$$EPE[f] = \text{Bias}^2 + \text{Variance} + \text{Noise}$$

$$\mathbb{E}[(\mathbb{E}f - \hat{f})^2] + \mathbb{E}[(f - \mathbb{E}f)^2] + \mathbb{E}[(y - \hat{f})^2]$$

Bayesian model selection

D - $M_1, M_2, M_3, M_4, \ldots$

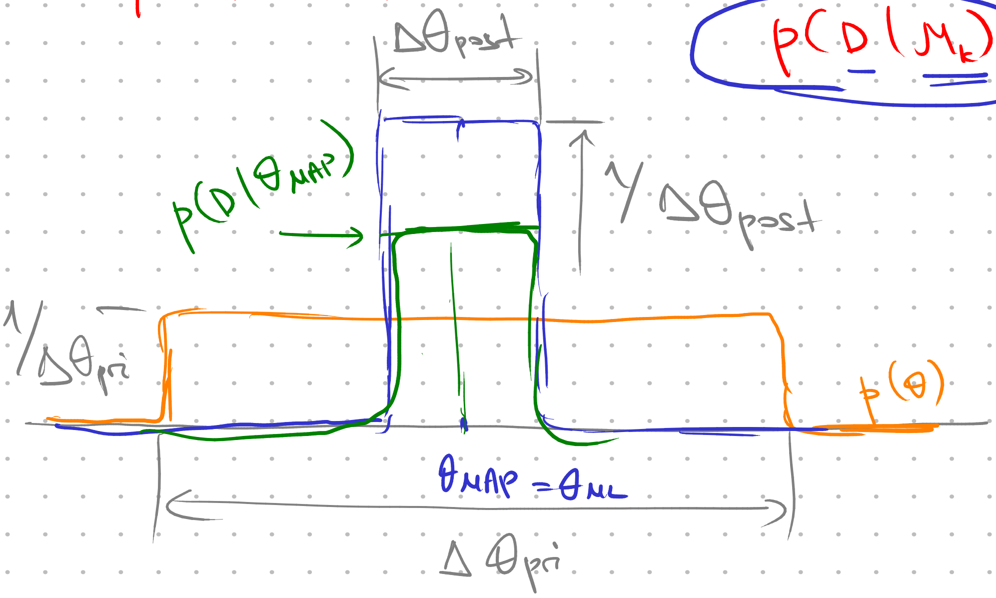
$p(D|\bar{\theta}_1, M_1)$ $\qquad$ $p(D|\bar{\theta}_3, M_3)$

$\bar{\theta}_1^{ML}$ $\qquad$ $\bar{\theta}_3^{ML}$

$p(\bar{\theta}_1|M_1)$

$$\bar{\theta}_1^{MAP} = \text{argmax}\; p(\bar{\theta}_1|M_1)\, p(D|\bar{\theta}_1, M_1)$$

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



$p(D|M_k) =$

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

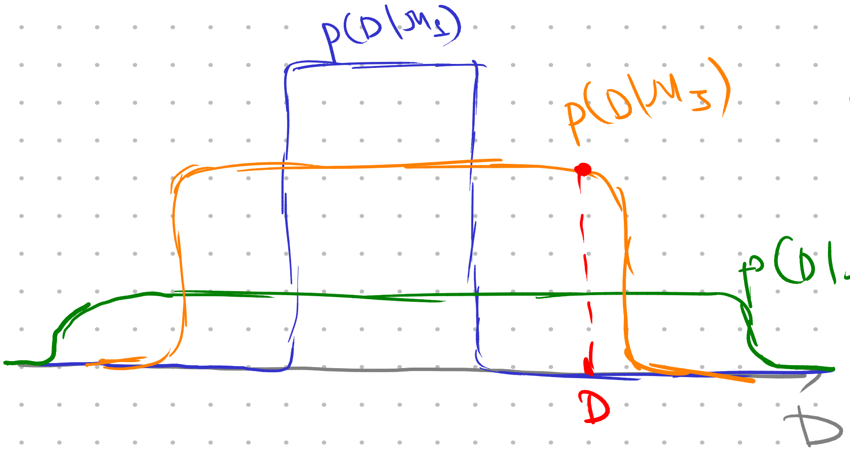
θ

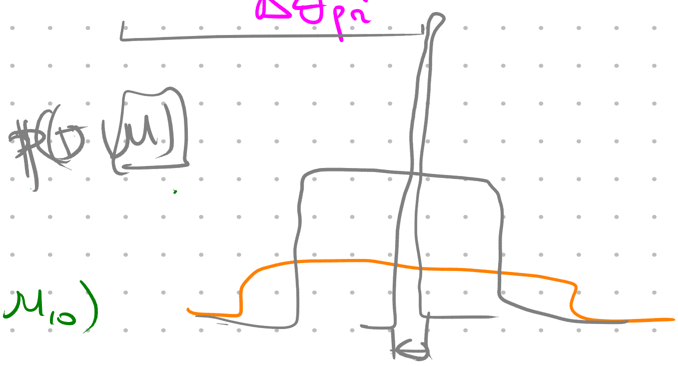
$$\approx \begin{cases} p(D|\theta_{MAP}), & \Delta\theta_{post} \\ 0 & \end{cases}$$

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

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

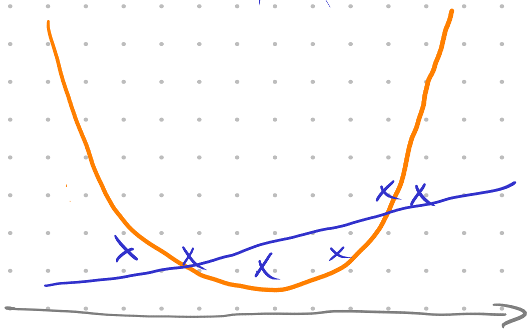
$$\ln p(D) \approx \underline{\ln p(D|\theta_{MAP})} + \underline{M \cdot \ln \frac{\Delta\theta_{post}}{\Delta\theta_{pri}}}$$





$$D \sim \underline{M_{true}} \qquad p(D|M_{true}) \overset{?}{\cancel{=}} p(D|M)$$

$\{(\bar{x}, y)\}$ $\quad y_n \sim p(y|\bar{x}_n, M_{true})$



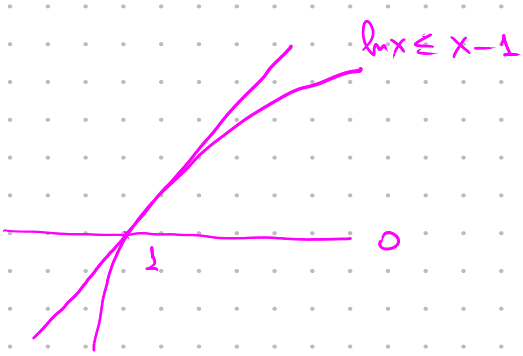
$$\mathbb{E}_{D \sim p(D|M_{true})}\left[\ln \frac{p(D|M_{true})}{p(D|M)}\right] =$$

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

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

$$KL(p\|q) = \int p(\bar{x}) \ln \frac{p(\bar{x})}{q(\bar{x})} \, d\bar{x} \overset{?}{\geq} 0$$

$$= -\int p(x) \ln \frac{q(\bar{x})}{p(\bar{x})} \, d\bar{x} \geq$$



$$\geq -\int p(x)\left(\frac{q(x)}{p(x)} - 1\right) dx =$$

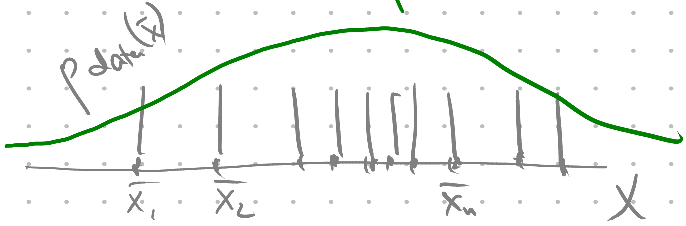
$$= -\int q(\bar{x}) dx + \int p(\bar{x}) d\bar{x} = 0$$

$$\ln p(D|\bar\theta) = \sum_{n=1}^{N} \ln p(\bar x_n|\bar\theta) \longrightarrow \max$$

$D = \{\bar x_1, \dots \bar x_N\}$

$$p_{model}(\bar x|\bar\theta) \approx p_{data}(\bar x) = \mathrm{Unif}(D) =$$

$$= \frac{1}{N} \cdot \sum_n \delta(\bar x - \bar x_n)$$



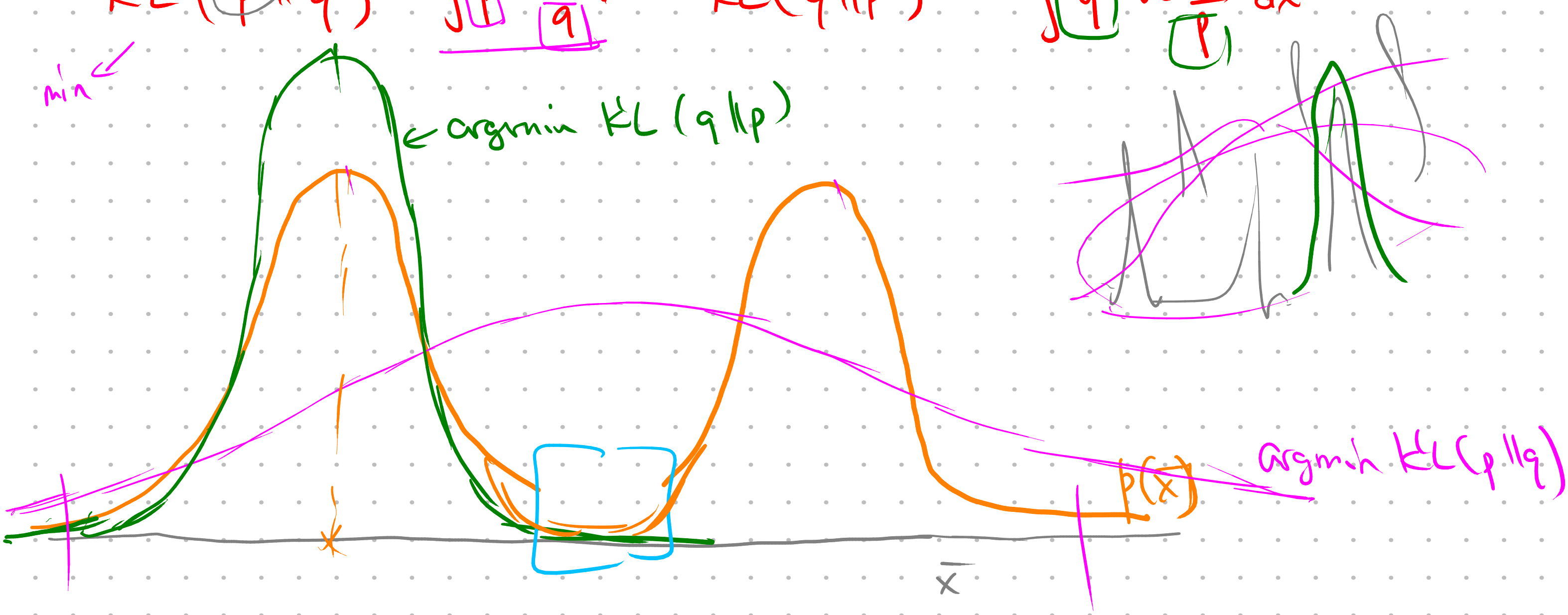
$KL(p_{model} \| p_{data})$

$$KL(p_{data} \| p_{model}) = \int p_{data}(\bar x) \ln \frac{p_{data}(\bar x)}{p_{model}(\bar x)} d\bar x =$$

$$= \int p_{data}(\bar x) \ln p_{data}(\bar x)\, d\bar x - \underline{\int p_{data}(\bar x) \ln p_{model}(\bar x)\, d\bar x} =$$

$$= \int \dots d\bar x - \sum_{n=1}^{N} \ln p_{model}(\bar x_n) \xrightarrow[\bar\theta]{} \min$$

$$KL(p\|q) = \int \boxed{p} \ln \frac{p}{\boxed{q}} dx \qquad KL(q\|p) = \int \boxed{q} \ln \frac{q}{\boxed{p}} dx$$



$$\underline{p(D)} = \int \underbrace{p(D|\theta)\,p(\theta)}_{\approx \mathcal{N}}\,d\theta$$

θ_{MAP}

$$p(D) = \int f(\bar\theta)\,d\bar\theta \approx \int f(\bar\theta_0)\, e^{-\frac{1}{2}(\bar\theta-\bar\theta_0)^T A(\bar\theta-\bar\theta_0)}\,d\bar\theta =$$

$$= f(\bar\theta_0)\sqrt{\frac{(2\pi)^M}{\det A}}$$

$A = \nabla\nabla \ln f(\bar\theta)$

max a posteriori

$$\ln p(D) \approx \ln p(D|\theta_{MAP}) + \underbrace{\ln p(\theta_{MAP}) + \frac{M}{2}\ln 2\pi - \frac{1}{2}\ln|A|}_{\text{Occam's factor}}$$

$$\boxed{\ln p(D) \approx \ln p(D|\theta_{MAP}) - \frac{1}{2} M \ln N} \quad \text{BIC}$$