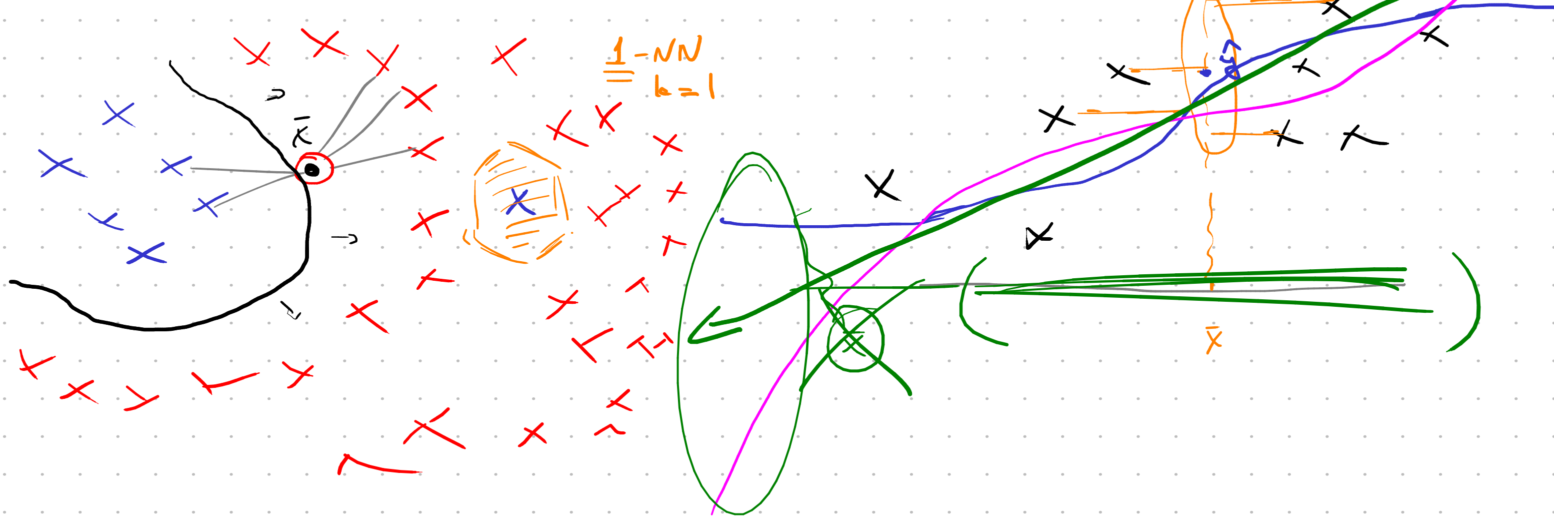
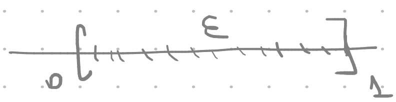
① Nearest neighbors   k-NN



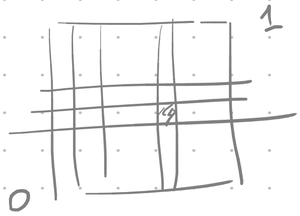
② Curse of dimensionality

— численное интегрирование

$d=1$



$d=2$



$1/\varepsilon^2$

$1/\varepsilon^d$

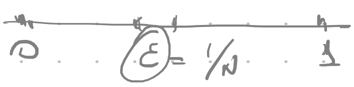
MCMC

$$p(\bar{x}\mid D) = \underline{\mathbb{E}_{p(\bar\theta\mid D)}}\left[p(\bar{x}\mid\bar\theta)\right] \approx \frac{1}{R}\sum p(\bar{x}\mid\bar\theta^{(r)})$$

$$\bar\theta^{(r)} \sim p(\bar\theta\mid D)$$

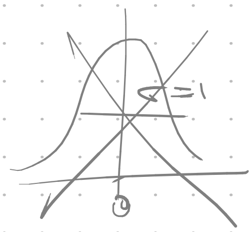
— ближайшие соседи

$d=1$

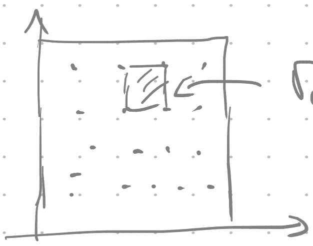


N точек

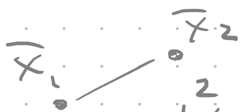
$\sqrt[d]{\varepsilon}$



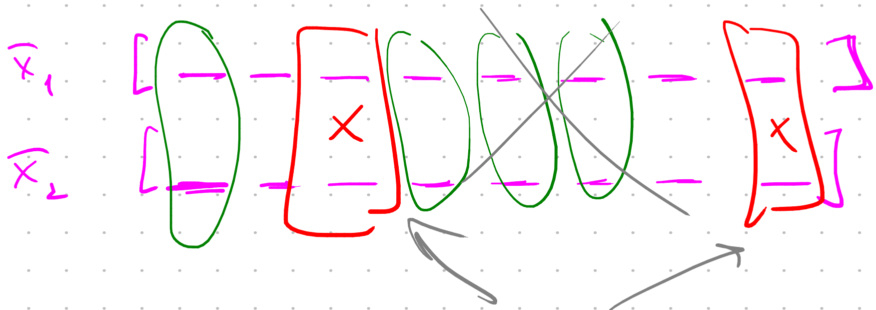
$d=2$

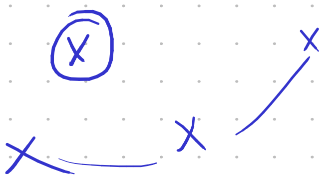


$\sqrt{\varepsilon}$



$$d^2(\bar{x}_1, \bar{x}_2) = \sum_{i=1}^{d} \underbrace{(x_{1i} - x_{2i})^2}$$

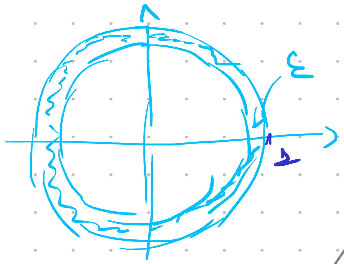




→ Эффект кожуры апельсина

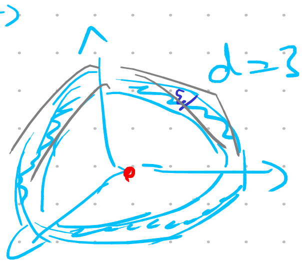
$d=1$

ε



πe^2

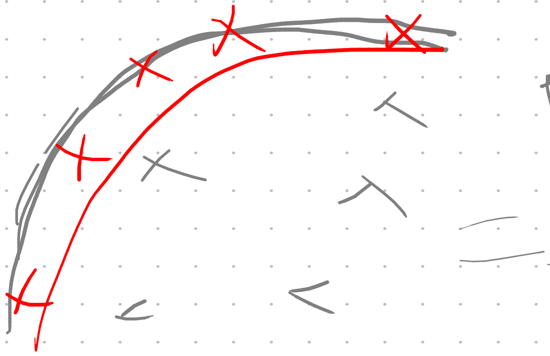
$$\frac{\pi\cdot 1-\pi(1-\varepsilon)^2}{\pi\cdot 1}=\underline{1-(1-\varepsilon)^2}=\underline{2\varepsilon-\varepsilon^2}$$

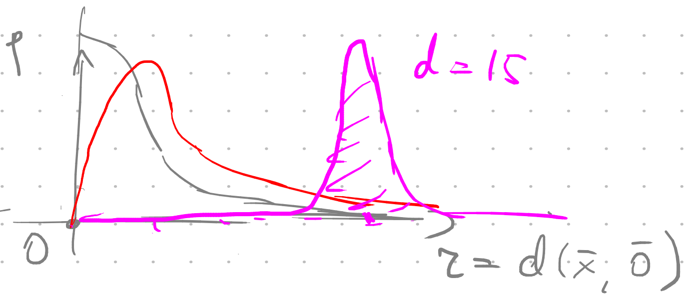


$d=3$

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

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





$$r^2=\sum^{d} x_i^2 \sim \mathcal{N}(x_i|0,1)$$

3 Statistical decision theory

$\bar{x}\in\mathbb{R}^d$, $y\in\mathbb{R}$ $\quad D=\{(\bar{x}_n,y_n)\}_{n=1}^{N}$ $\quad \underline{f}:\bar{x}\mapsto y$

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

exp pred_error

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

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

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

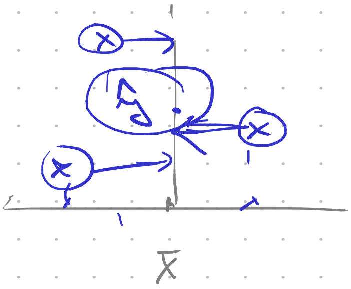
$$=\int\left[\int (y-f(\bar{x}))^2\,p(y|\bar{x})\,dy\right]p(\bar{x})\,d\bar{x}\xrightarrow[f]{}\min$$

$f(\bar{x})$ ↘ min

regression function

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

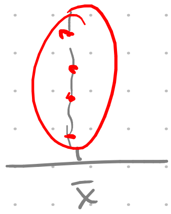
$$\int (x-\underset{=\mathbb{E}[x]}{a})^2\,p(x)\,dx \quad \mathbb{E}_{p(x)}\left[(x-a)^2\right]$$



$$\hat f(\bar x) = E_{p(y|\bar x)}[y] \approx \frac{1}{n}\sum_{i=1}^{n} y_i \approx, \quad y_i \sim p(y|\bar x)$$

хорошая ? хорошая

$$\approx \frac{1}{n}\sum_{i\in NN} y_i$$



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

$$EPE[f] = E_{p(\bar x, y)}\left[L(y, f(\bar x))\right]$$

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

$f(\bar x)$

1 1 ... 0 ... 1

$$= \int \left(\underbrace{\sum_{k=1}^{K} L(C_k, f(\bar x))\, p(C_k|\bar x)}_{\to \min_{f(\bar x)}} \right) p(\bar x)\, d\bar x$$

$$\boxed{\hat f(\bar x) = \arg\max_k p(C_k|\bar x)}$$ – optimal Bayes classifier

$$\hat f(\bar x) = \arg\max_k \sum_{s=1}^{K} L(C_s, C_k)\, p(C_s|\bar x)$$

Disease	Test: $t=0$	Test: $t=1$
$d=0$	0	1
$d=1$	1000	0

4 Bias-variance-noise decomposition

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

$$\pm \hat f(\bar x)$$

$$\left((y - \hat f) \pm (\hat f - f)\right)^2$$

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

$$\iint (y-\hat f(\bar x))(\hat f(\bar x) - f(\bar x))\, p(\bar x)\, p(y|\bar x)\, d\bar x\, dy$$

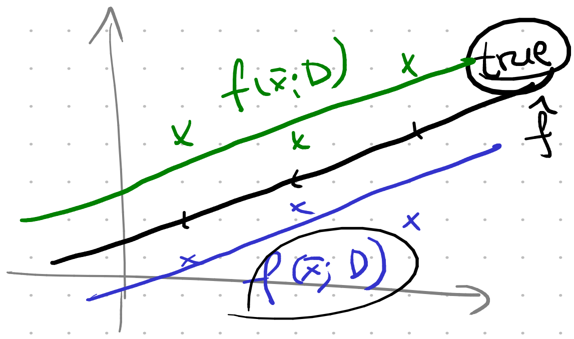
$$= \int \Big(\underbrace{\int (y-\hat f(\bar x))\, p(y|\bar x)\, dy}_{=0} \Big)(\hat f - f)\, p(\bar x)\, d\bar x$$

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

$f \to \min$

$D, \quad f(\bar x; D), \quad D \sim p(\bar x, y)$

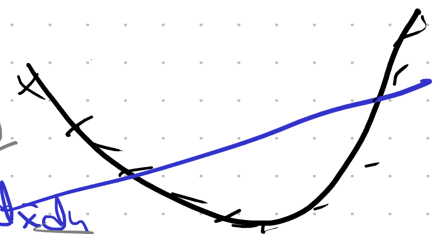
$E_D[f(\bar x)]$



$$E\left[\left(\hat f - f \pm E_D[f]\right)^2\right] =$$

$$= E\left[(\hat f - E_D[f])^2\right] - 2\int (\hat f - E_D f)(E_D f - f(\bar x; D))\, p\, d\bar x\, dy\, dD$$

$$+ E\left[(f - E_D[f])^2\right]$$



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

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

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