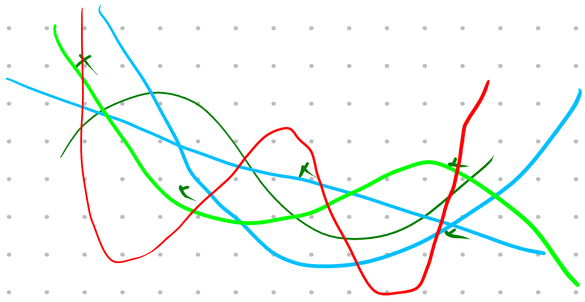


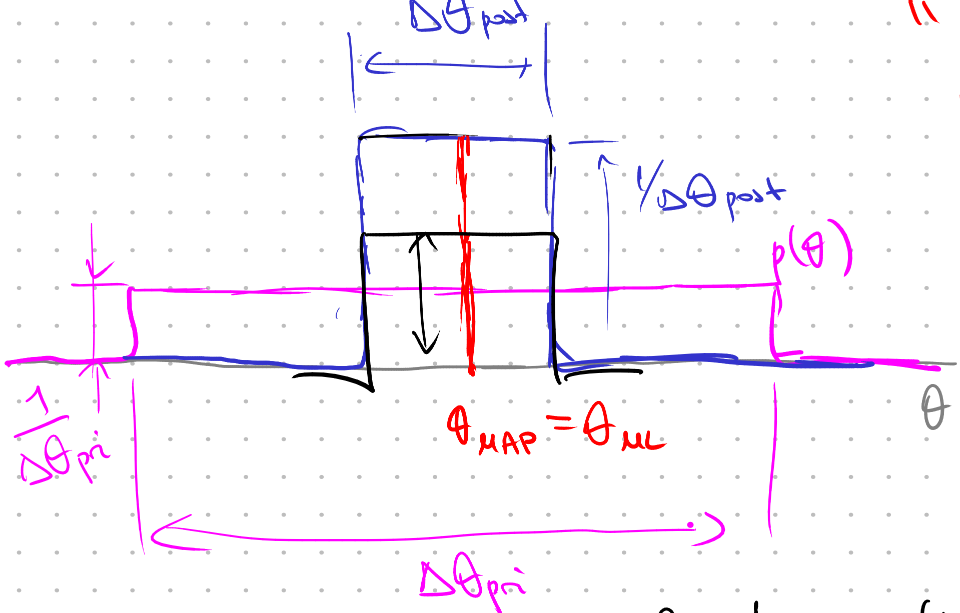
$D \quad M_1, M_2, \dots, M_K$

$$p(M_i \mid D) \propto \underbrace{\cancel{p(M_i)}}_{\text{prior}}\, p(D \mid M_i)$$

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

$$\underline{p(D \mid M_i)} = \int p(\theta \mid M_i)\, p(D \mid \theta, M_i)\, d\theta$$

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



M парам.

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

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

$$\ln p(D) = \ln p(D \mid \theta_{ML}) - M \cdot \ln \frac{\Delta\theta_{pri}}{\Delta\theta_{post}}$$

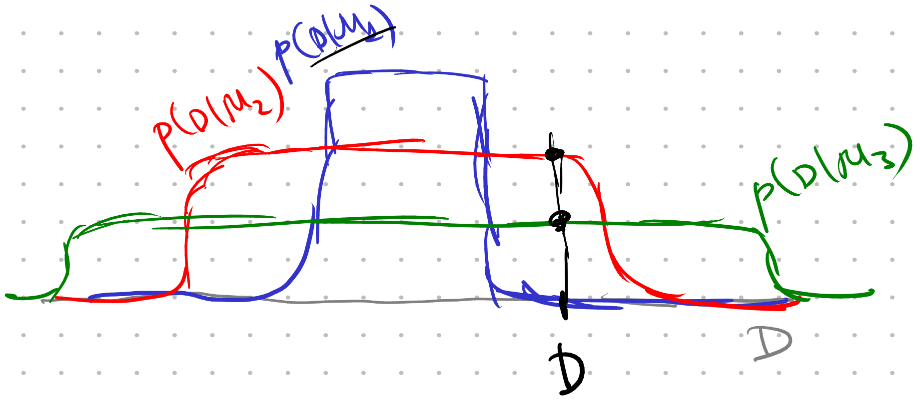
$p(\theta),\ p(\theta \mid D) \approx N(\ldots)$

BIC - Bayesian Information Criterion

число параметров

число точек в D

$$\ln p(D) \approx \ln p(D \mid \theta_{ML}) - \frac{1}{2} M \ln N$$



Kullback-Leibler divergence

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

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

Утв.

① $KL(p\|q) \geq 0$

② $KL(p\|q) = 0 \iff p = q$ a.e.

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

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

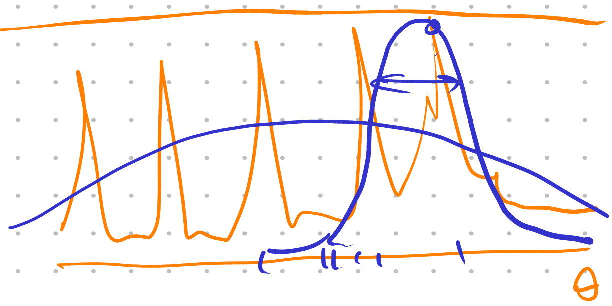
$$\int p_{data}(x) f(x) dx \approx \sum_{x \in D} f(x)$$

$$\ln p(D|\theta) \to \max \qquad \sum_{x \in D} \ln p(x|\theta) = \int p_{data}(x) \ln p(x|\theta) dx \xrightarrow[\theta]{} \max$$

$$KL(p_{data}\|p) = \int p_{data}(x) \ln \frac{p_{data}(x)}{p(x|\theta)} dx \xrightarrow[\theta]{} \min$$

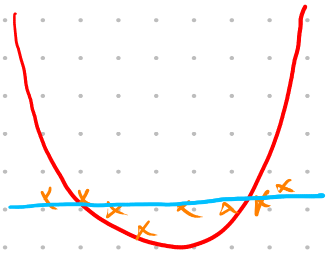
$$p(\theta) = \boxed{p(\theta|D)} \propto p(\theta) p(D|\theta)$$

$$\approx q(\theta)$$



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

Sanity check:



$$p(D|M_{true}) \Rightarrow p(D|M_{true}) \overset{?}{\geq} p(D|M)$$

$$\ln p(D|M_{true}) \overset{?}{\geq} \ln p(D|M)$$

$$\int p(D|M_{true}) \ln p(D|M_{true})\, dD - \int \ln p(D|M)\, p(D|M_{true})\, dD \overset{?}{\geq} 0$$

$$KL(p_{true}\|p) = \int p(D|M_{true}) \ln \frac{p(D|M_{true})}{p(D|M)}\, dD \geq 0$$

Equivalent kernel

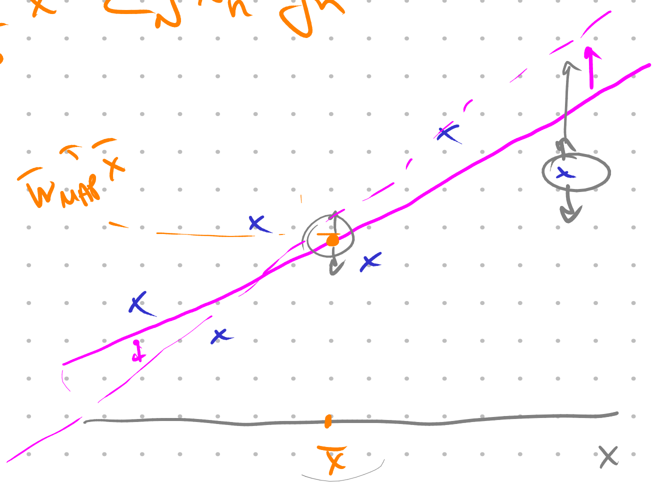
$$p(y|\bar{x}, D) = \mathcal{N}(y \,|\, \bar{w}_{MAP}^T \bar{x}, \sigma^2), \quad \sigma^2 = \sigma_0^2 + \bar{x}^T \Sigma_N \bar{x}$$

$$\bar{w}_{MAP} = \frac{1}{\sigma_0^2} \Sigma_N X^T \bar{y}$$

$$\bar{x}^T \bar{w}_{MAP} = \frac{1}{\sigma_0^2} \bar{x}^T \Sigma_N X^T \bar{y} = \sum_{n=1}^{N} \frac{1}{\sigma_0^2} \bar{x}^T \Sigma_N \bar{x}_n y_n$$

$$X^T \bar{y} = \sum_{n=1}^{N} \bar{x}_n y_n$$

$$\bar{x}^T \bar{w}_{MAP} = \sum_{n=1}^{N} \underbrace{k(\bar{x}, \bar{x}_n)}_{\text{kernel}} \cdot y_n$$



Эмпирический Байес   Empirical Bayes

$$\beta = \frac{1}{\sigma^2}$$

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

$$p(\bar{w}) = \prod_i \mathcal{N}(w_i \,|\, 0, \frac{1}{\alpha})$$

hyperparameters

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

non-informative priors

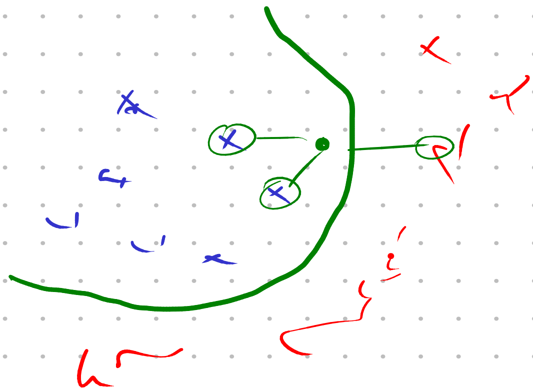
$$p(D|\alpha,\beta) = \int p(D|\bar{w},\beta)\, p(\bar{w}|\alpha)\, d\bar{w}$$

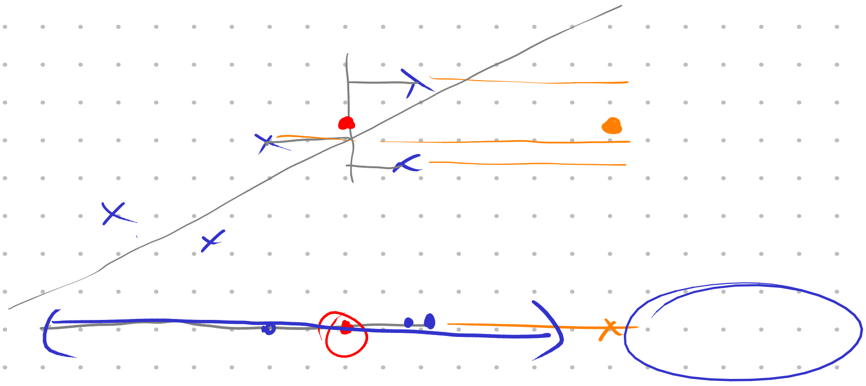
$$\boxed{\ln p(D|\alpha,\beta) = \frac{d}{2}\ln\alpha + \frac{N}{2}\ln\beta - \frac{1}{2}\ln\det A - \frac{N}{2}\ln 2\pi - \frac{\beta}{2}\sum_n (y_n - \bar{\mu}_N^T \bar{x}_n)^2 - \frac{\alpha}{2}\bar{\mu}_N^T \bar{\mu}_N}$$

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

$$\bar{\mu}_N = \dots$$

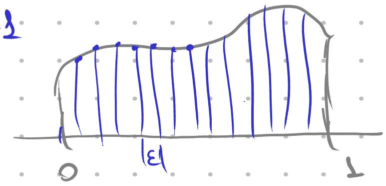
Nearest neighbors





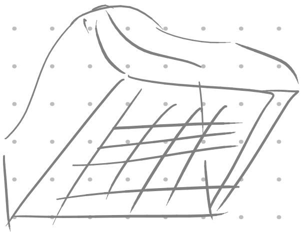
Curse of dimensionality

$d=1$



$1/\varepsilon$

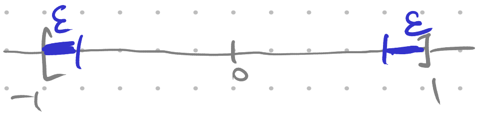
$d=2$



$1/\varepsilon^2$

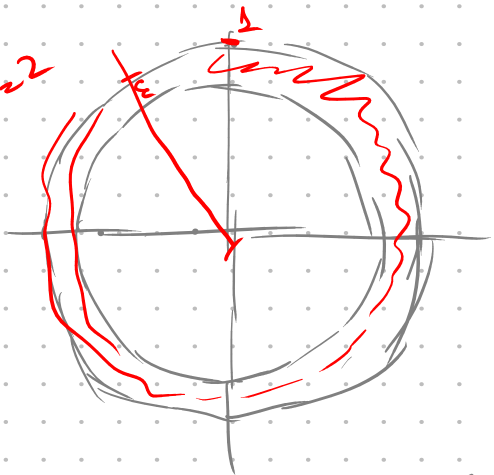
$$d(x,y) = \sum_{i=1}^{d} (x_i - y_i)^2$$

$d=1$



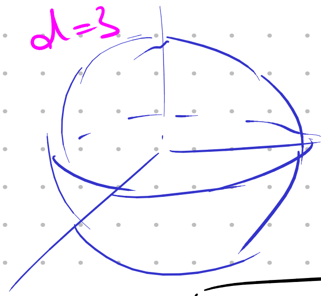
ε

$d=2$



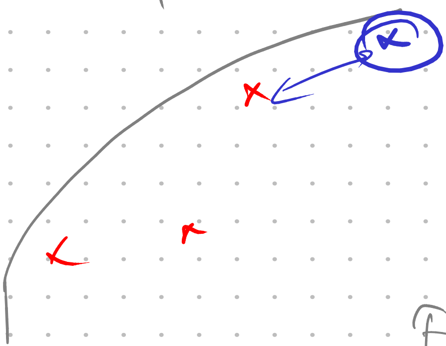
$$\frac{\pi\cdot 1^2 - \pi(1-\varepsilon)^2}{\pi} = 1-(1-\varepsilon)^2$$

$d=3$



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

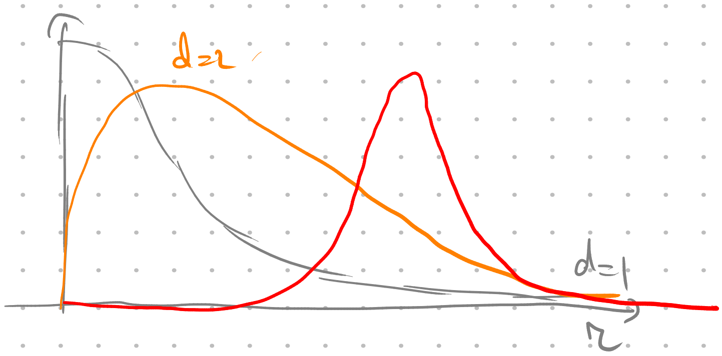
$$\boxed{1-(1-\varepsilon)^d} \quad d\to\infty$$



$$\bar{x} \sim \mathcal{N}(\bar{0}, \mathbb{I})$$

$$d(\bar{x}, 0) = \sum x_i^2$$

$$x_i \sim \mathcal{N}(0,1)$$



Statistical decision theory

$\bar{x} \in \mathbb{R}^d \qquad D = \{(\bar{x}_n, y_n)\}$

$y \in \mathbb{R} \qquad p(\bar{x}, y)$

$f(\bar{x}) \qquad f: \bar{x} \mapsto y$

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

$= p(\bar{x})\, p(y|\bar{x})$

$$\underset{\text{exp. pred. error}}{EPE[f]} = \mathbb{E}_{p(\bar{x},y)}\left[(f(\bar{x}) - y)^2\right] = \int p(\bar{x},y)(f(\bar{x})-y)^2 \, d\bar{x}\, dy \xrightarrow[f]{} \min$$

$$= \underbrace{\int \underbrace{\left(\int (f(\bar{x})-y)^2 p(y|\bar{x})\, dy\right)}_{f(\bar{x}) \to \min} p(\bar{x})\, d\bar{x}}_{f \to \min}$$

$\int (z-a)^2 p(z)\, dz \xrightarrow[a]{} \min$

$a = \mathbb{E}[z]$

regression function

$$\boxed{\hat{f}(\bar{x}) = \mathbb{E}_{p(y|\bar{x})}[y]}$$

$$\hat{f}(\bar{x}) = \mathbb{E}_{p(y|\bar{x})}[y] \approx \frac{1}{R}\sum_{r=1}^{R} y_r \approx \frac{1}{R}\sum_{r=1}^{R} y'_r$$

y'_r – k-NN на $\bar{x}$

$$L(y, y') = \begin{cases} 1, & y \neq y' \\ 0, & y = y' \end{cases}$$

$$EPE[f] = \mathbb{E}_{p(\bar{x},y)}[L(y, f(\bar{x}))] = \int \left(\sum_{k=1}^{K} L(y, f(\bar{x}))\, p(y_k|\bar{x})\right) p(\bar{x})\, d\bar{x}$$

$$\boxed{\hat{f}(\bar{x}) = \arg\max_k p(y_k|\bar{x})}$$

Optimal Bayes classifier

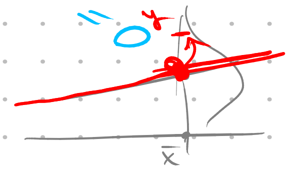
$$EPE[f] = \iint p(\bar{x},y)\,(f(\bar{x})-y)^2\, d\bar{x}\,dy =$$

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

$$\underbrace{\int (\hat{f}(\bar{x})-y)\,p(y|\bar{x})\,dy}_{=0}$$

$$= \iint p(\bar{x},y)\,(f(\bar{x})-\hat{f}(\bar{x}))^2\, d\bar{x}\,dy + \underbrace{2\iint \overset{p(\bar{x})\,p(y|\bar{x})}{p(\bar{x},y)}\,(f(\bar{x})-\hat{f}(\bar{x}))(\hat{f}(\bar{x})-y)\, d\bar{x}\,dy}_{=0}$$

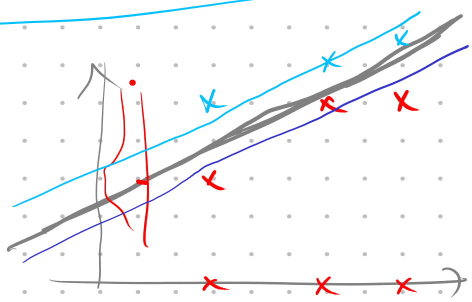
$$+ \iint p(\bar{x},y)\,(\hat{f}(\bar{x})-y)^2\, d\bar{x}\,dy$$



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

$$f(\bar{x}) = f(\bar{x};D)$$

$$D \sim p(\bar{x},y)$$



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

$$= \int p(\bar{x},y)\,(f-\mathbb{E}_D f)^2\, d\bar{x}\,dy + 2\int p(\bar{x},y)\,(f-\mathbb{E}_D f)\underbrace{(\mathbb{E}_D f-\hat{f})}_{\to 0}\, d\bar{x}\,dy + \int p(\bar{x},y)\,(\mathbb{E}_D f-\hat{f})^2\, d\bar{x}\,dy$$

$$EPE[f] = \mathbb{E}_{p(\bar{x},y)}\left[(f(\bar{x};D)-\mathbb{E}_D f(\bar{x};D))^2\right] \qquad \text{variance}$$

$$+ \mathbb{E}_{p(\bar{x},y)}\left[(\mathbb{E}_D f(\bar{x};D)-\hat{f}(\bar{x}))^2\right] \qquad \text{bias}$$

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

Bias-variance-noise decomposition