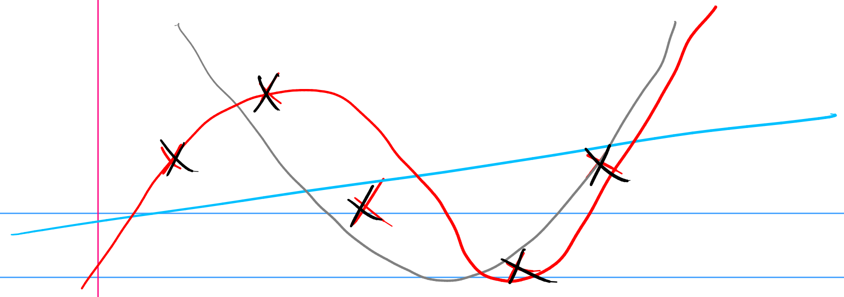
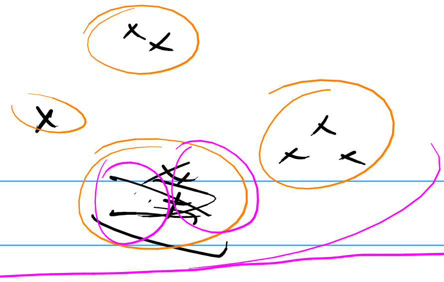




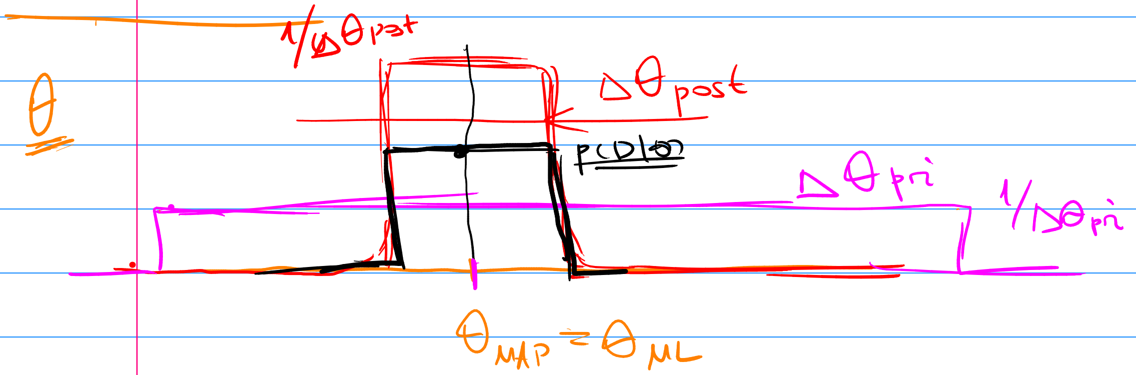
D

$\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_K$ prior

$$p(\mathcal{M}_i \mid D) \propto p(\mathcal{M}_i)\, p(D \mid \mathcal{M}_i)$$

$$p(\theta \mid D, \mathcal{M}_i) = \frac{p(\theta \mid \mathcal{M}_i)\, p(D \mid \theta, \mathcal{M}_i)}{p(D \mid \mathcal{M}_i)}$$

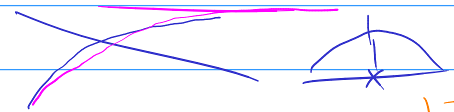
$$p(D \mid \mathcal{M}_i) = \int p(\theta \mid \mathcal{M}_i)\, p(D \mid \theta, \mathcal{M}_i)\, d\theta$$



$$\bar{\theta} = (\theta_1, \theta_2, \dots, \theta_M)$$

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

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

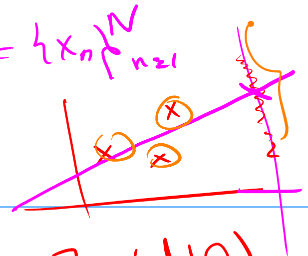


$$\ln p(D) \approx \ln p(D \mid \theta^*) - M \cdot \ln \frac{\Delta\theta_{pri}}{\Delta\theta_{post}}$$

$$p(D) = \int\!\!\int_{\Delta\theta_{post}} \left(\frac{1}{\Delta\theta_{pri}}\right)^M p(D \mid \theta_{ML})\, d\theta_1 \dots d\theta_M = p(D \mid \theta_{ML}) \left(\frac{\Delta\theta_{post}}{\Delta\theta_{pri}}\right)^M$$

BIC – Bayesian Inform. Criterion

$D = \{x_n\}_{n=1}^N$

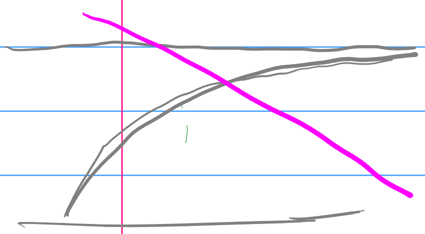


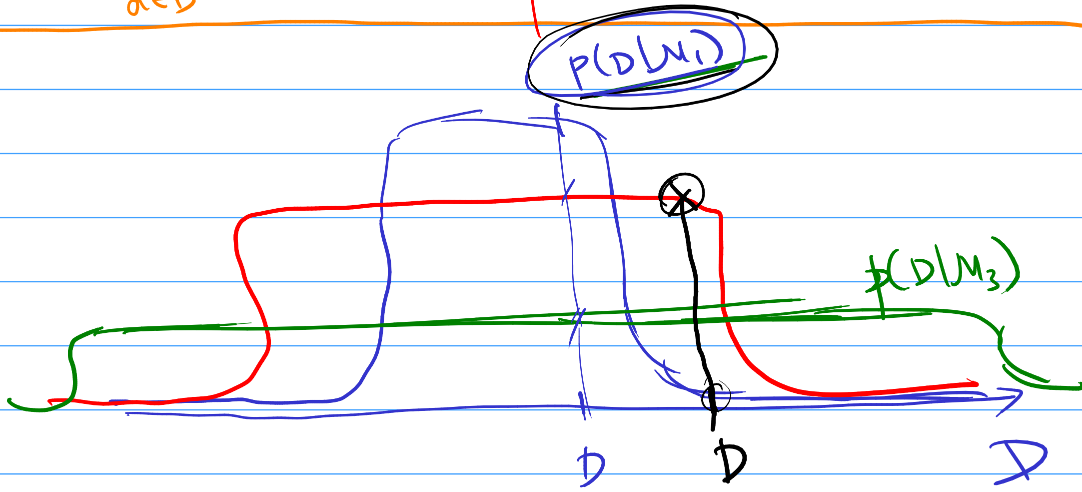
$$\ln p(D) \approx \ln p(D|\theta^*) - \frac{1}{2} M \ln N$$

$$\ln p(D|\theta^*) = \sum_{d \in D} \ln p(d|\theta^*)$$

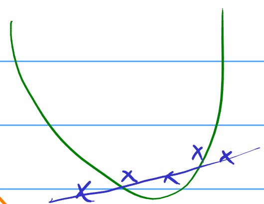
$$p(D|\theta) = \prod_{d \in D} p(d|\theta)$$

$\mathcal{M}_1$ $\mathcal{M}_2$ $\mathcal{M}_3$





$$\boxed{p(D|\mathcal{M}_{true})}$$



$$p(D|\mathcal{M}_{true}) \overset{?}{\gtrless} p(D|\mathcal{M})$$

$$p(D|\mathcal{M}_{true}) = \int p(\bar w|\mathcal{M}_{true})\, p(D|\bar w, \mathcal{M}_{true})\, d\bar w$$

$$\ln \frac{p(D|\mathcal{M}_{true})}{p(D|\mathcal{M})} \overset{?}{\gtrless} 0$$

$$p(y|\bar x, D)$$

$$\int p(D|\mathcal{M}_{true}) \ln \frac{p(D|\mathcal{M}_{true})}{p(D|\mathcal{M})}\, dD = KL\big(p(D|\mathcal{M}_{true}) \,\|\, p(D|\mathcal{M})\big)$$

Kullback-Leibler divergence

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

$\boxed{KL(p\|q) \geq 0}$ упр

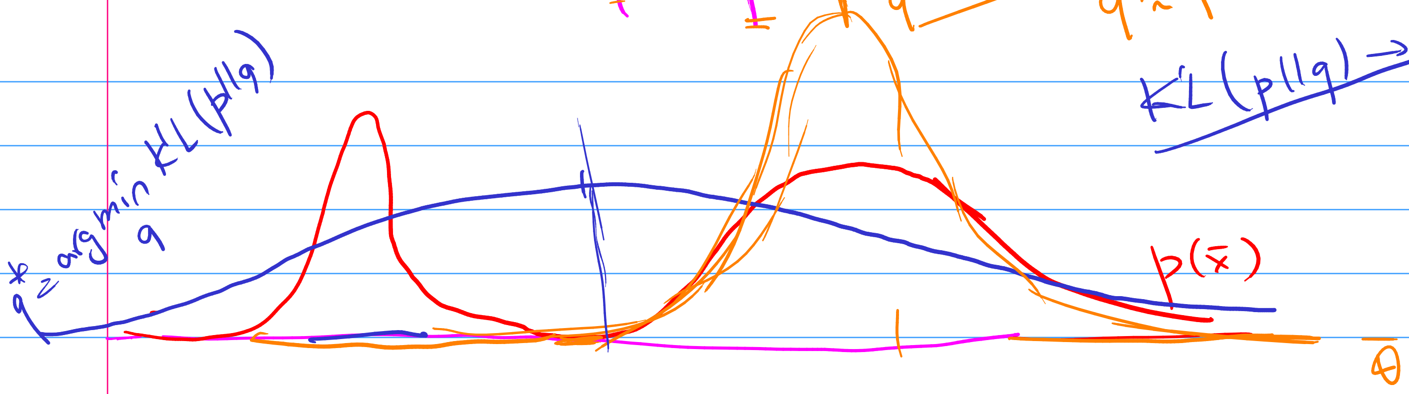
$KL(p\|q) = 0 \Leftrightarrow p \overset{a.e.}{=} q$

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

$q^* = \arg\min_q KL(q\|p)$

$q \approx p$

$KL(p\|q) \to \min$



$KL(q\|p) \to \min$

$$JSD(p,q) = KL\left(p \,\Big\|\, \frac{p+q}{2}\right) + KL\left(q \,\Big\|\, \frac{p+q}{2}\right)$$

$p(D|\mu)$

$$p(D|\bar{w}) = \prod \mathcal{N}(y_n | \bar{x}_n^T \bar{w}, \beta^{-1})$$

$$p(\bar{w}) = \prod \mathcal{N}(w_i | 0, \alpha^{-1})$$ — hyperparameters

$$p(\alpha, \beta | D) \propto p(D|\alpha,\beta) \, p(\alpha,\beta)$$ $\approx$ const

$$\left| p(D|\alpha,\beta) \right| \xrightarrow[\alpha,\beta]{} \max$$

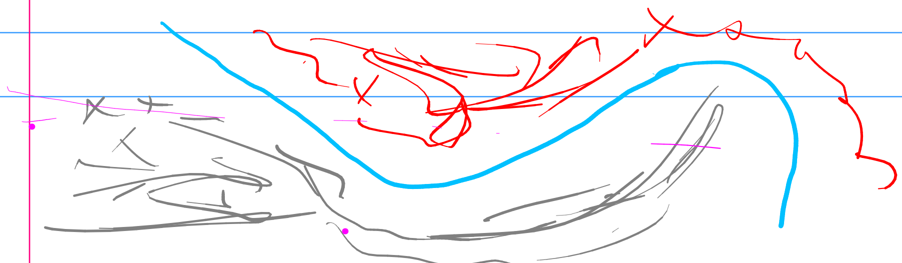
$$p(D|\alpha,\beta) = \int p(D|\bar{w},\beta) \, p(\bar{w}|\alpha) \, d\bar{w}$$

$$\ln p(D|\alpha,\beta) = \frac{M}{2} \ln \alpha + \frac{N}{2} \ln \beta - \frac{1}{2} \ln \det A - \cancel{\frac{N}{2} \ln 2\pi}$$
$$- \frac{\beta}{2} \sum_n (y_n - \bar{\mu}_N^T \bar{x}_n)^2 - \frac{\alpha}{2} \bar{\mu}_N^T \bar{\mu}_N$$

$$\xrightarrow[\alpha,\beta]{} \max$$

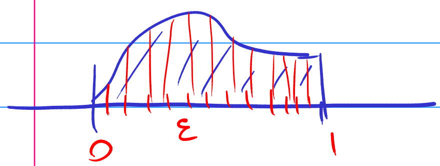
$$A = \beta X^T X + \alpha I$$

$\boxed{\text{Empirical Bayes}}$



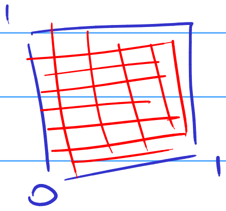
Curse of dimensionality

$d=1$



$1/\varepsilon$ $\quad d \quad$ $1/\varepsilon^d$

$d=2$



$1/\varepsilon^2$

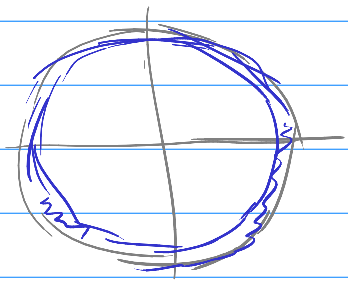
$$p(x\mid D) = \int p(x\mid\theta)\,p(\theta\mid D)\,d\theta$$

$$\propto p(\theta)\,p(D\mid\theta)$$

$d=1$

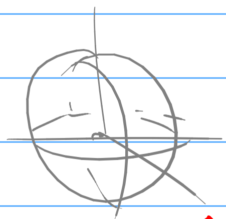
ε

$d=2$



$$\frac{\pi\cdot 1^2 - \pi(1-\varepsilon)^2}{\pi\cdot 1^2} = 1-(1-\varepsilon)^2$$

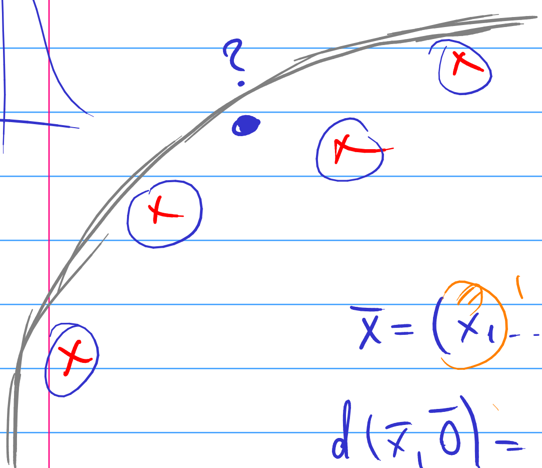
$d=3$

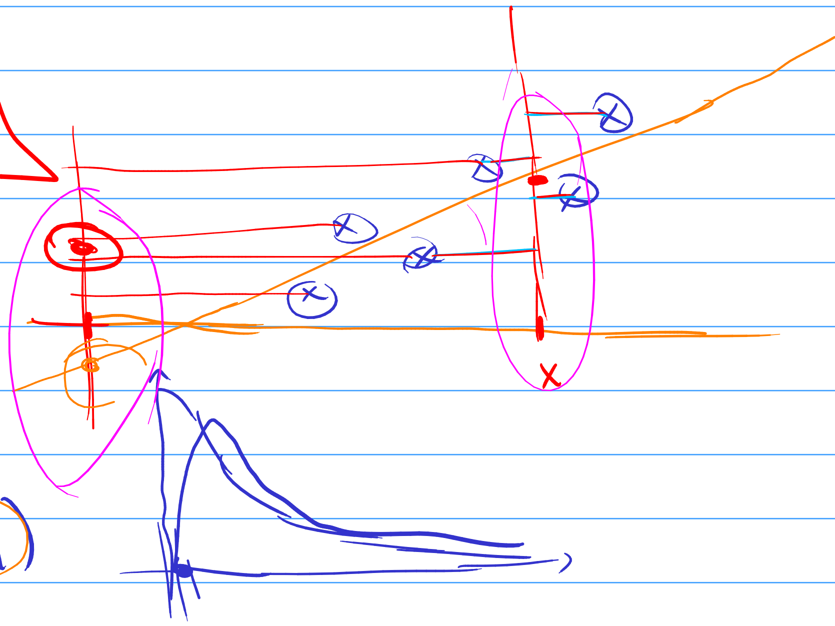


$$\frac{4/3\,\pi\cdot 1^3 - 4/3\,\pi(1-\varepsilon)^3}{4/3\,\pi\cdot 1^3} = 1-(1-\varepsilon)^3$$

d:

$$\boxed{1-(1-\varepsilon)^d}$$





$$\bar{x} = (x_1 \dots x_d)$$

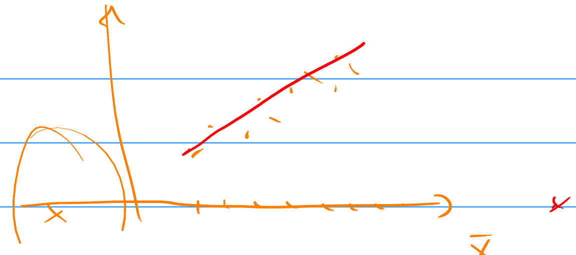
$$d(\bar{x},\bar{0}) = x_1^2 + x_2^2 + \dots + x_d^2$$

$$d(\bar{x}, \bar{x}_n) = \sum (x_i - x_{in})^2$$

$\bar{x} \in \mathbb{R}^d \qquad y \in \mathbb{R} \qquad p(\bar{x}, y) \;\Big|\; p(y|\bar{x}) \cdot p(\bar{x})$

$f: \bar{x} \longmapsto y$

$L(y, f(\bar{x})) = (y - f(\bar{x}))^2$



$$EPE[f] = E_p L = \iint (y - f(\bar{x}))^2 p(\bar{x}, y)\, d\bar{x}\, dy \xrightarrow[f]{} \min$$

$$= \int \left[\int (y - f(\bar{x}))^2 p(y|\bar{x})\, dy \right] p(\bar{x})\, d\bar{x}$$

regression function $\qquad f(\bar{x}) \to \min \qquad\qquad f(x) \to \min$

$$\hat{f}(\bar{x}) = E_{p(y|\bar{x})}[y] \qquad\qquad \sum_y p(y|\bar{x}) \cdot [y \neq f(\bar{x})]$$

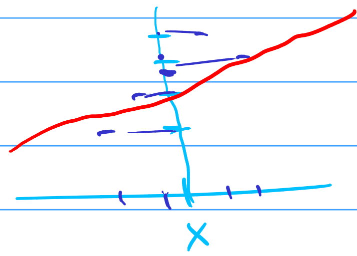
$$L(y, f(\bar{x})) = [y \neq f(\bar{x})]$$

$$EPE[f] = \iint [y \neq f(\bar{x})] p(\bar{x}, y)\, d\bar{x}\, dy = \int \left[\int [y \neq f(\bar{x})] p(y|\bar{x})\, dy \right] p(\bar{x})\, d\bar{x}$$

$$\hat{f}(\bar{x}) = \operatorname*{argmax}_y p(y|\bar{x})$$

$$\hat{f}(\bar{x}) = E_{p(y|\bar{x})}[y] \approx \frac{1}{R} \sum_{r=1}^{R} y_r \approx, \qquad y_r \sim p(y|\bar{x})$$

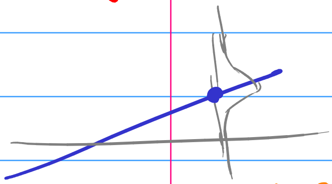
$$\approx \frac{1}{R} \sum_{\bar{x}_r \in kNN(\bar{x})} y_r$$



$$EPE[f] = \iint (y - f(\bar{x}))^2 p(\bar{x}, y)\, d\bar{x}\, dy =$$

$$= \iint (y - \hat{f}(\bar{x}) + \hat{f}(\bar{x}) - f(\bar{x}))^2 \ldots =$$

$$= \iint (y-\hat f(\bar x))^2 p(\bar x, y)\, d\bar x\, dy \cancel{- 2\iint (y-\hat f)(f-\hat f)\, p(\bar x,y)\, d\bar x\, dy} + \iint (\hat f - f)^2 p\, d\bar x\, dy$$



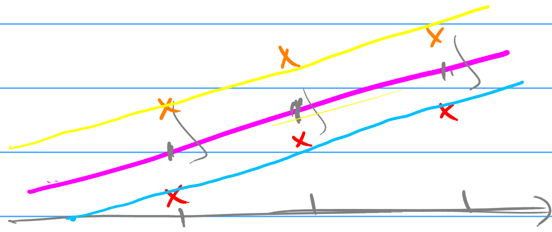
$$\parallel$$

$$-2\int \underbrace{\left[\int (y-\hat f(\bar x))\, p(y|\bar x)\, dy\right]}_{=0} (f-\hat f)\, p(\bar x)\, d\bar x$$

$$= \overbrace{\mathbb{E}_{p(\bar x,y)}\left[(y-\hat f(\bar x))^2\right]}^{\text{NOISE}} + \overbrace{\mathbb{E}_p\left[(f(\bar x)-\hat f(\bar x))^2\right]}$$

$$\iint (f(\bar x; D) - \hat f(\bar x))^2\, p(\bar x,y)\, d\bar x\, dy =$$

$$\mathbb{E}_{D\sim p(\bar x,y)}[f(\bar x;D)] = \mathbb{E}_D f$$



$$= \int (f - \mathbb{E}_D f + \mathbb{E}_D f - \hat f)^2\, p(\bar x)\, d\bar x =$$

$$= \int (f(\bar x) - \mathbb{E}_D f)^2\, p(\bar x)\, d\bar x + \overbrace{\cancel{2\int (f(\bar x;D) - \mathbb{E}_D f)(\mathbb{E}_D f - \hat f)\, p(\bar x)\, d\bar x}}^{=0} + \int (\hat f - \mathbb{E}_D f)^2\, p(\bar x)\, d\bar x$$

$$EPE[f] = \mathbb{E}_p\left[(\hat f(\bar x) - \mathbb{E}_D[f])^2\right] + \qquad \text{Bias}^2$$

$$+ \mathbb{E}_p\left[(f(\bar x;D) - \mathbb{E}_D[f])^2\right] + \qquad \text{Variance}$$

$$+ \mathbb{E}_p\left[(\hat f(\bar x) - y)^2\right] \qquad \text{Noise}$$