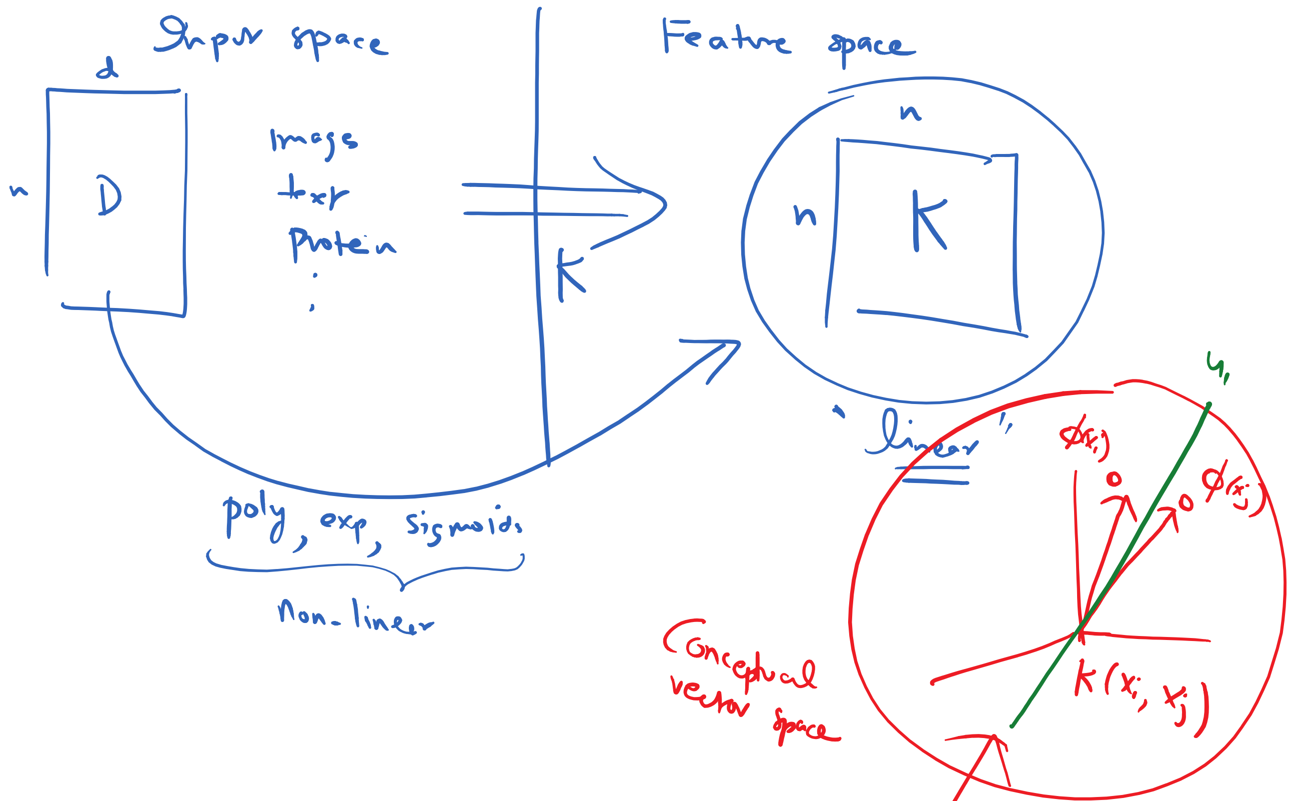


Kernel PCA

use only **K**

K → Symmetric
K → Positive Semi-Definite

we ~~do~~ want to find $\vec{u}_1$: direction in feature space that maximizes variance
minimizes MSE

we can still project onto $\vec{u}_1$

$$\boxed{\vec{u}_1 = \sum_{i=1}^{n} c_i \, \phi(x_i)}$$

$\phi(x_j)$ project onto $\vec{u}_1$:

$$\underbrace{\left(\frac{\phi(x_j)^T u_1}{u_1^T u_1} \right)}_{\text{Value}} u_1 \qquad u_1^T u_1 = 1$$

$$\underbrace{\qquad\qquad}_{\text{Value } a_j}$$

$u_1^T u_1 = 1$

$$a_j = \phi(x_j)^T u_1$$

$$= \phi(x_j)^T \left(\sum_{i=1}^{n} c_i \phi(x_i) \right)$$

$$= \sum c_i \, \phi(x_j)^T \phi(x_i)$$

$$\boxed{a_j = \sum_{i=1}^{n} c_i \, k(x_j, x_i)}$$

we can use only K.

$$\boxed{K\vec{c} = n.\lambda \, \vec{c}}$$

$$\vec{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \in \mathbb{R}^n$$

$$\underline{u_1} = \sum_{i=1}^{n} c_i \phi(x_i)$$

$$K\vec{c}_1 = \eta_1 \vec{c}_1$$

$$\eta_1 = n \cdot \lambda_1$$

⓪ Create K

① Center K

② $(C, \Lambda) = \text{eigen}(K)$

$$\underbrace{C = \begin{pmatrix} | & | & & | \\ \vec{c}_1 & c_2 & \cdots & c_n \\ | & | & & | \end{pmatrix}}_{\text{all eigenvectors}}$$

$$\Lambda = \begin{pmatrix} n\lambda_1 & & & \\ & n\lambda_2 & & \\ & & \ddots & \\ & & & n\lambda_d \end{pmatrix}$$

③ <u>project</u>: $\begin{pmatrix} | & | \\ \vec{c}_1 & c_2 \\ | & | \end{pmatrix}$

new data in 2D : $\phi(x_1) \begin{pmatrix} a_{11}, & a_{12} \\ a_{21}, & a_{22} \\ \vdots & \vdots \end{pmatrix}$

$n \times 2$

a_{11} is projection of $\phi(x_1)$ onto $\vec{c}_1$

a_{12} is projection of $\phi(x_1)$ onto $\vec{c}_2$

A_2

a_{11} is projection of $\phi(x_1)$ onto $\vec{u}_2$

$$\phi(x_1)^T \left(\sum_{i=1}^{n} c_i \phi(x_i) \right)$$

$$a_{11} = \boxed{\sum_{i=1}^{n} c_i \cdot k(x_1, x_i)}$$

$$= \vec{c}_1 \cdot K_1$$

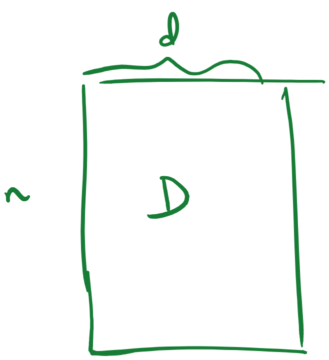
1st row of K matrix

---

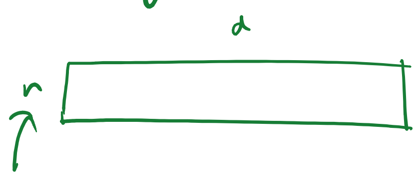
linear kernel

Σ $\qquad$ K

exponential K

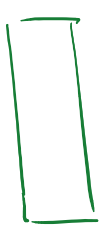


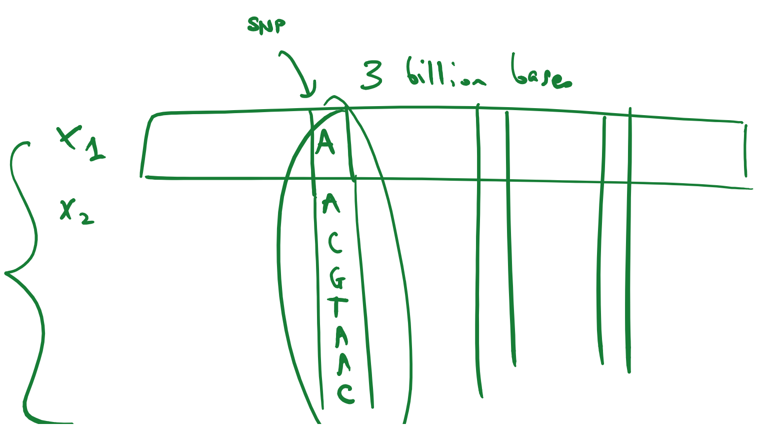
d can very large

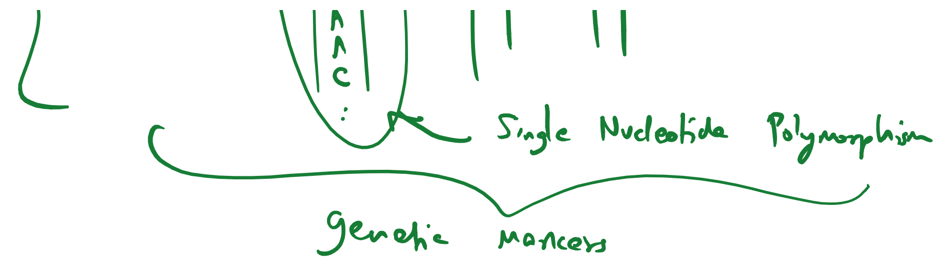


$n = 100$ $\qquad$ $d = 100{,}000$

$n \gg d$



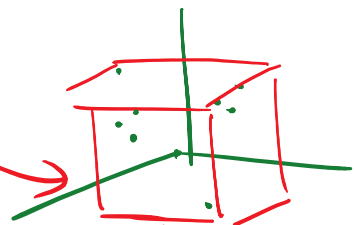




High Dimensional Space

"curse" "benefits"

Universe vs. sample

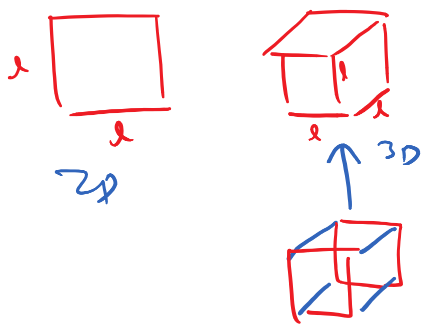


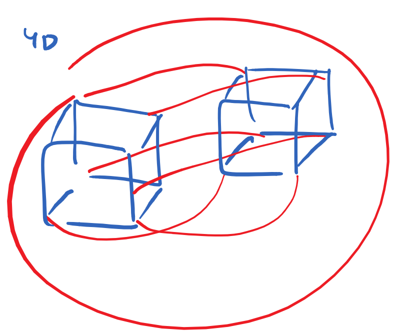
hypercube H_d

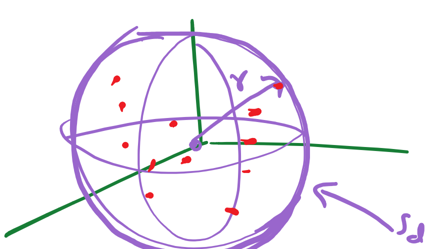
edge len: l

Volume

$$V(H_d) = l^d$$





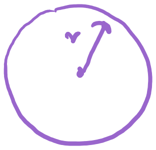


Vol is the same { hyper ball ← inside as well; hyper sphere ← only the surface }

$$B_d = \{ \vec{x} \mid \|x\|_2^2 \le r^2 \}$$

$$S_d = \{ \vec{x} \mid \|x_2\|^2 = r^2 \}$$

the same [hyper sphere $\longleftarrow$ only the surface $S_d = \{ \vec{x} \mid \|x_2\|^2 = r^2 \}$



2D

$$\text{vol}(S_2) = \pi r^2$$

3D

$$\text{vol}(S_3) = \frac{4}{3}\pi r^3$$

$\vdots$

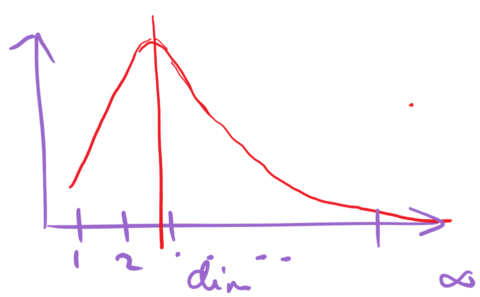
$$\text{vol}(S_d) = K_d \, r^d = \underbrace{\left(\frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)} \right)}_{K_d} r^d$$

$$\underbrace{\Gamma\left(\frac{d}{2}+1\right)}_{\text{gamma function}} = \begin{cases} \left(\frac{d}{2}\right)! & \text{if } d \text{ is even} \\ \sqrt{\pi}\left(\frac{d!!}{2^{(d+1)/2}}\right) & \text{if } d \text{ is odd} \end{cases}$$

$$d!! = d\cdot(d-2)(d-4)\cdots$$

Vol of the unit hypersphere

$r = 1$

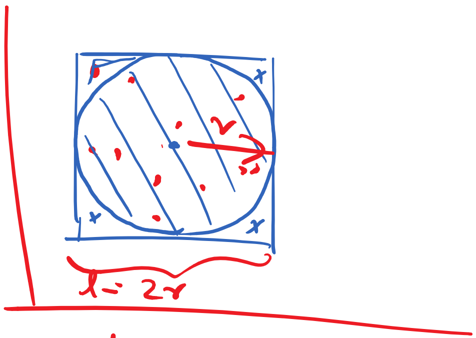


$d = 5.263$

$$\lim_{d\to\infty} \text{vol}(S_d) = \lim_{d\to\infty} \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)} = 0$$

by about 20 dimensions

vol is close to zero.

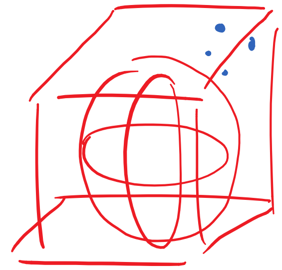


$d = 2$

$4 \quad 2^d$

$$\frac{vol(S_2)}{vol(H_2)} = \frac{\pi r^2}{4r^2} = \frac{\pi}{4}$$

78.5%



$2^3 = 8$ "corners"

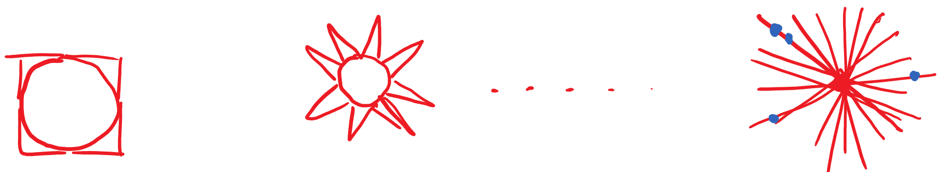
$$\frac{4/3\,\pi r^3}{8r^3} = 52.4\%$$

$$\lim_{d \to \infty} \frac{vol(S_d)}{vol(H_d)} \to 0$$

2^d corners

Points are in the corners

2D

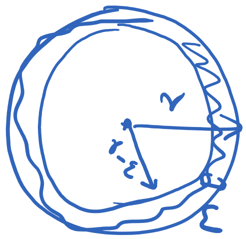


2D 3D ...

"Similarity"

$$n << 2^d$$

Points are
On the surface



ε is very small $\varepsilon > 0$

Vol of the thin shell

$$\frac{\mathrm{vol}(\mathrm{shell}_\varepsilon)}{\mathrm{vol}(\mathrm{outer})} = \frac{\cancel{K_d}\, r^d - \cancel{K_d}(r-\varepsilon)^d}{\cancel{K_d}\, r^d}$$

$$= 1 - \left(\frac{r-\varepsilon}{r}\right)^d$$

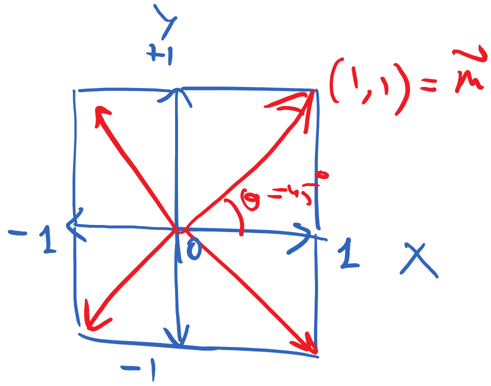
$$\lim_{d \to \infty} \frac{\mathrm{Vol}(\mathrm{shell}_\varepsilon)}{\mathrm{Vol}(\mathrm{outer})} = 1 - \underbrace{\cancel{\left(1 - \frac{\varepsilon}{r}\right)^d}}_{0} = 1$$

$0.9999\ldots 9$

scattering effect

---

explosion of "new" dimensions



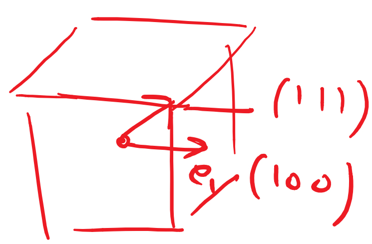
$X = \vec{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $Y = \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

2D

2^d "corners" $\longrightarrow$ 2^{d-1} "axes"

main diagonal : $\begin{pmatrix} 1 \\ 1 \end{pmatrix} =$

$$\boxed{\cos \theta_{\vec{m}, \vec{e}_1} = \frac{\vec{m}^T \vec{e}_1}{\|\vec{m}\| \, \|\vec{e}_1\|}}$$



$$\lim_{d \to \infty} \cos\theta = \frac{1}{\sqrt{d}} = 0$$

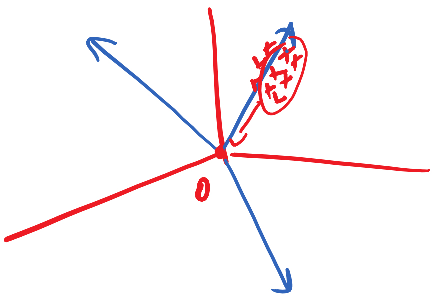
$$\rightarrow \theta = 90^\circ$$

2^{d-1} new "orthogonal" axes

$+$

d

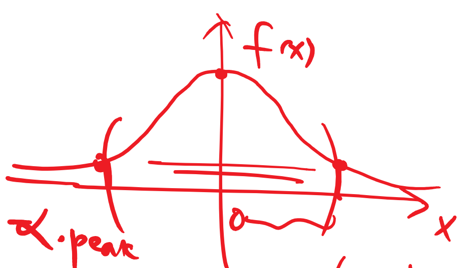
lower-dimensional Orthogonal projections



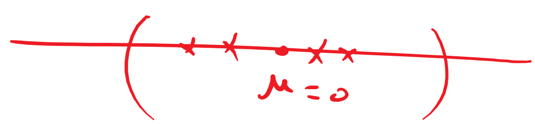
$$\underbrace{2^{d-1} + d}_{\text{total \# of "orthogonal" dimensions}}$$

Density of Normal

1D



$\alpha = 50\%$



$$f(x) = \frac{1}{(\sqrt{2\pi})^d} e^{\left(-\frac{x^T x}{2}\right)}$$

$\mu = 0$

$\sigma = 1$

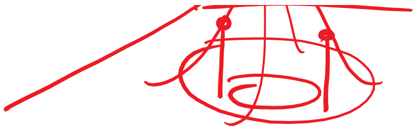
$$\text{Peak density} = f(0) = \frac{1}{(\sqrt{2\pi})^d}$$

2D

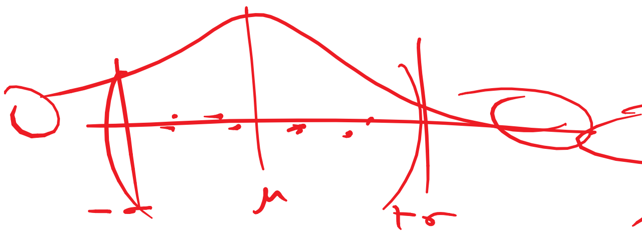
→ 50%.

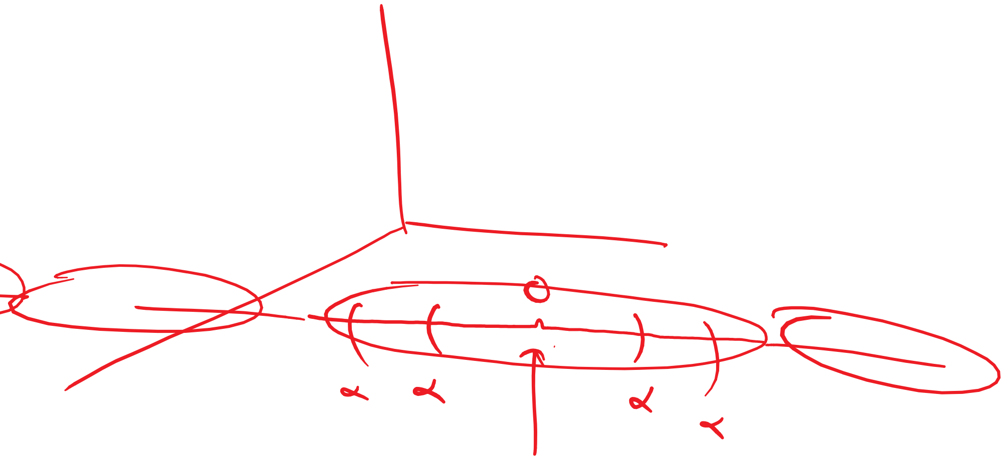
$d \to \infty$

50%.



$d \to \infty$





most of the prob mass
lies in the tails