① Bayesian model selection

D $\qquad \mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_k$

$\qquad \bar\theta_1, \bar\theta_2, \dots, \bar\theta_k$

$$p(\mathcal{M}_k|D) = \frac{p(\mathcal{M}_k)\, p(D|\mathcal{M}_k)}{p(D)} \qquad p(\mathcal{M}_k) = 1/k$$

„лин регр $d=4$"

$$p(\theta|D) = \frac{p(\theta)\, p(D|\theta)}{p(D)}$$

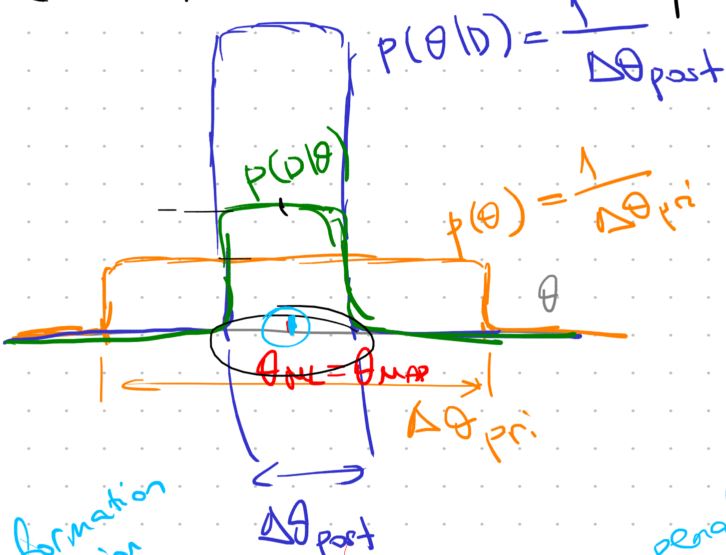
$$p(\bar\theta_k|D, \mathcal{M}_k) = \frac{p(\bar\theta_k|\mathcal{M}_k)\, p(D|\bar\theta_k, \mathcal{M}_k)}{p(D|\mathcal{M}_k)}$$

надо оценить

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

② Простой пример

$$p(D) = \int p(\bar\theta)\, p(D|\bar\theta)\, d\bar\theta$$



$$p(D) = \int p(\theta)\, p(D|\theta)\, d\theta = $$

$$= \int_{\Delta\theta_{post}} \frac{1}{\Delta\theta_{pri}}\, p(D|\theta_{MAP})\, d\theta =$$

penalty

$$= p(D|\theta_{MAP})\, \frac{\Delta\theta_{post}}{\Delta\theta_{pri}}$$

information criterion

$$\log p(D) = \log p(D|\theta_{MAP}) - \log \frac{\Delta\theta_{pri}}{\Delta\theta_{post}}$$

$$\bar\theta = (\theta_1, \dots, \theta_d)$$

$$p(D) = \int p(\bar\theta)\, p(D|\bar\theta)\, d\bar\theta = \int \left(\frac{1}{\Delta\theta_{pri}}\right)^d p(D|\bar\theta_{MAP})\, d\bar\theta$$

$(\Delta\theta_{post})^d$

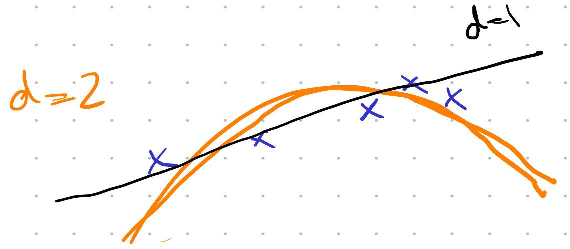
$$\log p(D) = \log p(D|\bar\theta_{MAP}) - d \cdot \log \frac{\Delta\theta_{pri}}{\Delta\theta_{post}}$$

число параметров

③ Sanity check

$D \sim \mathcal{M}_{true}$

$$\mathbb{E}_{D\sim\mathcal{M}_{true}}\left[p(D|\mathcal{M}_{true}) \geq p(D|\mathcal{M})\right]$$



$$E_{D\sim M_{true}}\left[\log\frac{p(D|M_{true})}{p(D|M)}\right] = \quad \geq 0$$

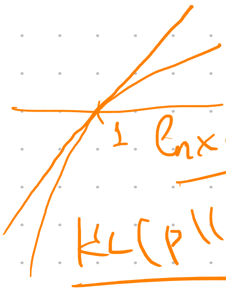
$$= \int p(D|M_{true})\log\frac{p(D|M_{true})}{p(D|M)}\,dD = KL\big(p(D|M_{true})\,||\,p(D|M)\big)$$

$KL(p||q) \neq KL(q||p)$

$$KL(p||q) = \int p(\bar{x})\log\frac{p(\bar{x})}{q(\bar{x})}\,d\bar{x}$$

$$= \int p(\bar{x})\log p(\bar{x})\,d\bar{x} - \int p(\bar{x})\log q(\bar{x})\,d\bar{x}$$

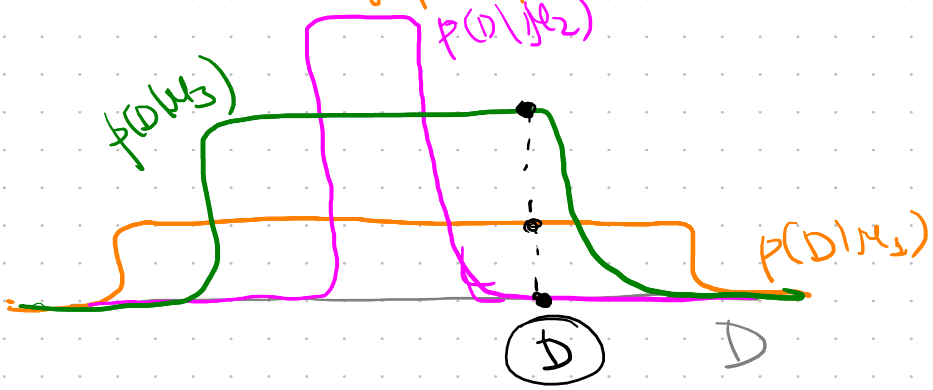
$$= H(p) - \underbrace{H(p,q)}_{\text{Cross-entropy}}$$



$\ln x \leq x-1$

$$KL(p||q) = \int p\log\frac{p}{q}\,d\bar{x} = -\int p\log\frac{q}{p}\,dx \geq$$

$$\geq -\int p\left(\frac{q}{p}-1\right)d\bar{x} = -\left(\int q\,d\bar{x} - \int p\,d\bar{x}\right) = 0$$



$p(D|M\ldots)$

④ Bayesian Information Criterion (BIC)

$$p(D) = \int p(\bar{\theta})p(D|\bar{\theta})\,d\bar{\theta} = \int p(\bar{\theta})\,e^{l(\bar{\theta})}\,d\bar{\theta}$$

$$l(\bar{\theta}) = \log p(D|\bar{\theta}) = \sum_{n=1}^{N}\log p(d_n|\bar{\theta})$$

$$l(\bar{\theta}) \approx l(\bar{\theta}_{ML}) - \frac{N}{2}(\bar{\theta}-\bar{\theta}_{ML})^T J(\bar{\theta}_{ML})(\bar{\theta}-\bar{\theta}_{ML}),\ \text{vge}$$

$$J(\bar{\theta}_{ML}) = \frac{1}{N}\left(-\frac{\partial^2\log p(D|\bar{\theta})}{\partial\theta_i\partial\theta_j}\Big|_{\bar{\theta}_{ML}}\right)$$

$$p(\bar{\theta}) \approx p(\bar{\theta}_{ML}) + (\bar{\theta}-\bar{\theta}_{ML})^T\nabla_{\bar{\theta}}p(\bar{\theta})\Big|_{\bar{\theta}_{ML}}$$

$$p(D) \approx \int \left(p(\bar\theta_{ML}) + \cancel{(\bar\theta-\bar\theta_{ML})^T \nabla_\theta p(\bar\theta_{ML})} \right) e^{l(\bar\theta_{ML}) - \frac{N}{2}(\_\,\_)^T J(\_\,\_)} d\bar\theta =$$

$$\propto \mathbb{E}_{\mathcal{N}(\_|\bar\theta_{ML}\_)} \{ \bar\theta - \bar\theta_{ML} \} = 0 \qquad \propto \mathcal{N}(\bar\theta | \bar\theta_{ML}, J^{-1})$$

$$= e^{l(\bar\theta_{ML})} p(\bar\theta_{ML}) \cdot \int e^{-\frac{N}{2}(\bar\theta-\bar\theta_{ML})^T J(\bar\theta_{ML})(\bar\theta-\bar\theta_{ML})} d\bar\theta =$$

$$= e^{l(\bar\theta_{ML})} p(\bar\theta_{ML}) (2\pi)^{d/2} \left(\det N J(\bar\theta_{ML})\right)^{-1/2} =$$

$$= e^{l(\bar\theta_{ML})} p(\bar\theta_{ML}) (2\pi)^{d/2} N^{-d/2} (\det J)^{-1/2}$$

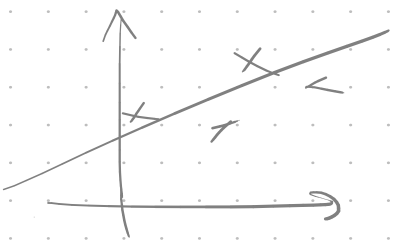
$$\log p(D) \approx \log p(D|\bar\theta_{ML}) + \underbrace{\log p(\bar\theta_{ML}) + \frac{d}{2}\log 2\pi - \frac{d}{2}\log N - \frac{1}{2}\log\det J(\bar\theta_{ML})}_{\text{Occam's factor}}$$

$$\boxed{BIC(\mathcal{M}) = -2\log p(D|\bar\theta_{ML}) + d\log N}$$

$d = \dim \bar\theta$

$N = |D|$

5 Empirical Bayes



hyperparameters

$$p(y|\bar x, \bar w) = \mathcal{N}(y|\bar w^T \bar x, \sigma^2) \qquad \beta = 1/\sigma^2$$

$$p(\bar w) = \mathcal{N}(\bar w | \bar 0, \sigma_0^2 I) \qquad \alpha = 1/\sigma_0^2$$

$$\log p(\bar w|\bar y, X) = \text{Const} - \frac{1}{2\sigma^2}\sum_n (y_n - \bar w^T \bar x_n)^2 - \frac{1}{2\sigma_0^2}\bar w^T \bar w$$

$$\log p(\bar w|\bar y, X, \alpha, \beta) = \text{Const} - \frac{\beta}{2}\sum_n (\_)^2 - \frac{\alpha}{2}\bar w^T \bar w$$

$\boxed{\text{MLE-II}}$

$$p(\bar w|D, \alpha, \beta) = \frac{p(\bar w|\alpha)\, p(D|\bar w, \beta)}{p(D|\alpha,\beta)} \qquad p(D|\alpha,\beta) \xrightarrow[\alpha,\beta]{} \max$$

marginal likelihood

$$E(\bar{\omega}) = -\frac{\alpha}{2}\bar{\omega}^T\bar{\omega} - \frac{\beta}{2}(\bar{y} - X\bar{\omega})^T(\bar{y} - X\bar{\omega}) =$$

$$= -\frac{\alpha}{2}\bar{\omega}^T\bar{\omega} - \frac{\beta}{2}\left(\bar{y}^T\bar{y} - 2\bar{\omega}^T X^T\bar{y} + \bar{\omega}^T X^T X\bar{\omega}\right)$$

$$= -\frac{1}{2}(\bar{\omega} - \bar{\mu})^T A(\bar{\omega} - \bar{\mu}) - \frac{\beta}{2}\bar{y}^T\bar{y} + \frac{1}{2}\bar{\mu}^T A\bar{\mu}$$

$$\boxed{A = \beta X^T X + \alpha \cdot I}$$

$$A\mu = \beta \cdot X^T\bar{y}, \quad \bar{\mu} = \beta \cdot A^{-1}X^T y$$

$$E(\bar{\mu}) = -\frac{\alpha}{2}\bar{\mu}^T\bar{\mu} - \frac{\beta}{2}(\bar{y} - X\bar{\mu})^T(\bar{y} - X\bar{\mu})$$

$$= -\frac{\beta}{2}\left(\bar{\mu}^T X^T X\bar{\mu} + \bar{y}^T\bar{y} - 2\bar{\mu}^T X^T\bar{y}\right) - \frac{\alpha}{2}\bar{\mu}^T\bar{\mu}$$

$$\beta \cdot \bar{\mu}^T X^T\bar{y} = \bar{\mu}^T A \cdot (\beta \cdot A^{-1}X^T\bar{y}) = \mu^T A\bar{\mu} = \beta\bar{\mu}^T X^T X\bar{\mu} + \alpha \cdot \bar{\mu}^T\bar{\mu}$$

$$E(\bar{\omega}) = E(\bar{\mu}) - \frac{1}{2}(\bar{\omega} - \bar{\mu})^T A(\bar{\omega} - \bar{\mu})$$

$$-\frac{\beta}{2}\bar{y}^T\bar{y} + \frac{1}{2}\bar{\mu}^T A\bar{\mu} = -\frac{\beta}{2}\bar{y}^T\bar{y} + \frac{\alpha}{2}\bar{\mu}^T\bar{\mu} + \frac{\beta}{2}\bar{\mu}^T X^T X\bar{\mu} =$$

$$= -\frac{\beta}{2}\left(\bar{y}^T\bar{y} + \bar{\mu}^T X^T X\bar{\mu} - 2\bar{\mu}^T X^T\bar{y}\right) - \beta\bar{\mu}^T X^T\bar{y} + \beta \cdot \bar{\mu}^T X^T X\bar{\mu} + \frac{\alpha}{2}\bar{\mu}^T\bar{\mu}$$

$$= -\frac{\beta}{2}(\bar{y} - X\bar{\mu})^T(\bar{y} - X\bar{\mu}) - \frac{\alpha}{2}\bar{\mu}^T\bar{\mu}$$

$$\log p(\bar{y} | X, \alpha, \beta) = \log \int \left(\frac{\alpha}{2\pi}\right)^{d/2}\left(\frac{\beta}{2\pi}\right)^{N/2} e^{E(\bar{\omega})}\, d\bar{\omega} =$$

$$= \log\left[\left(\frac{\alpha}{2\pi}\right)^{d/2}\left(\frac{\beta}{2\pi}\right)^{N/2} \cdot e^{E(\bar{\mu})} \cdot \int e^{-\frac{1}{2}(-)A(-)}\, d\omega\right] =$$

$$= \frac{d}{2}\underbrace{\log \alpha} + \frac{N}{2}\underbrace{\log \beta} - \frac{1}{2}\log\det A - \frac{\alpha}{2}\bar{\mu}^T\bar{\mu} -$$

$$- \frac{\beta}{2}(\bar{y} - X\bar{\mu})^T(\bar{y} - X\bar{\mu}) - \frac{N+d}{2}\log 2\pi \xrightarrow[\alpha,\beta]{} \max$$