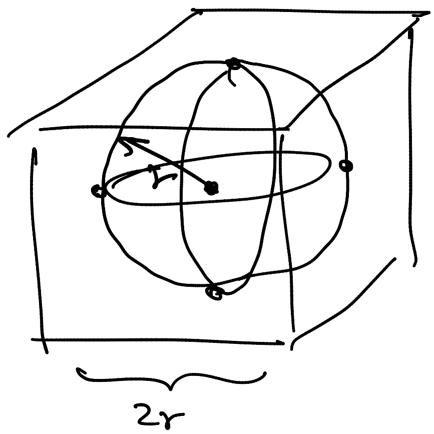
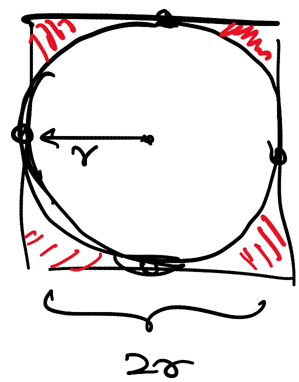
Lecture 8

Thursday, September 21, 2023 9:54 AM

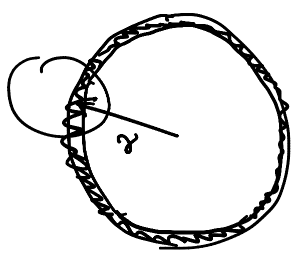




$$\lim_{d \to \infty} \text{ ratio of } \frac{\mathrm{Vol}(S_d(r))}{\mathrm{Vol}(H_d(2r))} = 0$$

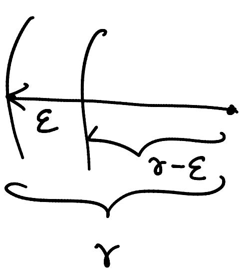
how many corners?

$O(2^d)$



$\varepsilon > 0$

e.g. $\varepsilon = 0.01$



Vol of ε- shell

$$\text{ratio} = \frac{\mathrm{Vol}(\text{outer}) - \mathrm{vol}(\text{inner})}{\mathrm{vol}(\text{outer})}$$

$$= \frac{\mathrm{vol}(S_d(r)) - \mathrm{vol}(S_d(r-\varepsilon))}{\mathrm{vol}(S_d(r))}$$

$$= \frac{\cancel{k_d}\, r^d - \cancel{k_d}\,(r-\varepsilon)^d}{\cancel{k_d}\, r^d}$$

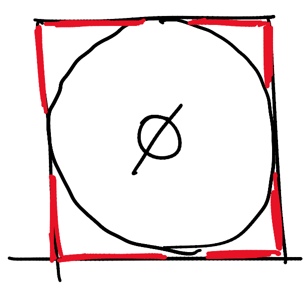
$$= 1 - \left(\frac{r-\varepsilon}{r}\right)^d$$

$$\lim_{d\to\infty} \text{ratio} = 1 - \underbrace{\left(1 - \frac{\varepsilon}{r}\right)}_{<1}^{d} \qquad r > \varepsilon > 0$$

0

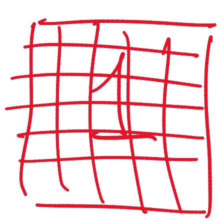
$$= 1$$

all points lie on the boundary (surface)



red = 100% of the volume!

28



28

$\longrightarrow$ 784

lower-dim representation

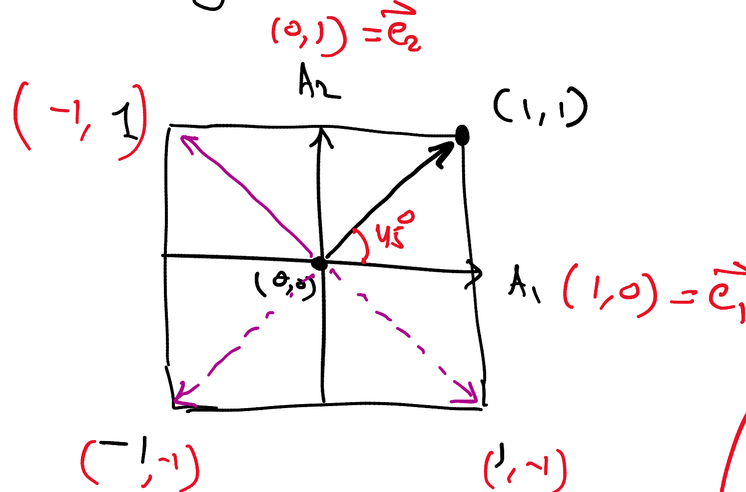
latent space

Latent space

dim << 784

Diagonals

↪ Orthogonal axes



$$\vec{1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

3D

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

$$\vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\cos\theta(\vec{e}_1, \vec{1}) = \frac{\vec{e}_1^T \vec{1}}{\|\vec{e}_1\| \, \|\vec{1}\|}$$

$$2d: \quad \cos\theta = \frac{1}{1 \cdot \sqrt{2}} = \frac{1}{\sqrt{2}} \qquad \begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\theta = 45^\circ$$

in d-dim

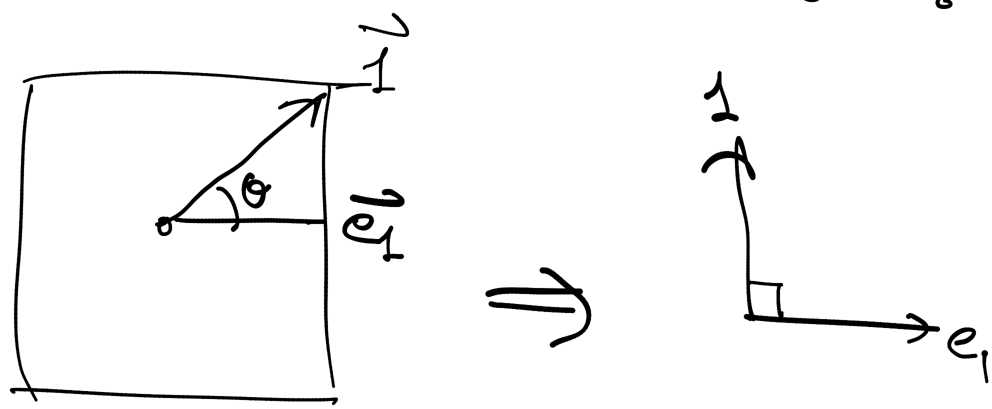
$$\cos\theta(\vec{e}_1, \vec{1}) = \frac{1}{1 \cdot \sqrt{d}}$$

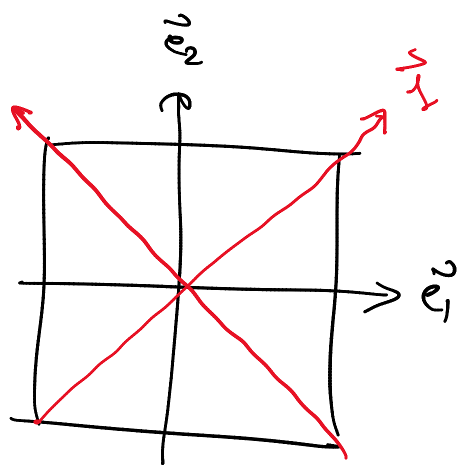
$$\|\vec{1}\| = \sqrt{\vec{1}^T \vec{1}} = \sqrt{d}$$

$$\lim \cos\theta = \lim \frac{1}{\sqrt{d}} = 0$$

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

$\Rightarrow \theta = 90^\circ$ or $\vec{e}_1$ and $\vec{1}$ become orthogonal!



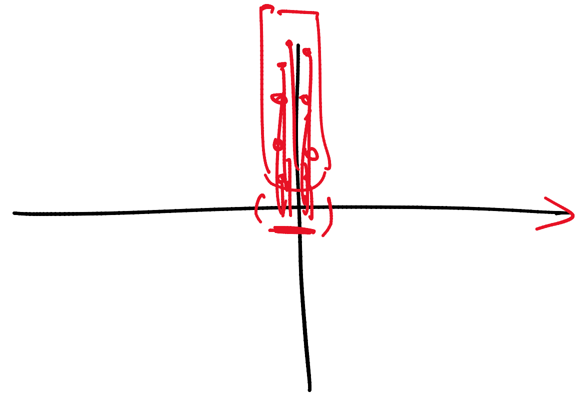


d – dim original dimensions / axes

every "diagonal" axes becomes orthogonal to primary ones

$$O\left(\underbrace{d + 2^{d-1}}\right)$$ additional axes!

big-oh!



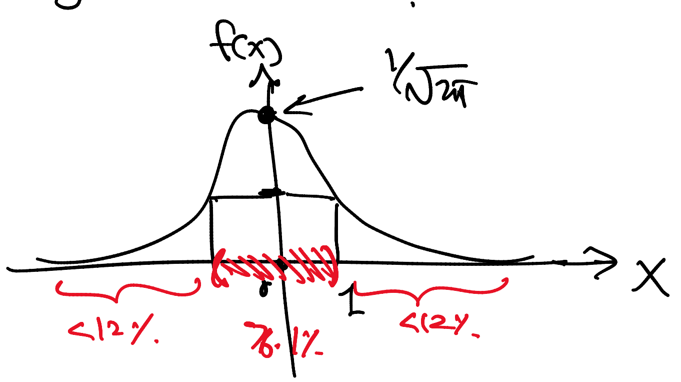
Multivariate Normal

Standard $\rightarrow \vec{\mu} = \vec{0}$

$\searrow \Sigma = I$

$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ in 3d

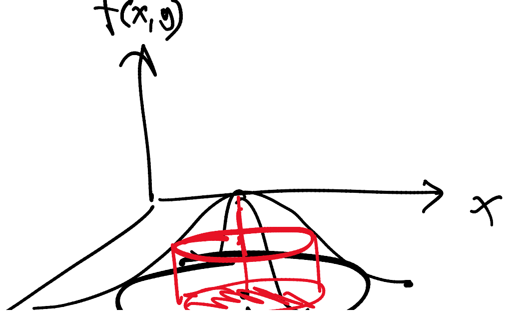
high-dim Gaussian?



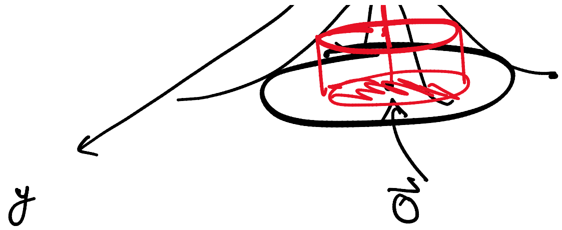
$\frac{1}{\sqrt{2\pi}}$

all x within α fraction of peak density

$\alpha = 50\%$ 76.1% of points lie within the α region



2d



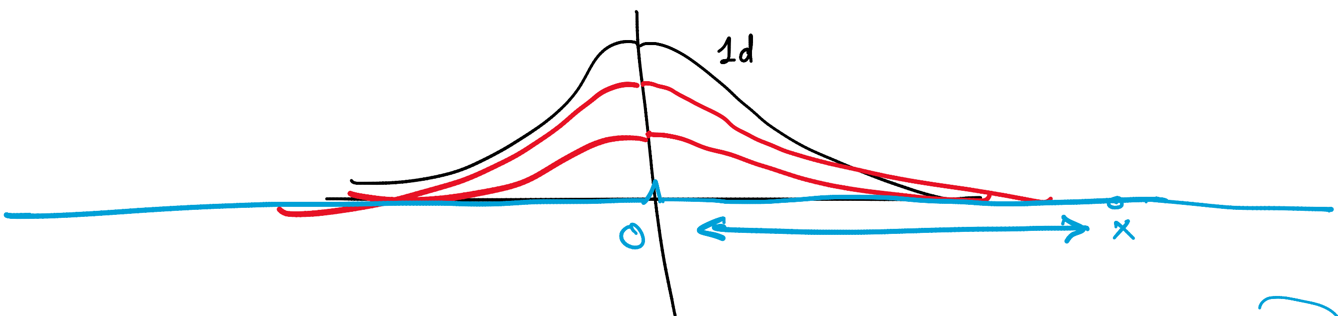
2d 50%

3d → 29.12%

as $d \to \infty$

there are no points with $\propto$ fraction of the peak

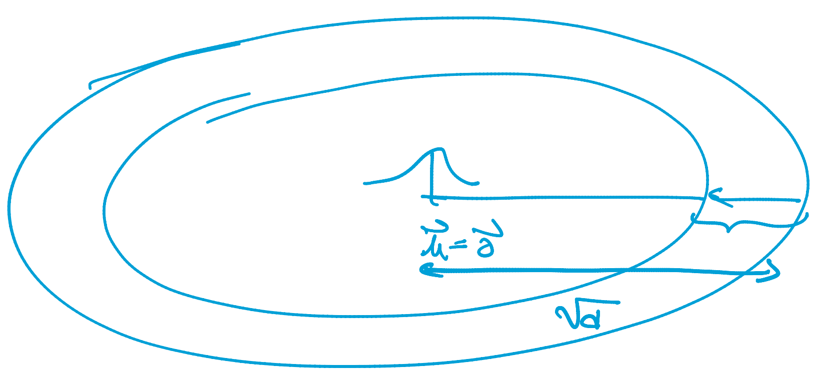
all points migrate to the "tails"



$$f(\vec{x} \mid \vec{0}, I) = \frac{1}{(\sqrt{2\pi})^d} \frac{1}{\sqrt{|I|}} e^{-\frac{\vec{x}^T I^{-1} \vec{x}}{2}} = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{\vec{x}^T \vec{x}}{2}}$$

$$|I| = \det(I) = 1 \qquad \vec{x} \in \mathbb{R}^d$$

$$I^{-1} = I$$



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

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

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

$$P\Big(\underbrace{\frac{f(\vec{x})}{f(\vec{0})}}\geq \alpha\Big)$$

find $\vec{x}$ with α fraction of peak (or mean)

$$P\left(e^{-\frac{\vec{x}^T\vec{x}}{2}} \geq \alpha\right)$$

$$P(\underbrace{\vec{x}^T x} \leq -2\ln\alpha)$$

$\alpha = 0.5$

$\alpha = 0.1$

$\vdots$

$\downarrow$

$$P\Big(\underbrace{\sum_{i=1}^{d} x_i^2} \leq -2\ln\alpha\Big)$$

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}$$

$\vec{x}^T\vec{x} = \|\vec{x}\|^2$

$\underbrace{\text{sum of } d \text{ squares}}$

χ^2 distribution with d degrees of freedom

$$P(\chi^2_d \leq -2\ln\alpha)$$

... cumulative distribution for χ^2

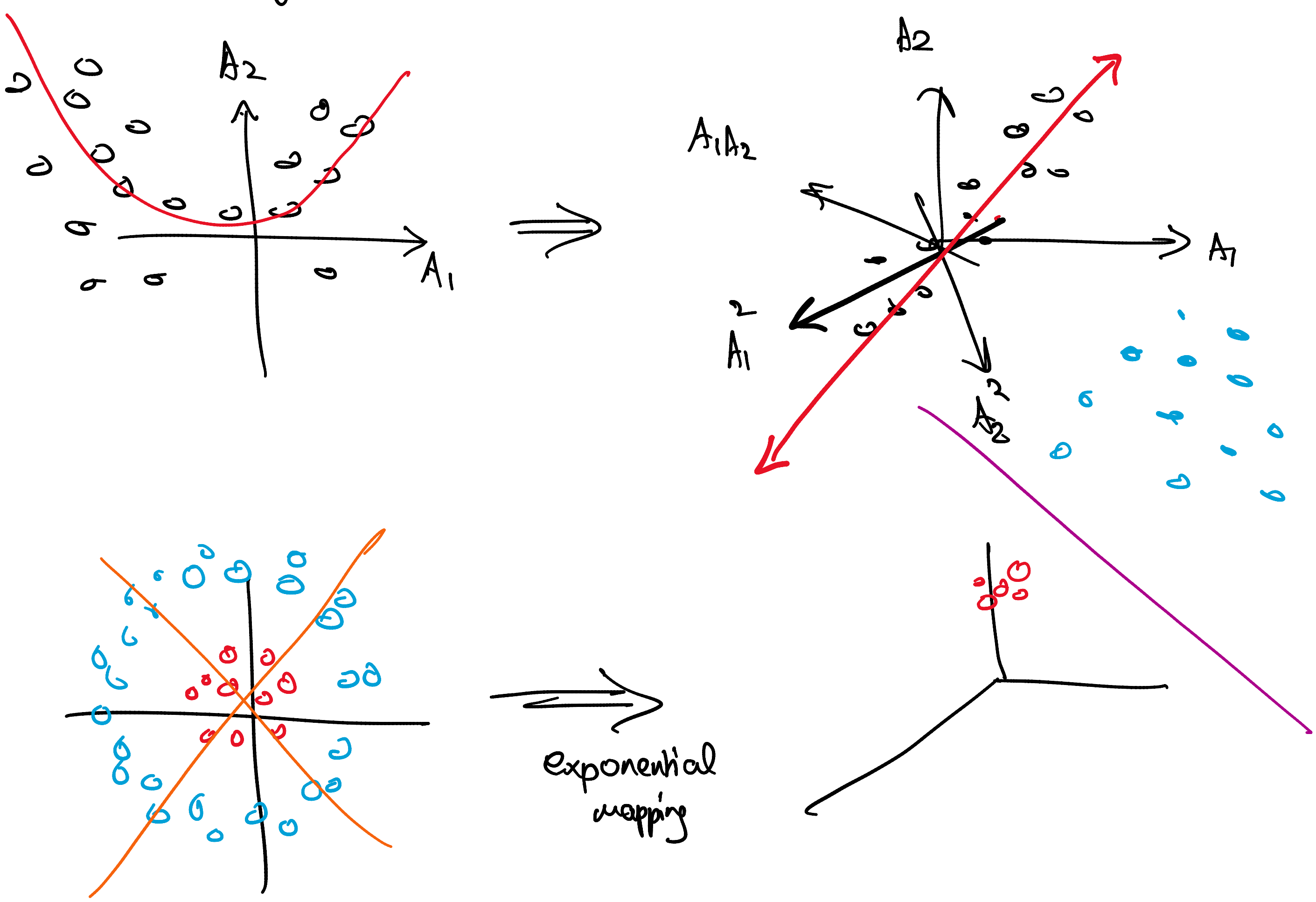
get from cumulative distribution for χ^2

$$\lim_{d \to \infty} P\left(\chi_d^2 \leq -2 \ln \alpha \right) = 0$$

$$\alpha > 0$$

We will not find points within α fraction no matter how small α is!

Benefit of high dimensions?



Linear regression

$X \equiv$ temp

$Y \equiv$ pressure

$$D = \begin{array}{cc} X & Y \\ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} & \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \end{array}$$

independent attribute

target variable (dependent)

input

$$f(x_i) = \hat{y}_i$$

prediction model

predict the target

given x_i

1) $\hat{y}_i = f(x_i)$ Compute the output

2) Compare $\hat{y}_i$ (prediction) with y_i (truth)

$$\min_{\theta} \; J = d(y_i, \hat{y}_i)$$

loss function / error

θ is parameters of the model f

regression $y_i, \hat{y}_i \in \mathbb{R}$

---

linear regression

d-dim $\vec{x}_i \in \mathbb{R}^d$

$\underbrace{\text{given } y_i \in \mathbb{R} \text{ (true)}}_{\text{supervised learning}}$

$$\hat{y}_i = f(\vec{x}_i) = w_1 x_{i1} + w_2 x_{i2} + \dots w_d x_{id} + b$$

supervised learning

$$\vec{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{id} \end{pmatrix}$$

$$\theta = \{w_1, w_2, \dots, w_d\} = \vec{w}$$

weight vector

b = bias

hyperplane!

$$\hat{y}_1 = f(x_1) = \vec{w}^T \vec{x}_1 + b$$

$$\hat{y}_2 = f(x_2) = \vec{w}^T \vec{x}_2 + b$$

$$\vdots$$

2d:

X	Y
x_1	y_1
x_2	y_2
$\vdots$	
x_n	y_n

$$\hat{y}_1 = w \cdot x_1 + b$$

$$\hat{y}_2 = w \cdot x_2 + b$$

$$\vdots$$

$$\hat{y}_n = w \cdot x_n + b$$

$$\hat{y}_i = f(x_i) = w \cdot x_i + b$$

$\mathcal{L} \equiv$ sum of squared errors

$$\mathcal{L} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\min_{w,b} J = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$= \sum_{i=1}^{n} \left(y_i - (w \cdot x_i + b) \right)^2$$

$$J = \sum_{i=1}^{n} \left(y_i^2 - 2 y_i \underbrace{(w x_i + b)} + (w x_i + b)^2 \right)$$

$$\frac{\partial J}{\partial b} = \sum_{i=1}^{n} \left(-2 y_i + 2 (w x_i + b) \right) = 0$$

$$= \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} w x_i + \sum_{i=1} b$$

$$\Rightarrow \quad n \cdot b = \sum_{i=1}^{n} y_i - w \left(\sum_{i=1}^{n} x_i \right)$$

$$\Rightarrow \quad b = \frac{1}{n} \sum y_i - w \cdot \left(\frac{1}{n} \sum x_i \right)$$

$$\boxed{b = \mu_y - w \cdot \mu_x}$$

$\mu_y = \text{mean}(Y)$

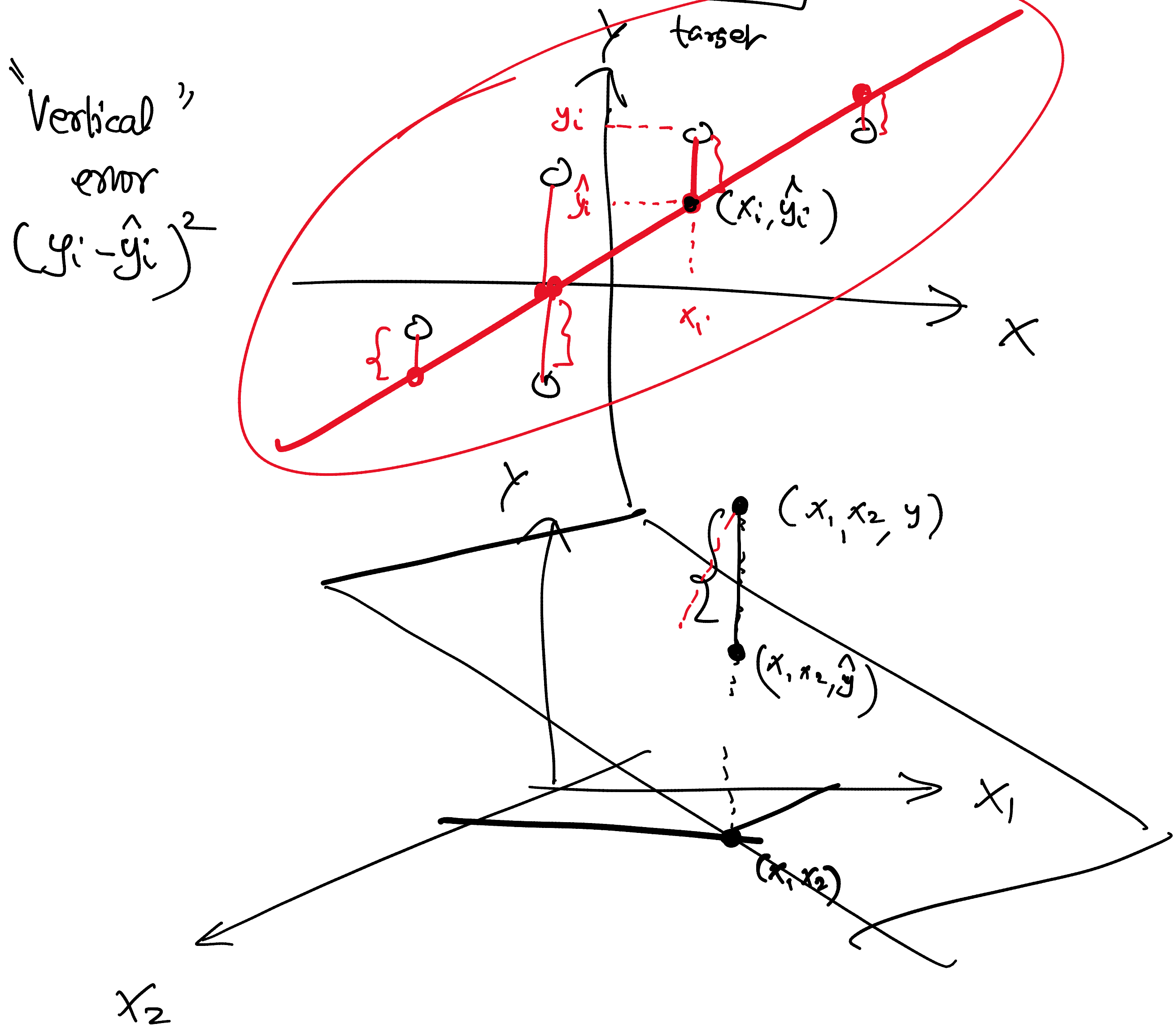
$\mu_x = \text{mean}(X)$

$$\frac{\partial J}{\partial w} = \sum_{i=1}^{n} \left(-2 y_i x_i + 2 (w x_i + b) \cdot x_i \right) = 0$$

$$\Rightarrow \quad w = \frac{\sum x_i y_i - n \cdot \mu_x \cdot \mu_y}{\sum x_i^2 - n \cdot \mu_x^2}$$

$$\boxed{\begin{aligned} w &= \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} \\ b &= \mu_y - w \cdot \mu_x \end{aligned}}$$

$$\boxed{b = \mu_y - w \cdot \mu_x}$$



Column perspective

$$n \text{ equations} \left\{ \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} \begin{matrix} = w\, x_1 + b \\ = w\, x_2 + b \\ \vdots \\ = w\, x_n + b \end{matrix} \right. \implies \boxed{\underset{n}{\underline{\hat{Y}}} = w \cdot \underset{n}{\underline{\vec{X}}} + b \cdot \vec{1}}$$

$$\hat{y}_n = w x_n + b$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\underbrace{\hat{Y} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix} \in \mathbb{R}^n \qquad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \vec{1} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}}_{\mathbb{R}^n}$$

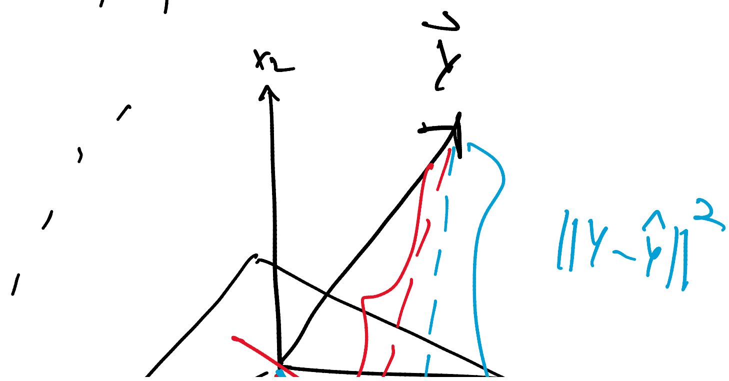
$$\vec{X},\ \vec{Y},\ \vec{1} \text{ and } \hat{\vec{Y}} \in \mathbb{R}^n$$

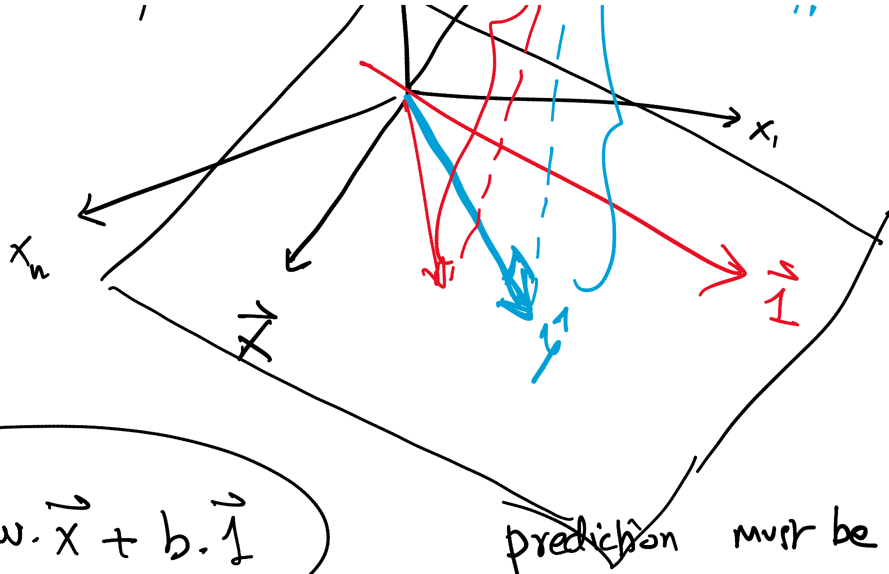
$$\hat{\vec{Y}} = w \cdot \vec{X} + b \cdot \vec{1}$$

$$\underbrace{Y \text{ vs. } \hat{Y}}_{\mathcal{L}}$$

$$\mathcal{L} = \|Y - \hat{Y}\|^2$$

$$= \left\| \begin{pmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{pmatrix} \right\|^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$



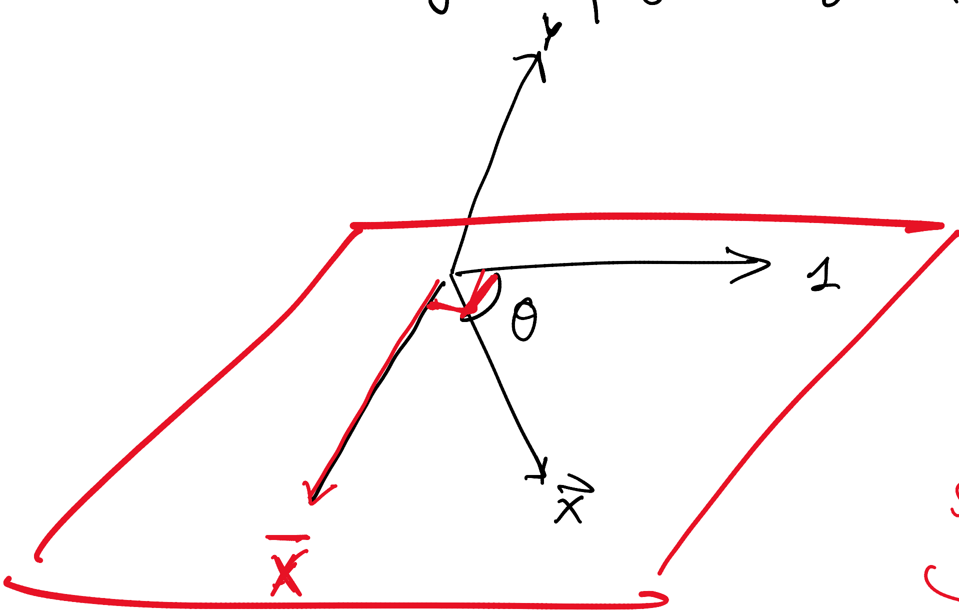


$$\hat{Y} = w \cdot \vec{X} + b \cdot \vec{1}$$

prediction must be a linear combination of $\vec{X}$ and $\vec{1}$ vector

$$\hat{Y} \in \text{span}\{\vec{X}, \vec{1}\}$$

$\hat{Y}$ is simply the projection of Y onto the space spanned by $\vec{X}$ and $\vec{1}$



$$\underbrace{\text{Span}\{\bar{X}, \vec{1}\}}_{\text{orthogonal basis!}} = \text{span}\{\vec{X}, \vec{1}\}$$

~~not~~

$$\boxed{\hat{Y} = \text{proj}_{\vec{X}}(Y) + \text{Proj}_{\vec{1}}(Y)}$$