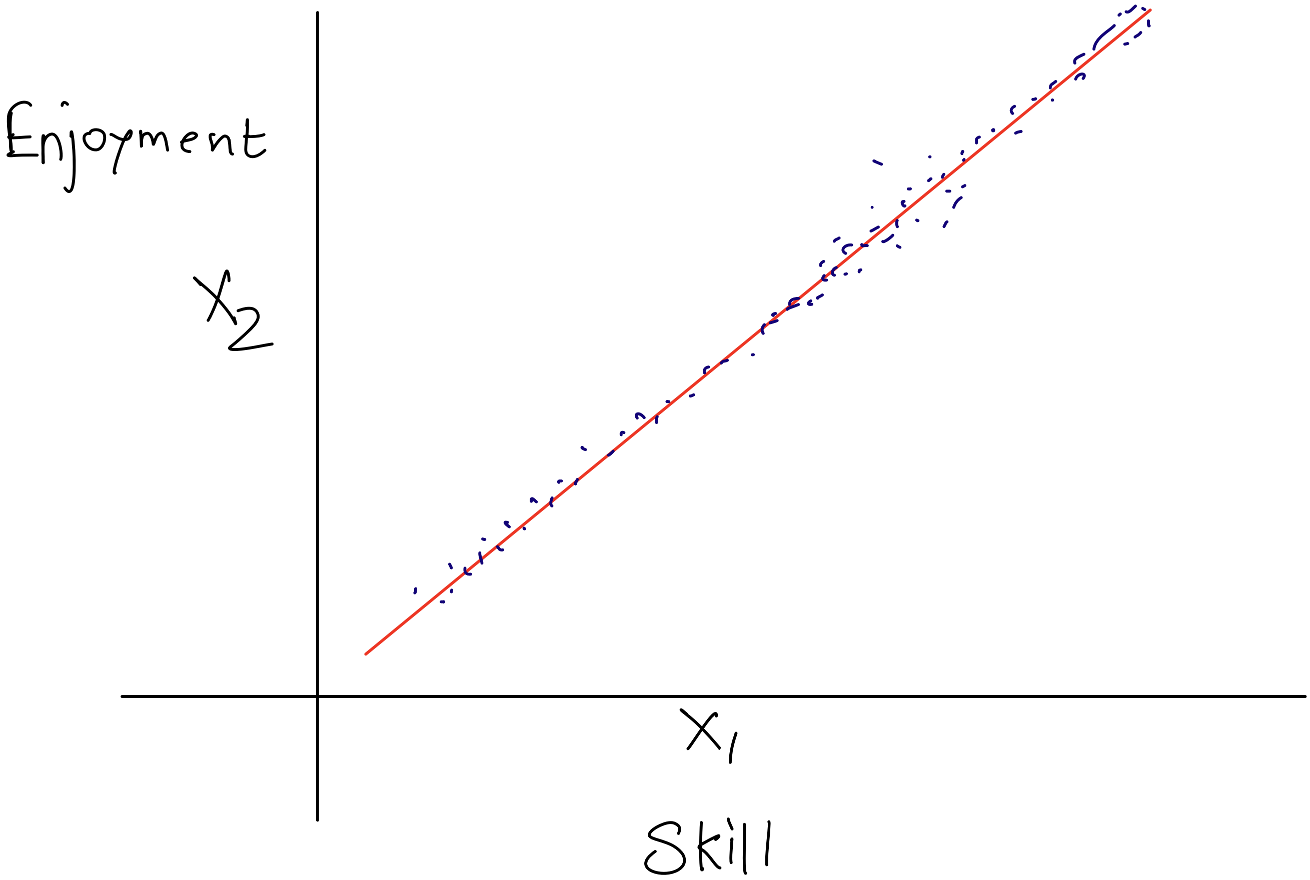
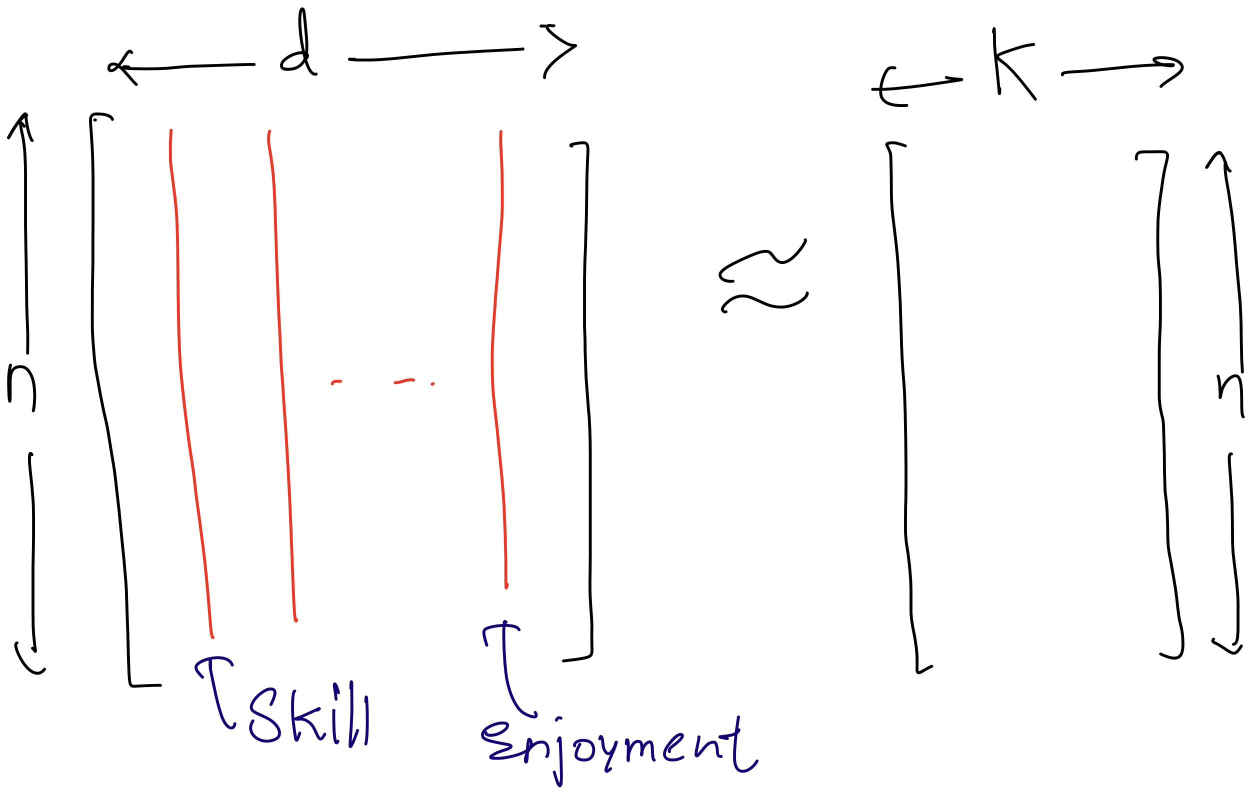
$$\{x^{(i)}\}_{i=1}^{n} \quad x^{(i)} \in \mathbb{R}^d$$



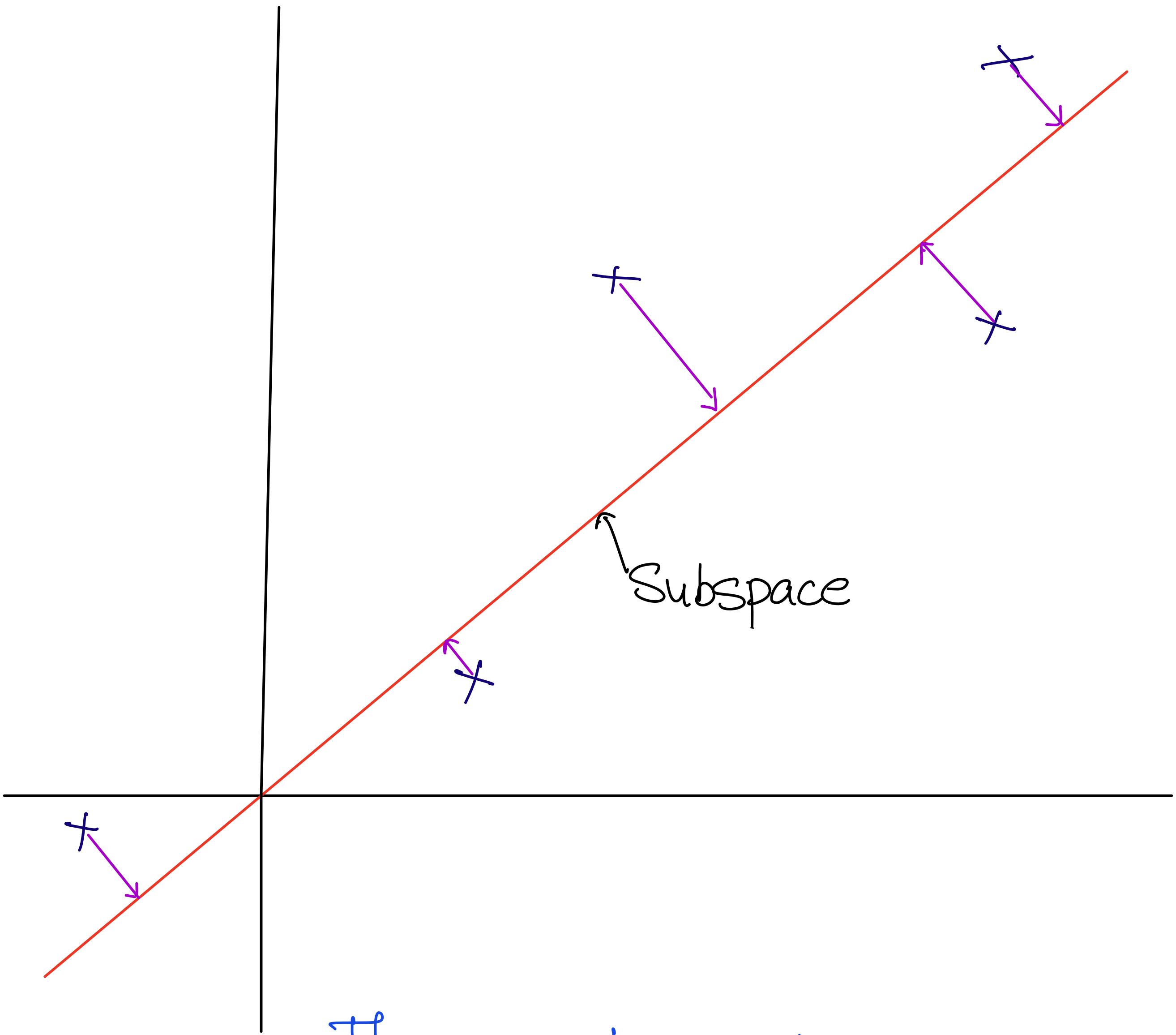


$$x_j^{(i)} := \frac{x_j^{(i)} - \mu_j}{\sigma_j}$$

$$\mu_j = \frac{1}{n}\sum_{i=1}^{n} x_j^{(i)}$$

$$\sigma_j^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_j^{(i)} - \mu_j\right)^2$$

Standardizing the dataset (Standardize each column independently)

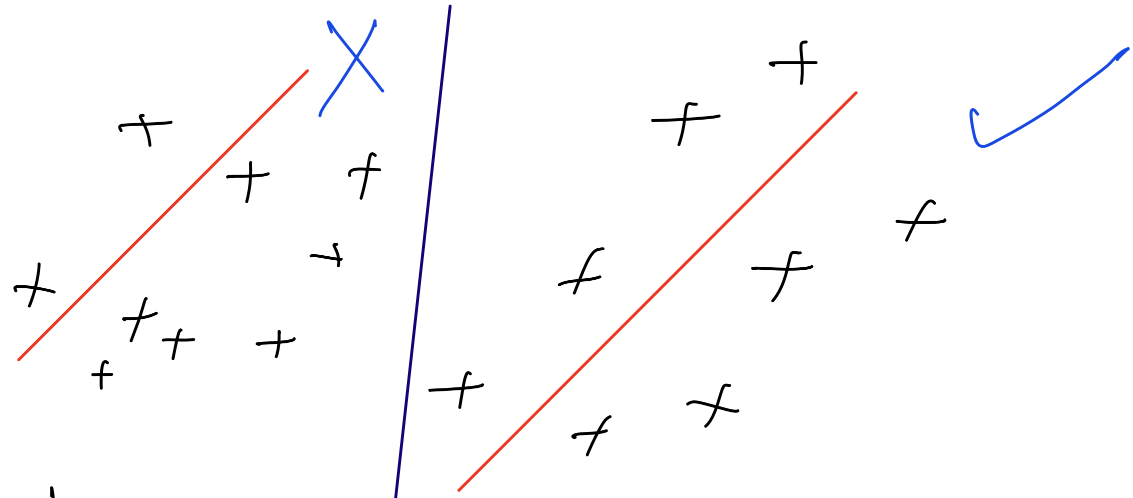


There can be ∞ subspaces

We find subspace which has maximum variance.

We want to lose no. of dimensions but not variance.

Eg.



Var is low as all points on one side

Var high as points are spreadout.

$u \in \mathbb{R}^d$ unit length (subspace)

$\text{Proj}(u)\,\vec{x}$

$$\frac{uu^T\vec{x}}{u^Tu} = \frac{(x^Tu)\,u}{1} = (x^Tu)\,u$$

What we want to do is max the norm.
i.e. we want to max dist from origin to the projected point for all points

$$u = \arg\max_u \frac{1}{n}\sum_{i=1}^{n} \|Proj(u)\,x\|_2^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} \|(x^Tu)u\|_2^2 \qquad \|u\| = 1$$

$$= \frac{1}{n}\sum_{i=1}^{n} (x^Tu)^2$$

$$\frac{1}{n}\sum_{i=1}^{n} (u^Txx^Tu)$$

$$u^T\left(\frac{1}{n}\sum_{i=1}^{n} xx^T\right)u$$

$$\arg\max_u \; u^T(\text{Samp Cov Matrix})\,u$$

$u \rightarrow$ Eigen vector corr. max eigen value of Sample Covar. Matrix

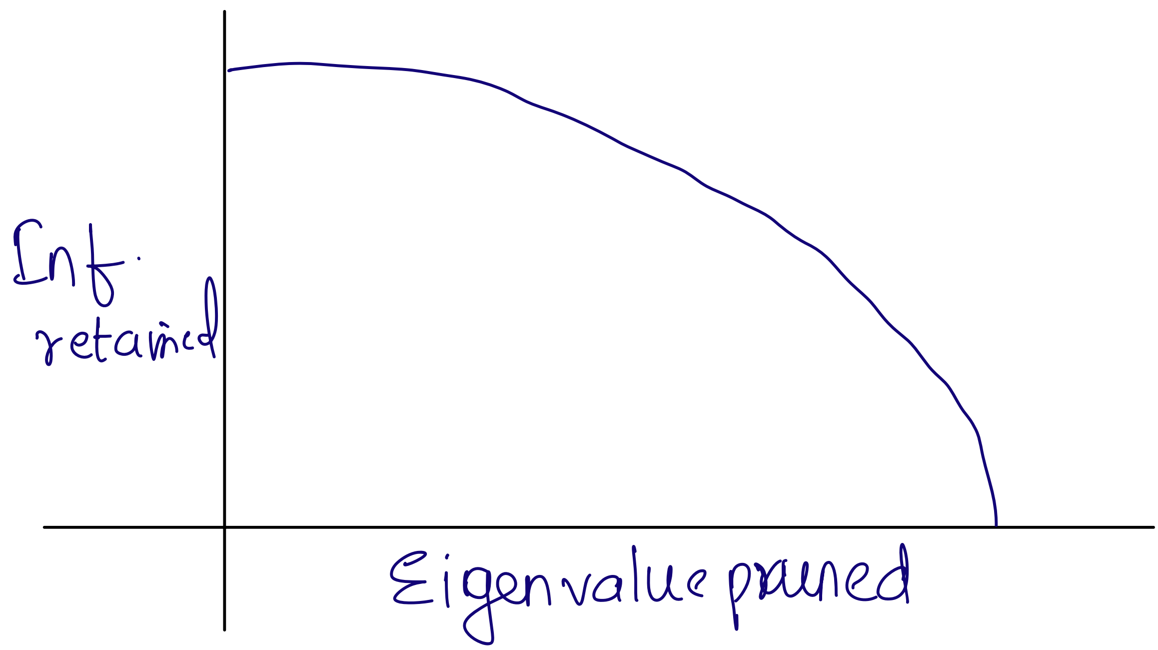
$X \rightarrow$ matrix of data

$X^T X \rightarrow$ find eigenvalue & eigenvectors

Depending on number of subspace you want to retain choose eigen vectors

Find K such that

$$\frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{d} \lambda_i} = 95\%$$



X^TX is square, sym, PSD

Eigen decom = SVD decom.

Maximizing variance = Minimizing residuals

	Probabilistic	Non Probabilistic
Clustering (Classification)	GMM	K-Means
Subspace (Regression)	Factor Analysis ↓ $d >> n$ $L \to z$ to X	PCA ↓ $n > d$ (not necessary) $U \to X$ to z

Power Iteration

$$u^{(0)} \longrightarrow (X^T X) u^0$$

$$u^{(1)} \longrightarrow \frac{(X^T X)\, u^{(0)}}{\| X^T X\, u^{(0)} \|}$$

$$(X^T X) u^{(1)} \longrightarrow \frac{u^{(2)}}{\| u^{(2)} \|}$$

ICA

$d \rightarrow$ speakers

$d \rightarrow$ microphones

$S \in \mathbb{R}^d \rightarrow$ collection of speeches

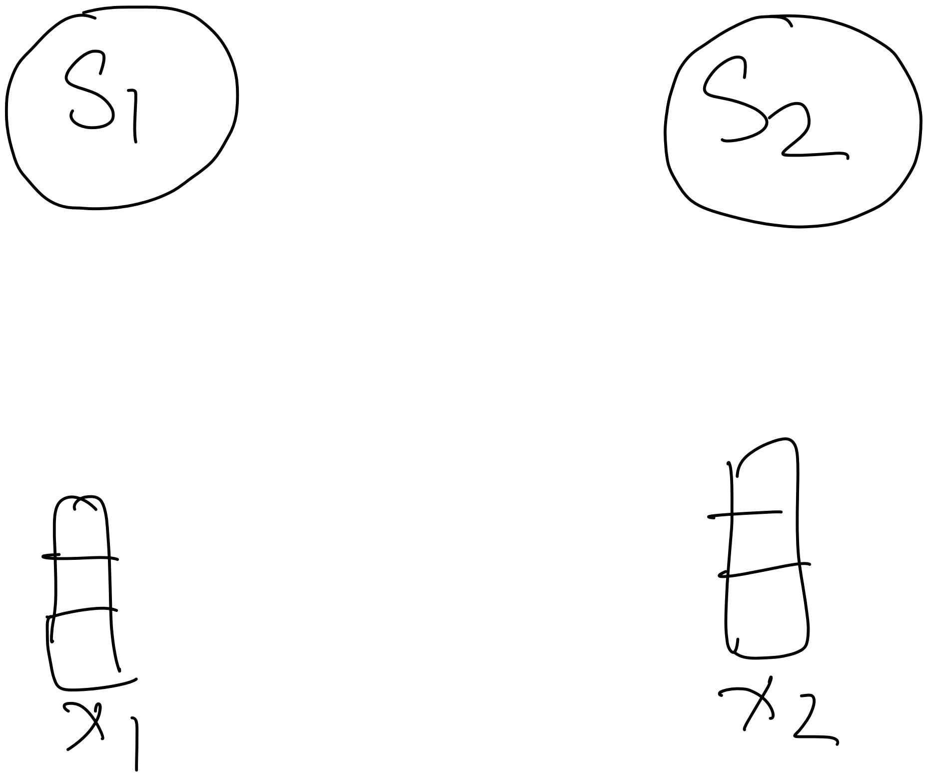
$x \in \mathbb{R}^d \longrightarrow$ recording

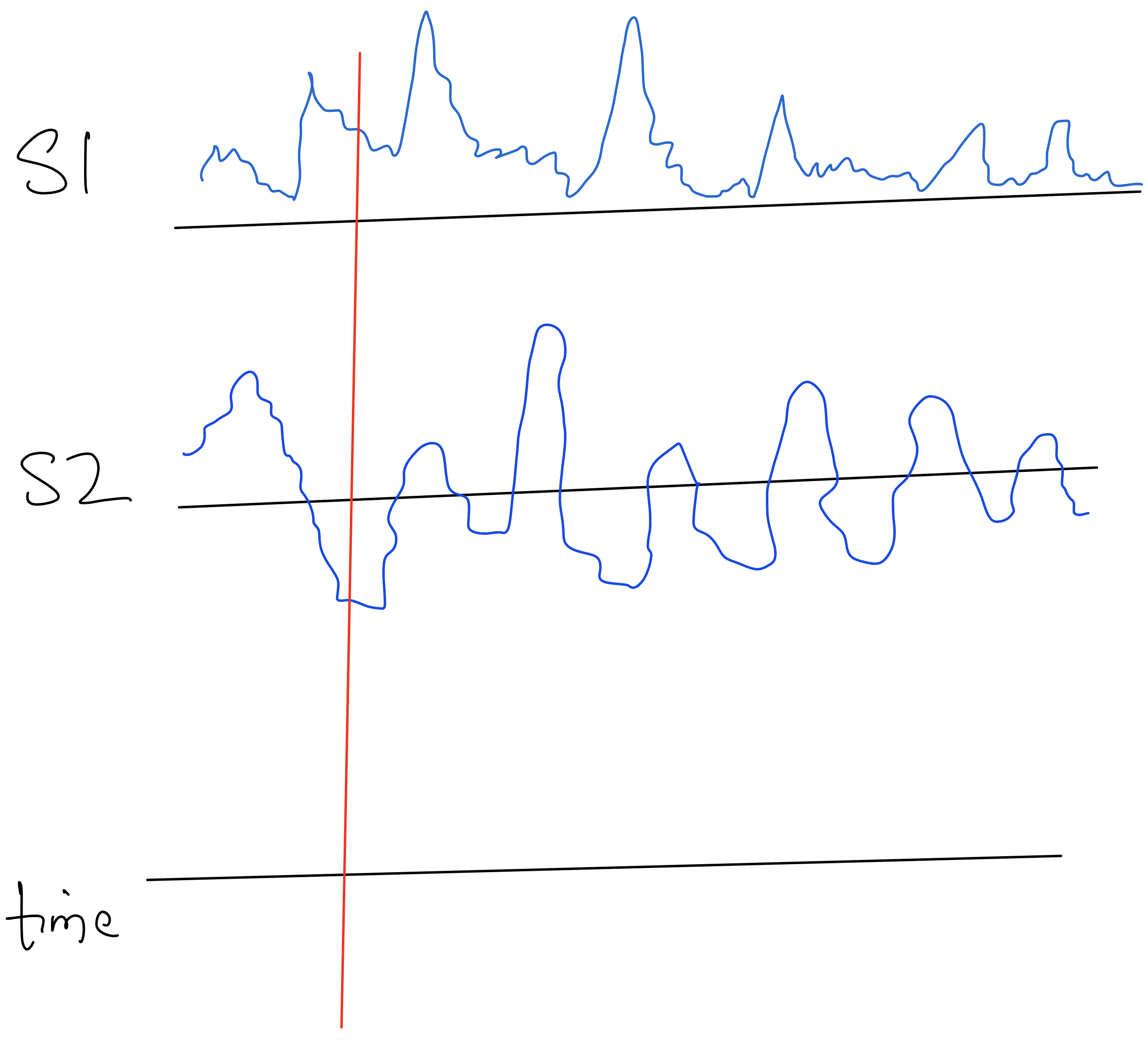
$x = \underbrace{A}_{\text{mixing matrix}} s$

$S_j^{(i)} \rightarrow j^{th}$ speaker at i^{th} time

$x_j^{(i)}$ → time

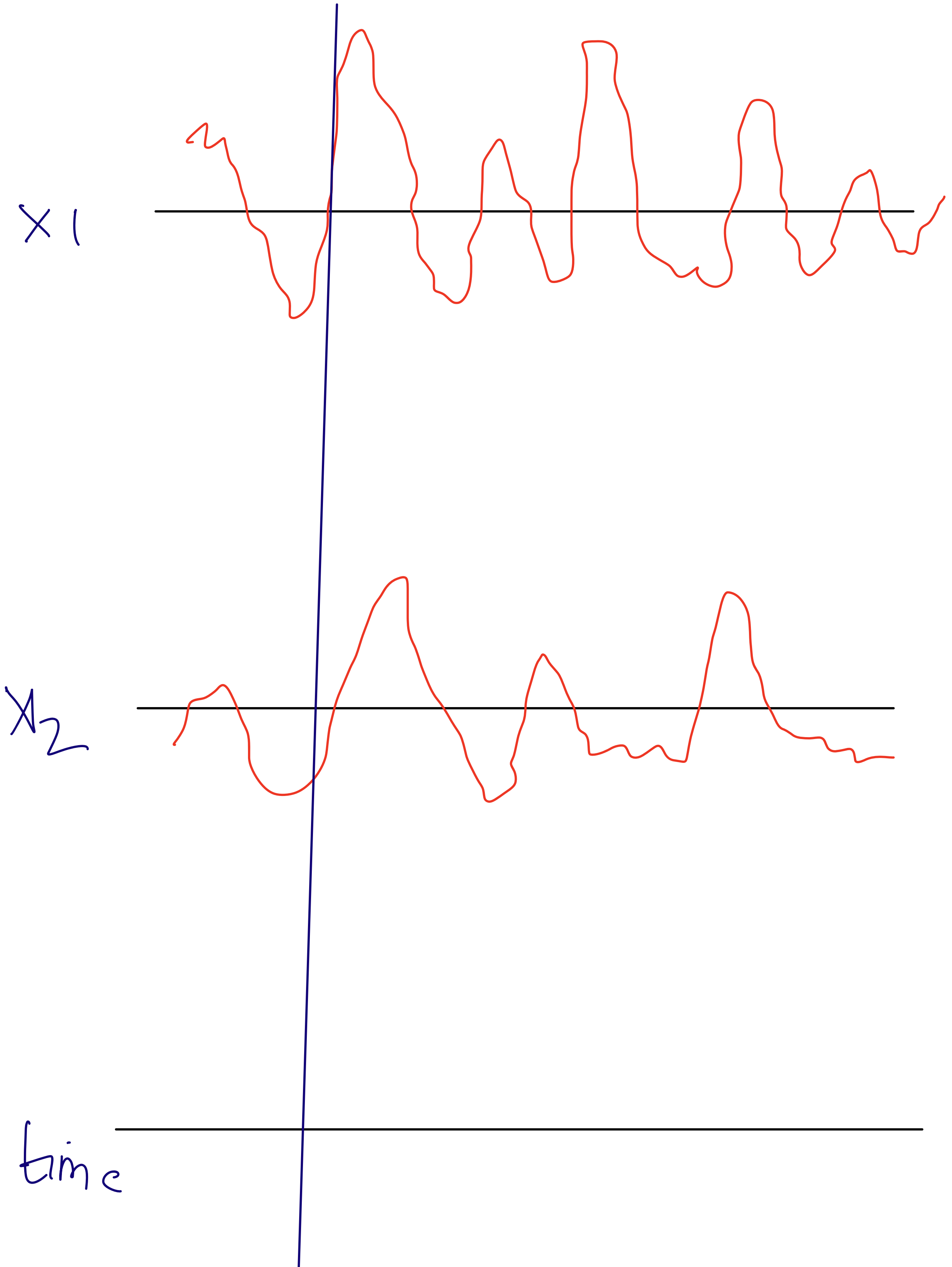
← microphone


S_1
S_2
x_1
x_2



$$S^{(i)} = (s_1^{(i)}, s_2^{(i)}) \in \mathbb{R}^d$$

$s_1^{(i)} \rightarrow$ amplitude for wave 1 at time i



$$X^{(i)} = (X_1^{(i)}, X_2^{(i)}) \in \mathbb{R}^d$$

$d = 2$

$$x^{(i)} = A s^{(i)}$$

$W = A^{-1}$ which is unmixing matrix

$$s^{(i)} \leftarrow W x^{(i)}$$

Assumption:

① $\# s = \# x$

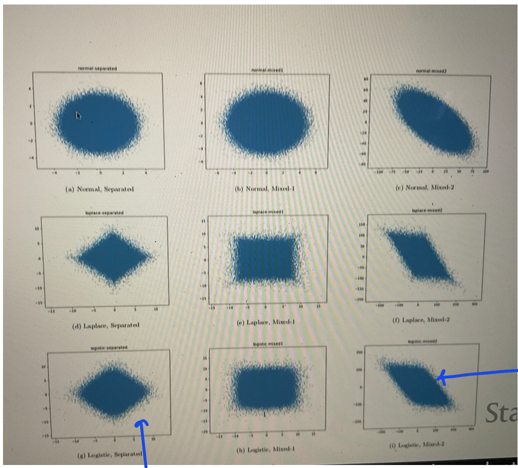
② $s = Wx$

③ $S_j \perp S_k \quad j \neq k$

④ S_j is NOT Gaussian

1st & 2nd are same.

1, 2, 3 all are different. We want to recover 1st from 3rd



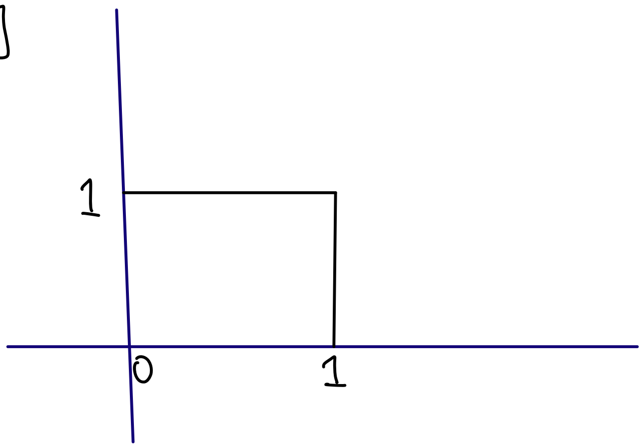
Start

Final

In 2nd & 3rd Case, we can map corners from one space to another

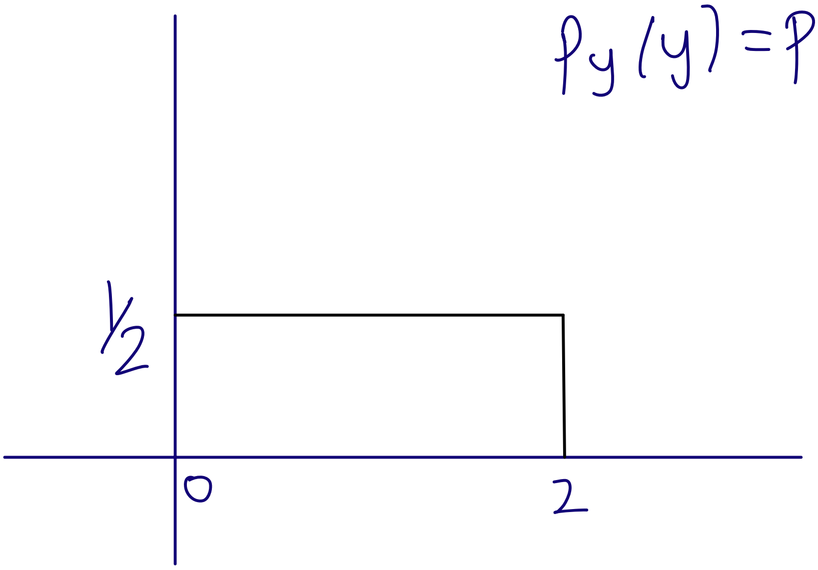
When we seperate using ICA, we may swap out the source even though we seperate.

X ~ Unif [0,1]



$Y = 2X$

$$p_y(y) = p_x\left(\frac{y}{2}\right) \cdot \frac{1}{2}$$



$$p_y(y) = p_x(Wx) \cdot \frac{1}{W}$$

$$p(x) = \prod_{j=1}^{d} p_s(w_j^T x) \cdot |W|$$

$$W = \begin{bmatrix} -w_j- \\ \\ \\ \end{bmatrix} \begin{bmatrix} x \\ \\ \end{bmatrix}$$ unmixing matrix

$P_s \sim$ logistic distribution

pdf of $f(x) = \sigma(x) \cdot (1 - \sigma(x))$

$$\ell(w) = \sum_{i=1}^{n} \left[\left(\sum_{j=1}^{d} \log\left[\sigma'(x^{(i)}) (1 - \sigma(x^{(i)})) \right) + \underbrace{\log|w|}_{\text{Jacobian}} \right]$$