1}^{N}\left(-\frac{1}{2\sigma_z^2}(x_i-\Theta)^2\right)-\frac{1}{2\sigma^2}(\Theta-\mu)^2$$

$$\ln f(\Theta) = -N\ln\sigma_z - \ln\sigma - (N+1)\ln\sqrt{2\pi} - \frac{1}{2\sigma_z^2}\sum_{i=1}^{N}\left((x_i-\Theta)^2\right) - \frac{1}{2\sigma^2}(\Theta-\mu)^2$$

$$\frac{\partial\ln f(\Theta)}{\partial\Theta} = \frac{1}{\sigma_z^2}\sum_{i=1}^{N}(x_i-\Theta) - \frac{1}{\sigma^2}(\Theta-\mu) = 0$$

$$\frac{\sum_{i=1}^{N}x_i - n\Theta}{\sigma_z^2} - \frac{\Theta-\mu}{\sigma^2} = 0$$

$$\sigma^2\sum_{i=1}^{N}x_i - \sigma^2 n\Theta - \sigma_z^2\Theta + \sigma_z^2\mu = 0$$

$$\sigma^2\sum_{i=1}^{N}x_i + \sigma_z^2\mu = \Theta(\sigma^2 n + \sigma_z^2)$$

$$\Theta = \frac{\sigma^2\sum_{i=1}^{N}x_i + \sigma_z^2\mu}{\sigma^2 n + \sigma_z^2} = \frac{\sum_{i=1}^{N}x_i + \frac{\sigma_z^2}{\sigma^2}\mu}{n + \frac{\sigma_z^2}{\sigma^2}}$$

↳ dobrej jakości pomiary, mało rzetelna wiedza

$\sigma_z < \sigma$ ← większa σ dla wiedzy

$$\Theta = \frac{\sum_{i=1}^{N}x_i}{n} = \bar{X} \qquad \frac{\sigma_z^2}{\sigma^2}\approx 0$$

↳ dobra wiedza, stare pomiary $\sigma_2 > \sigma$ $\frac{\sigma^2}{\sigma_2^2} \approx 0$

$$\Theta = \frac{\frac{\sigma^2}{\sigma_2^2}\sum_{i=1}^{N} x_i + \mu}{\frac{\sigma^2}{\sigma_2^2} n + 1} = \mu$$

$$\Theta = \underbrace{\frac{n\sigma^2}{\sigma^2 n + \sigma_2^2}}_{W_1}\bar{X} + \underbrace{\frac{\sigma_2^2}{\sigma^2 n + \sigma_2^2}}_{W_2}\mu$$

$\bar{x}$ to estymator max $\hat{\Theta}_{ML}$

Zadania:

① Pokaż, że estymator MAP jest średnią ważoną estymatora ML i wiedzy apriorycznej wyrażonej przez wartość oczekiwaną μ.

$$\hat{\Theta}_{ML} = \arg\max L(\Theta)$$

$$\hookrightarrow L(\Theta) = \prod_{i=1}^{N} p_X(x_i)$$

Map jako średnia ważona estymatora ML i wiedzy z wart oczekiwanej μ.

$$\hat{\Theta}_{MAP} = \arg\max_{\Theta}\left(L(\Theta)\cdot p(\Theta)\right) = \arg\max_{\Theta}\left(\log L(\Theta)\cdot \log p(\Theta)\right) =$$

$$= \arg\max(\log L(\Theta)) + \arg\max(\log p(\Theta)) = \hat{\Theta}_{ML} + \mu$$

$$\hat{\Theta}_{MAP} = \lambda\cdot\hat{\Theta}_{ML} + (1-\lambda)\mu$$

$$\hat{\Theta}_{MAP} = \underbrace{\frac{n\sigma^2}{\sigma^2 n + \sigma_2^2}}_{W_1}\underbrace{\bar{X}}_{\hat{\Theta}_{ML}} + \underbrace{\frac{\sigma_2^2}{\sigma^2 n + \sigma_2^2}}_{W_2}\mu$$