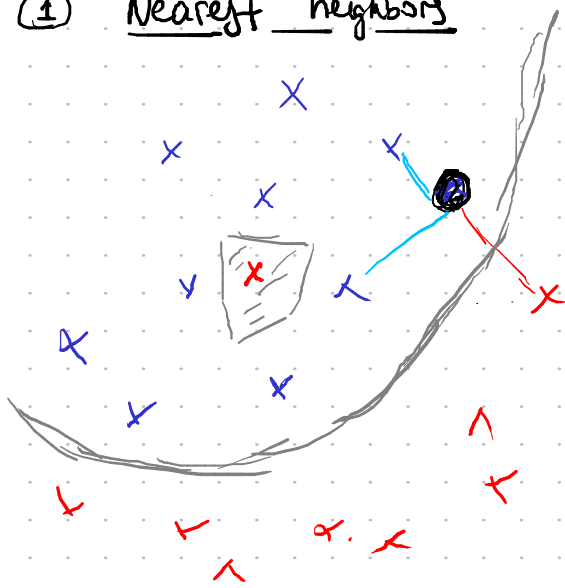


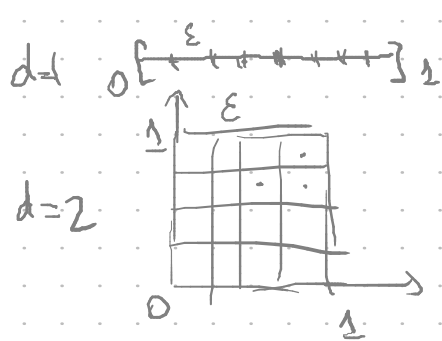
① Nearest neighbors

$k=3$



② Curse of dimensionality

— численное интегрирование



$$\int_0^1 f(x) dx \quad 1/\epsilon$$

$$1/\epsilon^2$$

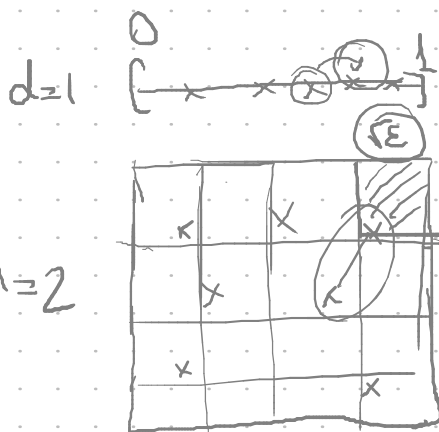
$$d \quad 1/\epsilon^d$$

$$p(\bar{x}|D) = \int p(\bar{x}|\bar{\theta})p(\bar{\theta}|D)d\bar{\theta} = \\ = E_{p(\bar{\theta}|D)}[p(\bar{x}|\bar{\theta})] \approx \frac{1}{R} \sum_{i=1}^R p(\bar{x}|\bar{\theta}^{(i)})$$

МСМС

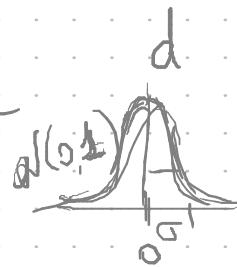
$$\bar{\theta}^{(i)} \sim p(\bar{\theta}|D)$$

— ближайшие соседи



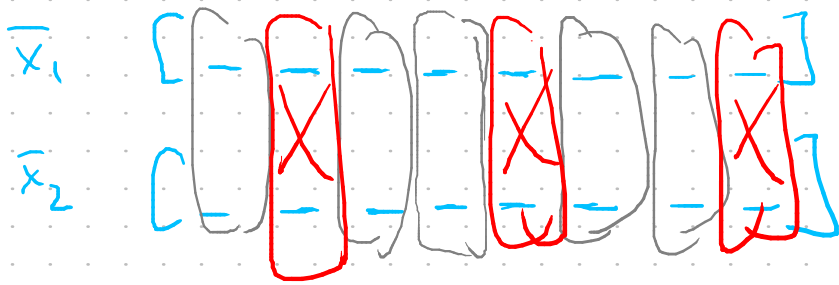
$$N \quad \epsilon = 1/N$$

$$O(\sqrt{\epsilon}) \\ O(\sqrt{1/N})$$



$$O(\epsilon^{1/d}) = O(N^{-1/d}) \\ \mathcal{N}(\bar{\theta}, I)$$

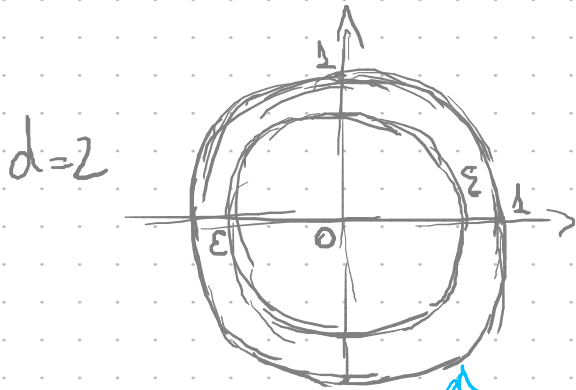
$$w(\bar{\theta}, \begin{pmatrix} 10 \\ 0 \end{pmatrix}) \\ \sqrt{d^2 = \sum x_i^2}$$



— эффект коллапса аненосиона

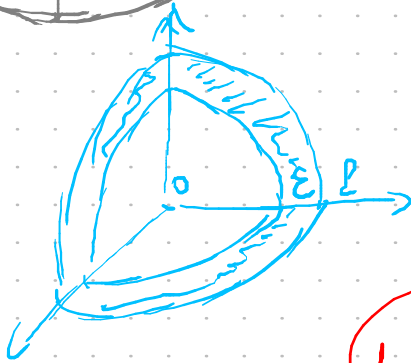


ϵ



$$\frac{\pi r^2 - \pi r^2 (1-\epsilon)^2}{\pi r^2} = 1 - (1-\epsilon)^2 = 2\epsilon - \epsilon^2$$

$d=3$



~~$\frac{4}{3}\pi r^3$~~

$$1 - (1-\epsilon)^3 = 3\epsilon - 3\epsilon^2 + \epsilon^3$$

$1 - (1-\epsilon)^d \rightarrow 1$

③ Statistical decision theory

$\bar{x} \in \mathbb{R}^d, y \in \mathbb{R}$

$D = \{(\bar{x}_n, y_n)\}_{n=1}^N$

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

expected
pred. error
↓

$L(y, \hat{y}) = (y - \hat{y})^2$

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

$EPE[f] = \mathbb{E}_{p(\bar{x}, y)} [(y - f(\bar{x}))^2] =$

$= \int (y - f(\bar{x}))^2 p(\bar{x}) p(y|\bar{x}) dy d\bar{x} =$

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

$\downarrow f(\bar{x})$
min

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

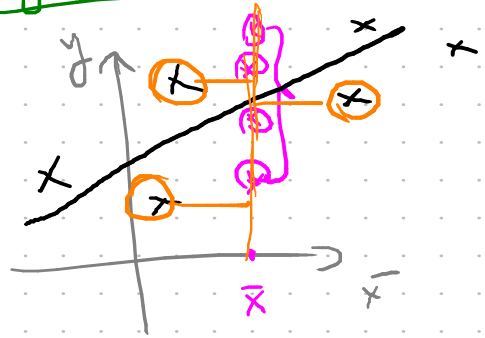
regression function

$$\hat{f}(\bar{x}) = E_{p(y|\bar{x})}[y] \approx$$

$$\approx \frac{1}{R} \sum_r y_r \approx \frac{1}{k} \sum_n y_n$$

$$y_r \sim p(y|\bar{x})$$

$$x_{optimal} \quad n: \bar{x}_n \in kNN(\bar{x})$$



$$y \in \{1, \dots, K\} \quad D = \{(\bar{x}_n, y_n)\} \quad f: \bar{x} \mapsto y$$

$$L(y, f(\bar{x})) = \begin{cases} 0 & f(\bar{x}) = y \\ 1 & \text{---} \end{cases}$$

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

$$= \int \left(\sum_{k=1}^K L(k, f(\bar{x})) p(k|\bar{x}) \right) p(\bar{x}) d\bar{x} \xrightarrow{f} \min$$

	Test	
L	0	1
0	0	1
1	1000	0

$\downarrow f(\bar{x})$
min

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

optimal Bayes classifier

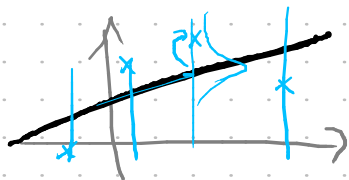
$$\hat{f}(\bar{x}) = \operatorname{argmin}_{f(\bar{x})} \sum_y L(y, f(\bar{x})) p(y|\bar{x})$$

④ Bias - variance - noise decomposition

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

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

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



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

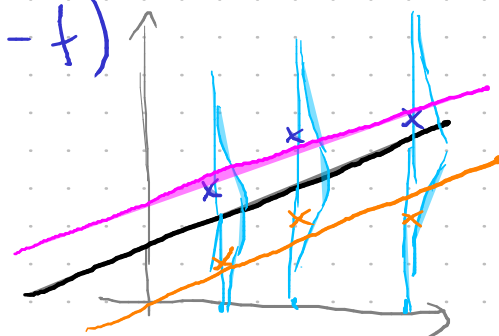
$$EPE[f] =$$

$$= \underbrace{\int (y - \hat{f}(\bar{x}))^2 p(\bar{x}, y) dy d\bar{x}}_{\text{noise}} + \underbrace{\int (\hat{f}(\bar{x}) - f(\bar{x}))^2 p(\bar{x}, y) dy d\bar{x}}_{(\hat{f} - E_D f - f)}$$

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

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

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



$$\begin{aligned} \int (\hat{f} - f) p dy d\bar{x} &= \int (\hat{f}(\bar{x}) - E_D f(\bar{x}; D))^2 p dy d\bar{x} + \\ &+ 2 \int \underbrace{(\hat{f} - E_D f)}_0 \underbrace{(E_D f - f)}_0 p dy d\bar{x} \\ &+ \int (E_D f - f)^2 p dy d\bar{x} \end{aligned}$$

$$\begin{aligned}
 EPE[f] &= \int (\hat{f}(\bar{x}) - E_D f(\bar{x}; D))^2 p(\bar{x}, y) dy d\bar{x} && \text{bias} \\
 &+ \int (\underline{f(\bar{x}; D)} - \underline{E_D f(\bar{x}; D)})^2 p(\bar{x}, y) dy d\bar{x} && \text{variance} \\
 &+ \int (y - \hat{f}(\bar{x}))^2 p(\bar{x}, y) dy d\bar{x} && \text{noise}
 \end{aligned}$$

⑤ Байесовский подход для регрессии

$$p(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} (x - \mu)^2} \quad D = \{x_1, x_2, \dots, x_n\}$$

$\tau = \frac{1}{\sigma^2}$
 precision

$\underline{\mu, \sigma^2}$

$$p(x | \mu, \tau) = \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau}{2} (x - \mu)^2}$$

$$\left. \begin{aligned}
 &p(D | \mu, \tau) \\
 &p(\mu, \tau) \\
 &p(\mu, \tau | D) \\
 &p(x | D)
 \end{aligned} \right\}$$

$$\begin{aligned}
 \log p(x_1, \dots, x_n | \mu, \tau) &= \sum_n \log p(x_n | \mu, \tau) \\
 &= \underline{\frac{N}{2} \log \tau} - \underline{\frac{N}{2} \log(2\pi)} - \frac{\tau}{2} \sum_n (x_n - \mu)^2
 \end{aligned}$$

1) $\tau = \text{const}$, $\mu = ?$

$$\log p(D | \mu) = \text{const} - \frac{\tau}{2} \sum_n (x_n - \mu)^2$$

$$p(\mu) = \mathcal{N}(\mu | \mu_0, \tau_0)$$

$$\log p(\mu | D) = \underbrace{\text{const} - \frac{\tau}{2} \sum_n (x_n - \mu)^2}_{\log p(D | \mu)} - \underbrace{\frac{\tau_0}{2} (\mu - \mu_0)^2}_{\log p(\mu)} =$$

$$= \text{const} - \frac{\tau}{2} \left(N \cdot \underline{\mu^2} - 2\mu \cdot \sum_n x_n + \sum_n x_n^2 \right) - \frac{\tau_0}{2} \left(\underline{\mu^2} - 2\mu_0 \mu + \underline{\mu_0^2} \right)$$

$$= \text{const} - \frac{1}{2} \cdot \mu^2 (\tau N + \tau_0) - \frac{1}{2} \cdot 2\mu \left(\tau \cdot \sum_n x_n + \mu_0 \right)$$

$$= \text{const} - \frac{\tau_0 + N\tau}{2} \left(\mu - \frac{\mu_0 + \tau \sum_n x_n}{\tau_0 + N\tau} \right)^2$$

$$\left\{ \begin{array}{l} \tau_N = \tau_0 + N\tau \\ \mu_N = \frac{\mu_0 + \tau \sum_n x_n}{\tau_0 + N\tau} \end{array} \right.$$

2) $\mu = \text{const}$, $\tau = ?$, $\tau = 1/\sigma^2$, $\tau > 0$, const

$$\log p(D|\tau) = \text{const} + \frac{N}{2} \log \tau - \frac{\tau}{2} \sum_n (x_n - \mu)^2$$

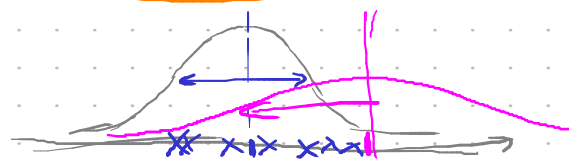
$$p(\tau | \alpha_0, \beta_0) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \tau^{\alpha_0-1} e^{-\beta_0 \tau} \quad \text{Gamma distribution}$$

$$\begin{aligned} \log p(\tau | D) &= \text{const} + \frac{N}{2} \log \tau - \frac{\tau}{2} \sum_n (x_n - \mu)^2 + (\alpha_0 - 1) \log \tau - \beta_0 \tau \\ &= \text{const} + \left(\alpha_0 + \frac{N}{2} - 1 \right) \log \tau - \tau \left(\beta_0 + \frac{1}{2} \sum_n (x_n - \mu)^2 \right) \end{aligned}$$

$$p(\tau | D) = \text{Gam} \left(\tau \mid \alpha_0 + \frac{N}{2}, \beta_0 + \frac{1}{2} \sum_n (x_n - \mu)^2 \right)$$

3) $\underline{\mu, \tau = ?}$

$p(\mu, \tau) = ?$



$$p(\mu, \tau) \neq p(\mu) p(\tau) = \mathcal{N}(\mu | \mu_0, \tau_0) \cdot \text{Gam}(\tau | \alpha_0, \beta_0)$$

$$\log p(\mu, \tau) = \log p(\mu) + \log p(\tau)$$

$$\log p(\underline{D}|\underline{\mu}, \tau) = \text{Const} + \frac{N}{2} \log \tau - \frac{\tau}{2} \sum_n (x_n - \mu)^2$$

$$p(\underline{\mu}, \tau) = \underline{p(\tau) p(\underline{\mu}|\tau)} = p(\tau) p(\underline{\mu}|\tau)$$

$$\left\{ \begin{array}{l} p(\tau) = \text{Gam}(\tau | \alpha_0, \beta_0) \\ p(\underline{\mu}|\tau) = \mathcal{N}(\underline{\mu} | \underline{\mu}_0, \lambda_0 \tau) \end{array} \right. \quad \log p(\tau) = \text{Const} + (\alpha_0 - 1) \log \tau - \beta_0 \tau$$

$$\log p(\underline{\mu}, \tau | \underline{D}) = \text{Const} + \frac{N}{2} \log \tau - \frac{\tau}{2} \sum_n (x_n - \mu)^2 + (\alpha_0 - 1) \log \tau - \beta_0 \tau + \frac{1}{2} \log \tau - \frac{\lambda_0 \tau}{2} (\mu - \mu_0)^2$$

$\log p(\underline{D}|\underline{\mu}, \tau)$
 $\log p(\tau)$
 $\log p(\underline{\mu}|\tau)$

$$- \frac{1}{2} \left(\tau (N \mu^2 - 2 \mu \sum x_n + \sum x_n^2) + \lambda_0 \tau (\mu^2 - 2 \mu \mu_0 + \mu_0^2) \right)$$

$$= - \frac{\tau}{2} \left(\mu^2 (N + \lambda_0) - 2 \mu (\sum x_n + \lambda_0 \mu_0) + \sum x_n^2 + \lambda_0 \mu_0^2 \right)$$

$$= - \frac{\tau (N + \lambda_0)}{2} \left(\mu - \frac{\lambda_0 \mu_0 + \sum x_n}{\lambda_0 + N} \right)^2 - \frac{\tau}{2} \left(\sum x_n^2 + \lambda_0 \mu_0^2 - \frac{(\lambda_0 \mu_0 + \sum x_n)^2}{\lambda_0 + N} \right)$$

$$p(\underline{\mu}, \tau | \underline{D}) = \text{Gam}(\tau | \alpha_0 + \frac{N}{2}, \beta_0 + \frac{1}{2} \sum x_n^2 + \lambda_0 \mu_0^2 - \frac{(\lambda_0 \mu_0 + \sum x_n)^2}{\lambda_0 + N}) \times \mathcal{N}(\underline{\mu} | \frac{\lambda_0 \mu_0 + \sum x_n}{\lambda_0 + N}, \frac{(N + \lambda_0) \tau}{2})$$

$$\sqrt{\frac{\tau}{2\pi}} \cdot e^{-\frac{\tau}{2} \left(\mu - \frac{\lambda_0 \mu_0 + \sum x_n}{\lambda_0 + N} \right)^2}$$

⑥ Posterior for μ ?

$$p(\mu|D) = p(\mu|x_1, x_2, \dots, x_n) = \int_0^\infty p(\mu, \tau|D) d\tau =$$

$$= \int_0^\infty \text{Gam}(\tau|\alpha_N, \beta_N) \mathcal{N}(\mu|\mu_N, \lambda_N \tau) d\tau =$$

$$= \text{Const} \int_0^\infty \tau^{\alpha_N-1} e^{-\beta_N \tau} \cdot \sqrt{\tau} \cdot e^{-\frac{\lambda_N \tau}{2} (\mu - \mu_N)^2} d\tau =$$

$$= \text{Const} \cdot \int_0^\infty \tau^{\alpha_N + \frac{1}{2} - 1} e^{-\tau \left(\beta_N + \frac{\lambda_N}{2} (\mu - \mu_N)^2 \right)} d\tau =$$

$$= \text{Const} \cdot \frac{\Gamma(\alpha_N + \frac{1}{2})}{\left(\beta_N + \frac{\lambda_N}{2} (\mu - \mu_N)^2 \right)^{\alpha_N + \frac{1}{2}}} = \frac{\text{Const}}{\left(\frac{1 + \frac{\lambda_N}{2\beta_N} (\mu - \mu_N)^2}{\frac{2\alpha_N + 1}{2}} \right)^{\frac{2\alpha_N + 1}{2}}}$$

$$p(t|D) = \frac{\Gamma\left(\frac{D+1}{2}\right)}{\sqrt{\pi D} \Gamma\left(\frac{D}{2}\right)} \cdot \left(\frac{1 + \frac{t^2}{D}}{\frac{D}{2}} \right)^{-\frac{D+1}{2}}$$

Student's
t-distribution

fat tails

⑦ Predictive distribution

$$p(x|x_1, \dots, x_N) = \int_0^\infty \int_{-\infty}^\infty p(x|\mu, \tau) p(\mu, \tau|x_1, \dots, x_N) d\mu d\tau =$$

$$= \int_0^\infty \int_{-\infty}^\infty \mathcal{N}(x|\mu, \tau) \cdot \text{Gam}(\tau|\alpha_N, \beta_N) \cdot \mathcal{N}(\mu|\mu_N, \lambda_N \tau) d\mu d\tau$$

$$= \int_0^\infty \left(\int_{-\infty}^\infty \mathcal{N}(x|\mu, \tau) \cdot \mathcal{N}(\mu|\mu_N, \lambda_N \tau) d\mu \right) \text{Gam}(\tau|\alpha_N, \beta_N) d\tau$$

$\mathcal{N}(\text{---})$

$$= \int_0^{\infty} \mathcal{N}(x | \mu_N, \frac{\lambda_N}{\lambda_N + 1} \tau) \text{Gam}(\tau | \alpha_N, \beta_N) d\tau =$$

Const

$$= \frac{1}{\left(1 + \frac{\lambda_N}{\beta_N(\lambda_N + 1)} (x - \mu_N)^2\right)^{\frac{2\alpha_N + 1}{2}}}$$