## ① Мотивация

$\mathcal{M}_1, \dots, \mathcal{M}_K$

$$p(D|\mathcal{M}_k) = \int p(D|\bar\theta_k, \mathcal{M}_k)\, p(\bar\theta_k|\mathcal{M}_k)\, d\bar\theta_k$$

$D = \{\bar x_1, \dots, \bar x_N\} \sim p_{data}(\bar x) \approx p(\bar x|\bar\theta_{ML})$, $\bar\theta \in \mathbb{R}^d$

$$\bar\theta_{ML} = \arg\max_{\bar\theta} p(D|\bar\theta)$$

$$KL(p_{data} \| p(\bar x|\bar\theta)) = \int p_{data}(\bar x) \log \frac{p_{data}(\bar x)}{p(\bar x|\bar\theta)} d\bar x =$$

$$= \underbrace{\mathbb{E}_{p_{data}}[\log p_{data}(\bar x)]}_{} - \underbrace{\mathbb{E}_{p_{data}}[\log p(\bar x|\bar\theta)]}_{\xrightarrow[\bar\theta]{} \max}$$

$$\mathbb{E}_{p_{data}(\bar x)}[\log p(\bar x|\bar\theta_{ML})] \overset{D\sim p_{data}}{\approx} \frac{1}{N}\sum_{n=1}^{N} \log p(\bar x_n|\bar\theta_{ML}) \xrightarrow[\bar\theta]{} \max \to \bar\theta_{ML}$$

- $\bar\theta_0 = \arg\max_{\bar\theta} \mathbb{E}_{p_{data}(\bar x)}[\log p(\bar x|\bar\theta)]$
- $\bar\theta_{ML}(D) \xrightarrow{N\to\infty} \bar\theta_0$
- асимпт. нормальность:

$$\sqrt{N}\cdot(\bar\theta_{ML} - \bar\theta_0) \longrightarrow \mathcal{N}(\bar 0, I(\bar\theta_0)^{-1})$$

в предп., что $p_{data}(\bar x) = p(\bar x|\bar\theta_0)$

## ② Информация Фишера / Fisher information

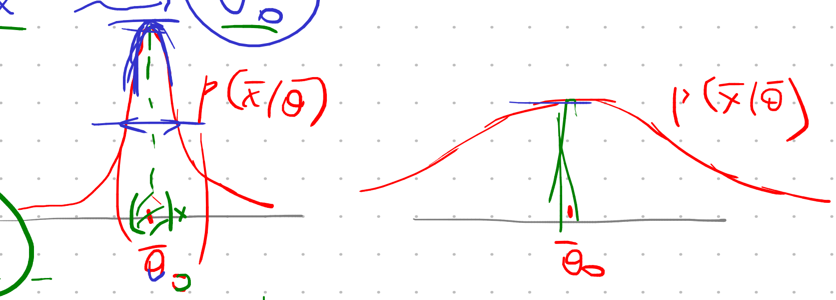
$p_{data}(\bar x) = p(\bar x|\bar\theta_0)$ $\quad \bar x \xrightarrow{?} \bar\theta_0$

$\bar\theta_0 = \arg\max_{\bar\theta} \mathbb{E}_{p_{data}}[p(\bar x|\bar\theta)]$

$\bar x \sim p_{data} = p(\bar x|\bar\theta_0)$

$\to \bar\theta_0 = \arg\max$

$\theta \in \mathbb{R}$




$$\mathbb{E}_{p(x|\theta_0)}\left[\frac{\partial \log p(\bar x|\theta_0)}{\partial\theta}\right] = \int \cancel{p(\bar x|\theta_0)} \frac{\partial p(\bar x|\theta)/\partial\theta|_{\theta_0}}{\cancel{p(\bar x|\theta_0)}} d\bar x = \frac{\partial}{\partial\theta}\int p(\bar x|\theta_0)\, d\bar x = 0$$

$$I(\theta) = \mathrm{Var}\left[\frac{\partial \log p(\bar x|\theta)}{\partial\theta}\right] = \mathbb{E}_{p(x|\theta)}\left[\left(\frac{\partial \log p}{\partial\theta}\right)^2\right] =$$

$$= \int \left(\frac{\partial \log p(\bar x|\theta)}{\partial\theta}\right)^2 p(\bar x|\theta)\, d\theta$$

$$\frac{\partial^2 \log p(\bar x|\theta)}{\partial\theta^2} = \frac{\partial}{\partial\theta}\left(\frac{\partial p/\partial\theta}{p}\right) = \frac{\partial^2 p/\partial\theta^2 \cdot p - (\partial p/\partial\theta)^2}{p^2} =$$

$$= \cancel{\frac{\partial^2 p/\partial\theta^2}{p}} - \left(\frac{\partial \log p(\bar x|\bar\theta)}{\partial\theta}\right)^2$$

$$\int \frac{\partial^2 f}{\partial x^2}\,dx$$

$$E_p\left[\frac{\partial^2 p/\partial\theta^2}{p}\right] = \int \cancel{p}\cdot\frac{\partial^2 p/\partial\theta^2}{\cancel{p}}\,d\bar x = \frac{\partial^2}{\partial\theta^2}\int p\,d\bar x = 0.$$

$$\boxed{I(\theta) = \mathrm{Var}\left[\frac{\partial \log p(\bar x|\theta)}{\partial\theta}\right] = -E_{p(\bar x|\theta)}\left[\frac{\partial^2 \log p(\bar x|\theta)}{\partial\theta^2}\right]}$$

$$I(\bar\theta)_{ij} = E_{p(\bar x|\bar\theta)}\left[\frac{\partial \log p(\bar x|\bar\theta)}{\partial\theta_i}\,\frac{\partial \log p(\bar x|\bar\theta)}{\partial\theta_j}\right]$$

|| если $p_{data} = p(\bar x|\bar\theta_0)$

$$J(\bar\theta)_{ij} = -E_{p(\bar x|\bar\theta)}\left[\frac{\partial^2 \log p(\bar x|\bar\theta)}{\partial\theta_i\,\partial\theta_j}\right]$$

(3) $$\bar\theta_0 = \arg\max_{\bar\theta} E_{p_{data}}\left[\log p(\bar x|\bar\theta)\right]$$

$\to$ $p_{data} \neq p(\bar x|\bar\theta_0)$, $\infty$

если $p_{data} = p(\bar x|\bar\theta_0)$, $\infty$

$$\sqrt{N}(\bar\theta_{ML} - \bar\theta_0) \to \mathcal{N}(\bar 0 \,|\, I(\bar\theta_0)^{-1})$$

$$\sqrt{N}(\bar\theta_{ML} - \bar\theta_0) \to \mathcal{N}(\bar 0,\ J^{-1}(\bar\theta_0)\cdot I(\bar\theta_0)\,J^{-1}(\bar\theta_0))$$

$\sqrt{N}$

$$\bar\theta_{ML} = \bar\theta_{ML}(D)$$

(4) Задача / пример

$$b(p_{data}) = E_{D\sim p_{data}}\left[\log p(D|\bar\theta_{ML}) - N\cdot E_{p_{data}(\bar x)}\left[\log p(\bar x|\bar\theta_{ML})\right]\right]$$

$$\log p(\bar y|\bar w, X) = -\frac{1}{2\sigma^2}\sum_n (y_n - \bar w^T \bar x_n)^2 + \mathrm{const}$$

$p_{data} = \mathcal{N}(\dots \bar w_{true})$

$$E_{p_{data}(\bar y)}\left[\log p(\bar y|\bar w, X)\right] = -\frac{1}{2\sigma^2} E_{\bar y}\left[\sum_n (y_n - \bar w^T \bar x_n)^2\right] =$$

$$= -\frac{1}{2\sigma^2}\left(\sum_n \underbrace{\mathbb{E}[y_n^2]}_{(\bar{w}_{true}^T \bar{x}_n)^2+\sigma^2} - 2\cdot\sum(\bar{w}^T\bar{x}_n)\cdot \underbrace{\mathbb{E}[y_n]}_{=\bar{w}_{true}^T\bar{x}_n} + \sum_n (\bar{w}^T\bar{x}_n)^2\right)$$

$$= -\frac{1}{2\sigma^2}\sum_{n=1}^{N}\left((\bar{w}_{true}^T\bar{x}_n)^2 + \sigma^2 - 2(\bar{w}^T\bar{x}_n)(\bar{w}_{true}^T\bar{x}_n) + (\bar{w}^T\bar{x}_n)^2\right)$$

$$= -\frac{N}{2} - \sum_{n=1}^{N}\left((\bar{w}-\bar{w}_{true})^T\bar{x}_n\right)^2$$

(5) $b(p_{data}) = \mathbb{E}_D[\log p(D|\bar{\theta}_{ML}) - N\cdot\mathbb{E}_{p_{data}}[\log p(\bar{x}|\bar{\theta}_{ML})]] =$

$$= \mathbb{E}_D[\log p(D|\bar{\theta}_{ML}^{(D)}) - \log p(D|\bar{\theta}_0)] + \quad B_1$$

$$+ \mathbb{E}_D[\log p(D|\bar{\theta}_0) - N\cdot\mathbb{E}_{p_{data}}[\log p(\bar{x}|\bar{\theta}_0)]] \quad B_2$$

$$+ \mathbb{E}_D[N\cdot\mathbb{E}_{p_{data}}[\log p(\bar{x}|\bar{\theta}_0)] - N\cdot\mathbb{E}_{p_{data}}[\log p(\bar{x}|\bar{\theta}_{ML}^{(D)})]] \quad B_3$$

$$B_2 = \mathbb{E}_D\left[\sum_n \log p(\bar{x}_n|\bar{\theta}_0)\right] - N\cdot\mathbb{E}_{p_{data}}[\log p(\bar{x}|\bar{\theta}_0)] = 0$$

$B_3$: $\quad \mathbb{E}_{p_{data}}[\log p(\bar{x}|\bar{\theta}_{ML})] = \eta(\bar{\theta}_{ML}) \approx$ Taylor $(\bar{\theta}_0)$

$$\approx \eta(\bar{\theta}_0) + \frac{1}{2}(\bar{\theta}_{ML}-\bar{\theta}_0)^T\left(\frac{\partial^2 \log p(\bar{x}|\bar{\theta})}{\partial\theta_i\partial\theta_j}\bigg|_{\bar{\theta}_0}\right)\cdot(\bar{\theta}_{ML}-\bar{\theta}_0)$$

$$= \eta(\bar{\theta}_0) - \frac{1}{2}(\bar{\theta}_{ML}-\bar{\theta}_0)^T J(\bar{\theta}_0)(\bar{\theta}_{ML}-\bar{\theta}_0)$$

$$B_3 \approx \mathbb{E}_D\left[N\cdot\eta(\bar{\theta}_0) - N\cdot\left(\eta(\bar{\theta}_0) - \frac{1}{2}(\_\_)^T J(\_\_)\right)\right] =$$

$$= \frac{N}{2}\mathbb{E}_D\left[(\bar{\theta}_{ML}-\bar{\theta}_0)^T \boxed{J(\bar{\theta}_0)}(\bar{\theta}_{ML}-\bar{\theta}_0)\right] \qquad \mathbb{E}[\bar{x}^T A\bar{x}] = \mathrm{Tr}(A\,\bar{x}\bar{x}^T)$$

$$= \frac{N}{2}\mathrm{Tr}\left(J(\bar{\theta}_0)\cdot\mathbb{E}_D[(\bar{\theta}_{ML}-\bar{\theta}_0)^T(\bar{\theta}_{ML}-\bar{\theta}_0)]\right) \approx$$

$$\approx \frac{1}{2}\mathrm{Tr}\left(\cancel{J(\bar{\theta}_0)\cdot J^{-1}(\bar{\theta}_0)}\, I(\bar{\theta}_0)\, J^{-1}(\bar{\theta}_0)\right)$$

$$B_3 \approx \frac{1}{2} \mathrm{Tr}\left(I(\bar\theta_0) \cdot J^{-1}(\bar\theta_0)\right)$$

$$l(\bar\theta) = \log p(D|\bar\theta)$$

$$B_1 = E_D\left[\log p(D|\bar\theta_{ML}) - \log p(D|\bar\theta_0)\right]$$

$$l(\bar\theta) \approx l(\bar\theta_{ML}) + \frac{1}{2}(\bar\theta - \bar\theta_{ML})^T \left.\frac{\partial^2 l(\bar\theta)}{\partial\bar\theta\,\partial\bar\theta^T}\right|_{\bar\theta_{ML}} (\bar\theta - \bar\theta_{ML})$$

$$\to -J(\bar\theta_0)$$

$$\frac{1}{N}\sum_{n=1}^{N} \left.\frac{\partial^2 \log p(\bar x_n|\bar\theta)}{\partial\bar\theta\,\partial\bar\theta^T}\right|_{\bar\theta_{ML}} \longrightarrow -J(\bar\theta_0)$$

Taylor

$$B_3 = E_D\left[\log p(D|\hat\theta_{ML}) - \log p(D|\bar\theta_0)\right] \approx \text{asymp.}$$

$$\approx \frac{N}{2} E_D\left[(\bar\theta - \bar\theta_{ML})^T J(\bar\theta_0)(\bar\theta - \bar\theta_{ML})\right] =$$

$$= \frac{N}{2} \mathrm{Tr}\left(J(\bar\theta_0) \cdot \overbrace{E_D\left[(\bar\theta - \bar\theta_{ML})^T(\bar\theta - \bar\theta_{ML})\right]}^{\frac{1}{N} J^{-1} \cdot I \cdot J}\right)$$

$$B_3 = \frac{1}{2} \mathrm{Tr}\left(I(\bar\theta_0) \cdot J(\theta_0)^{-1}\right)$$

$$b(p_{data}) = B_1 + B_2 + B_3 = \mathrm{Tr}\left(I(\bar\theta_0) \cdot J(\bar\theta_0)^{-1}\right)$$

Takeuchi:

$$TIC(M) = -2 \log p(D|\bar\theta_{ML}) + 2\mathrm{Tr}\left(\hat I(\bar\theta_0) \cdot \hat J(\bar\theta_0)^{-1}\right),$$

ye $$\hat I_{ij} = \frac{1}{N} \sum_{n=1}^{N} \left.\left(\frac{\partial \log p(\bar x_n|\bar\theta)}{\partial\theta_i} \frac{\partial \log p(\bar x_n|\bar\theta)}{\partial\theta_j}\right)\right|_{\bar\theta_{ML}}$$

$$\hat J_{ij} = - \text{''} - \frac{\partial^2 \log p(\bar x_n|\bar\theta)}{\partial\theta_i\,\partial\theta_j}$$

Ecли $p_{data} = p(\bar x|\hat\theta_0)$, то

Akaike inf. criterion

$$\boxed{AIC(M) = -2\log p(D|\bar{\theta}_M) + 2d}$$