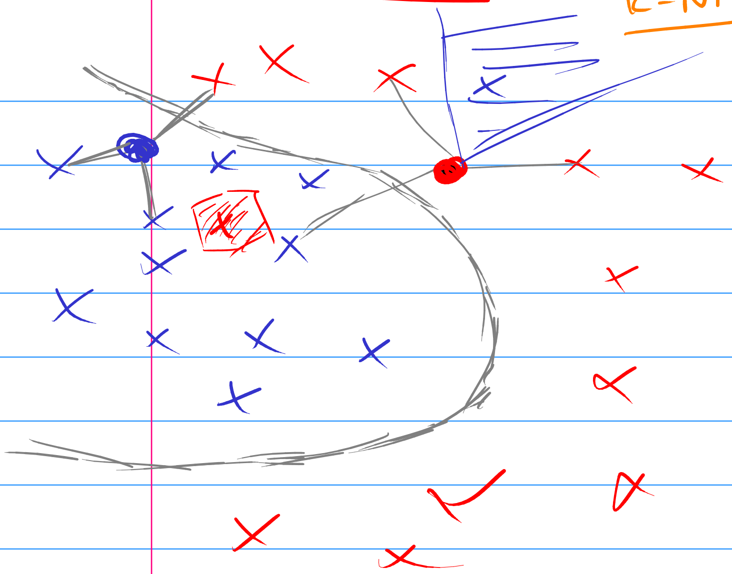
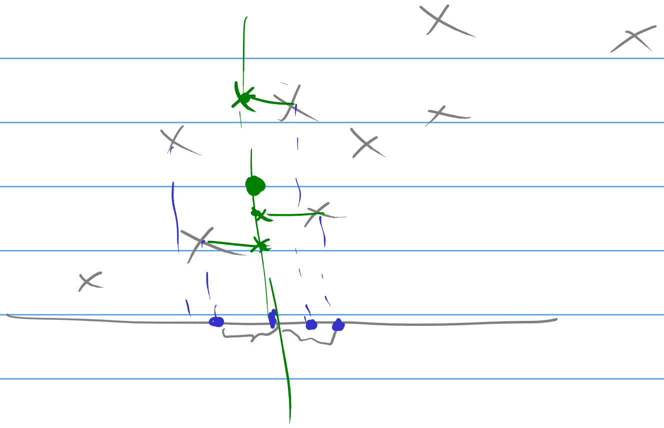
Nearest neighbors

k-NN



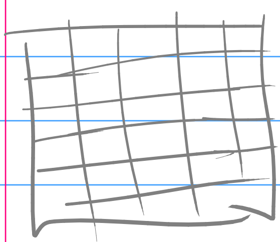


Curse of dimensionality

1) 

$1/\varepsilon$

$\int \dots \approx \sum$

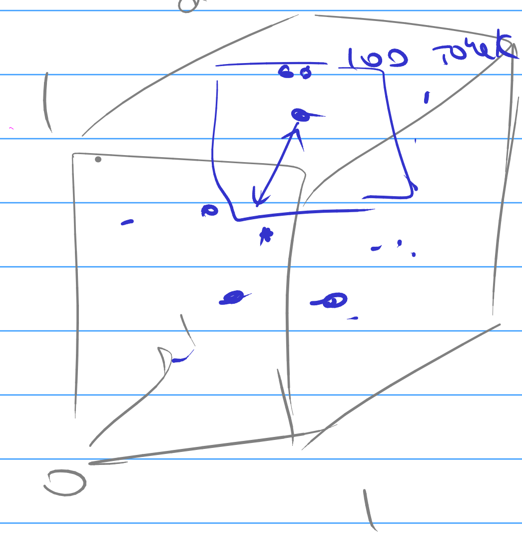


$1/\varepsilon^2$

d

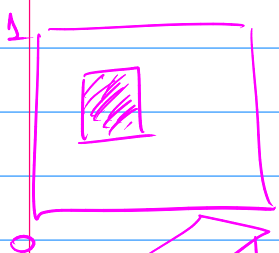
$1/\varepsilon^d$

$d=10$

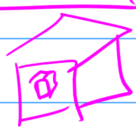


2) 

ε



$\sqrt{\varepsilon}$

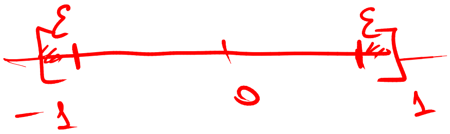


$\sqrt[3]{\varepsilon}$

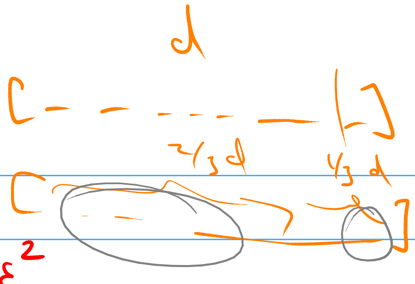
d

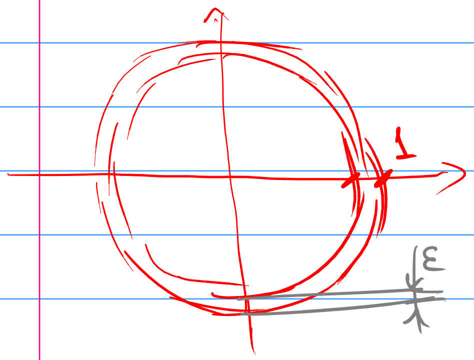
$\sqrt[d]{\varepsilon}$

3)

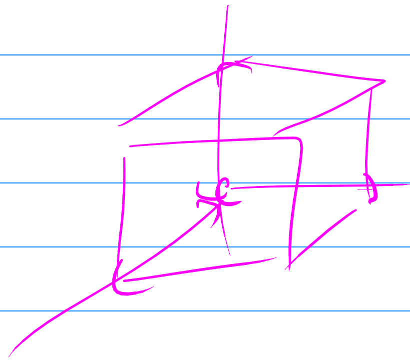


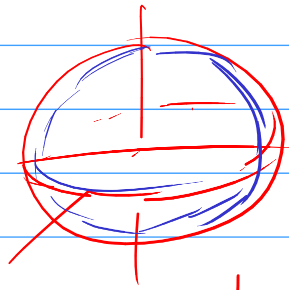
ε





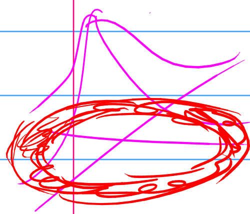
$$1-(1-\varepsilon)^2 = 2\varepsilon-\varepsilon^2$$



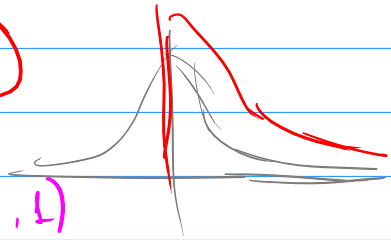


$$1-(1-\varepsilon)^3$$

d — $1-(1-\varepsilon)^d \xrightarrow[d\to\infty]{} 1$



$\bar{x} \sim \mathcal{N}(\bar{x} \mid \bar{0}, I)$ y_{np}



$(x_1, x_2, \ldots, x_d)$ $x_i \sim \mathcal{N}(x_i \mid 0, 1)$

$r^2(\bar{x}) = \boxed{\sum x_i^2} \longrightarrow \mathcal{N}(\_)$ $\boxed{x_i^2 \sim}$

Statistical decision theory

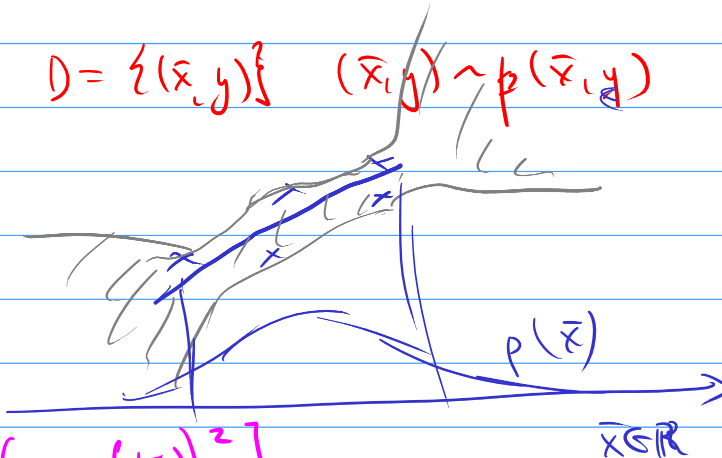
$\bar{x} \in \mathbb{R}^d$, $y \in \mathbb{R}$, $\boxed{p(\bar{x}, y)}$ $D = \{(\bar{x}, y)\}$ $(\bar{x}, y) \sim p(\bar{x}, y)$

$p(y \mid \bar{x})$

$\mathbb{R}^d \to \mathbb{R}$

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

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



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

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

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

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

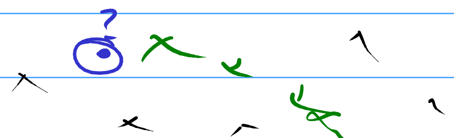
$$\xrightarrow[f(\bar x)]{} \min$$

$$\int (x-a)^2 \, p(x) \, dx \xrightarrow[a]{} \min \qquad \boxed{\hat a = \mathbb{E} x}$$

$$\mathbb{E}[(x-a)^2] =$$

$$= a^2 - 2a \, \mathbb{E}x + \mathbb{E}(x^2)$$

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



$$y \in \{c_1, c_2, \dots, c_K\}$$

$$L(y, f(\bar x)) = [y \neq f(x)] = \begin{cases} 0, & y = f(\bar x) \\ 1, & y \neq f(\bar x) \end{cases}$$

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

$$= \iint L(y, f(\bar x)) \, p(\bar x, y) \, dy \, dx =$$

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

$$= \int \left(\sum_{k=1}^{K} [f(\bar x) \neq c_k] \, p(c_k|\bar x) \right) p(\bar x) \, d\bar x$$

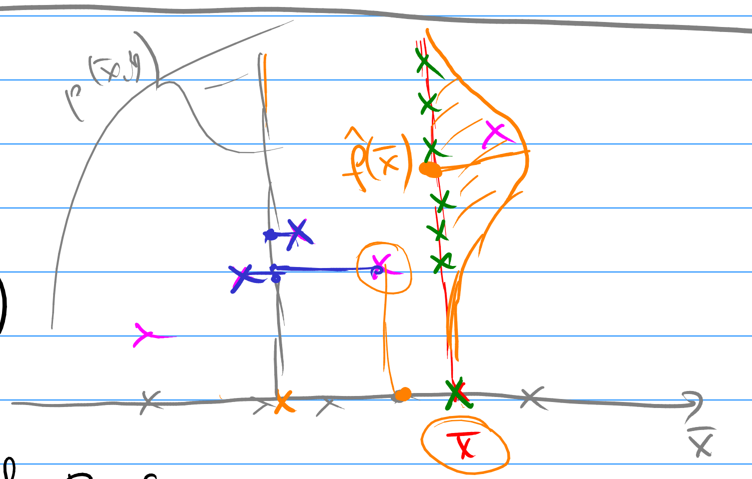
$$\xrightarrow[f(\bar x)]{} \min$$

$$\boxed{\hat f(\bar x) = \arg\min_k p(C_k|\bar x)}$$

оптимальный байесовский классификатор

$$\boxed{\hat f(\bar x) = \arg\min_k \sum_{k=1}^{K} L(C_k, f(\bar x))\, p(C_k|\bar x)}$$

		Test pos	Test neg
Dis	pos	0	1000
	neg	1	0



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

$$\approx \frac{1}{R}\sum_{r=1}^{R} y_r, \quad \text{где} \quad y_r \sim p(y|\bar x)$$

$$\approx \frac{1}{R}\sum_{r=1}^{R} y_r, \quad \text{где } y_r \text{ взяты в } R \text{ ближ.-соседях } \bar x$$

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

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

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

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

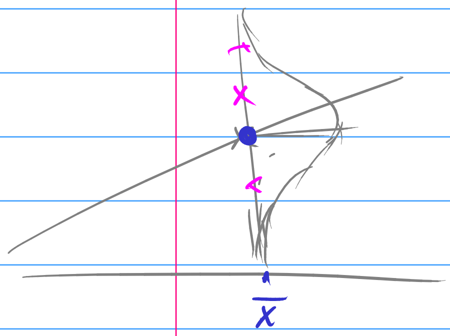
$$\Vert$$

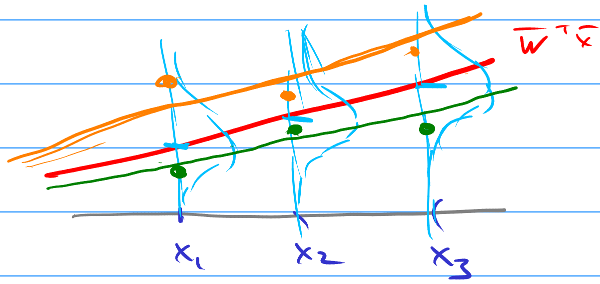
$$2\iint (y - \underbrace{\hat f(\bar x)}_{=\mathbb{E}[y|\bar x]})\, p(y|\bar x) \cdot (\hat f(\bar x) - f(\bar x))\, p(\bar x)\, d\bar x\, dy$$

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

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

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





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

$$E\left[(\hat f(\bar x) - f(\bar x; D))^2\right] = E\left[\left(\left(\hat f(\bar x) - E_{D'}[f(\bar x; D')]\right) + \left(E_{D'}[f(\bar x, D')] - f(\bar x; D)\right)\right)^2\right] =$$

$$= E\left[(\hat f - E_D f)^2\right] + \cancel{2E\left[(\hat f - E_D f)(E_D f - f(\bar x))\right]} + E\left[(E_D f - f)^2\right]$$

$$\iint \left(f(\bar x, D) - E_D f(x; D')\right) p(\bar x, y)\, d\bar x\, dy$$

$$EPE[f] = E\left[(\hat f - E_D f)^2\right] + \qquad \text{Bias}$$

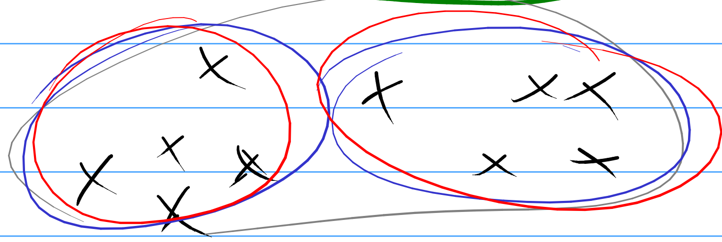
$$+ E\left[(f - E_D f)^2\right] + \qquad \text{variance}$$

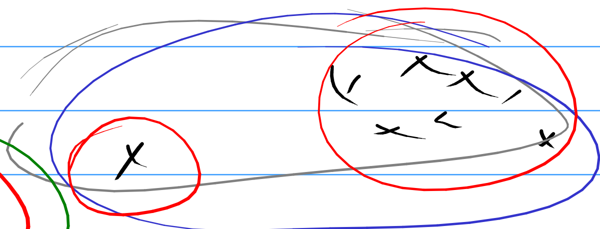
$$+ E\left[(y - \hat f)^2\right] \qquad \text{noise}$$

Bayesian model selection

$M_1, M_2, \ldots M_k$ — модели

θ, $p(D|\theta)\,p(\theta)$



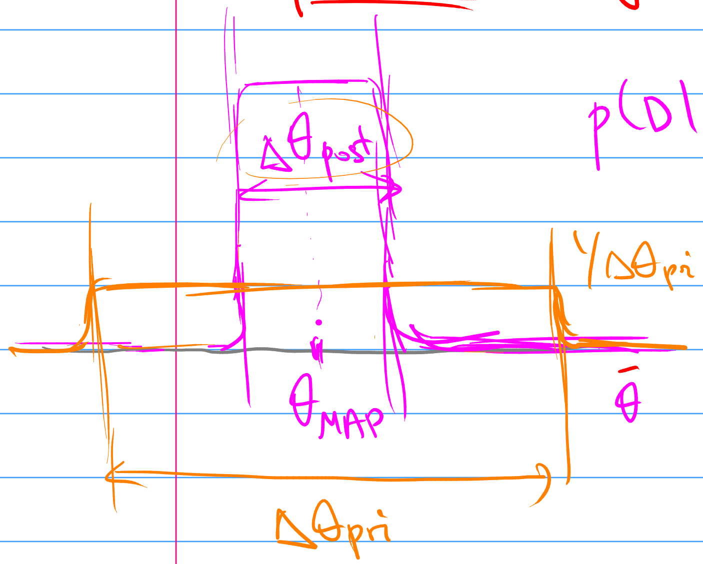


D

$$p(M_k | D) \propto p(M_k)\, p(D | M_k)$$

$$p(\theta|D,\mathcal{M}_k) = \frac{p(D|\theta,\mathcal{M}_k)\,p(\theta|\mathcal{M}_k)}{p(D|\mathcal{M}_k)}$$ evidence

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



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

$$= \underbrace{\int_{\Delta\theta_{post}} p(D|\theta_{MAP},\mathcal{M}_k)\,p(\theta|\mathcal{M}_k)\,d\theta}$$

$$= p(D|\theta_{MAP},\mathcal{M}_k)\,\frac{1}{\Delta\theta_{pri}}\underbrace{\int 1}_{\Delta\theta_{post}}$$

$$\ln p(D) \approx \ln p(D|\theta_{MAP}) - \ln\left(\frac{\Delta\theta_{pri}}{\Delta\theta_{post}}\right)$$

$$\ln p(D) \approx \ln p(D|\theta_{MAP}) - d \cdot \underbrace{\ln\left(\frac{\Delta\theta_{pri}}{\Delta\theta_{post}}\right)}$$