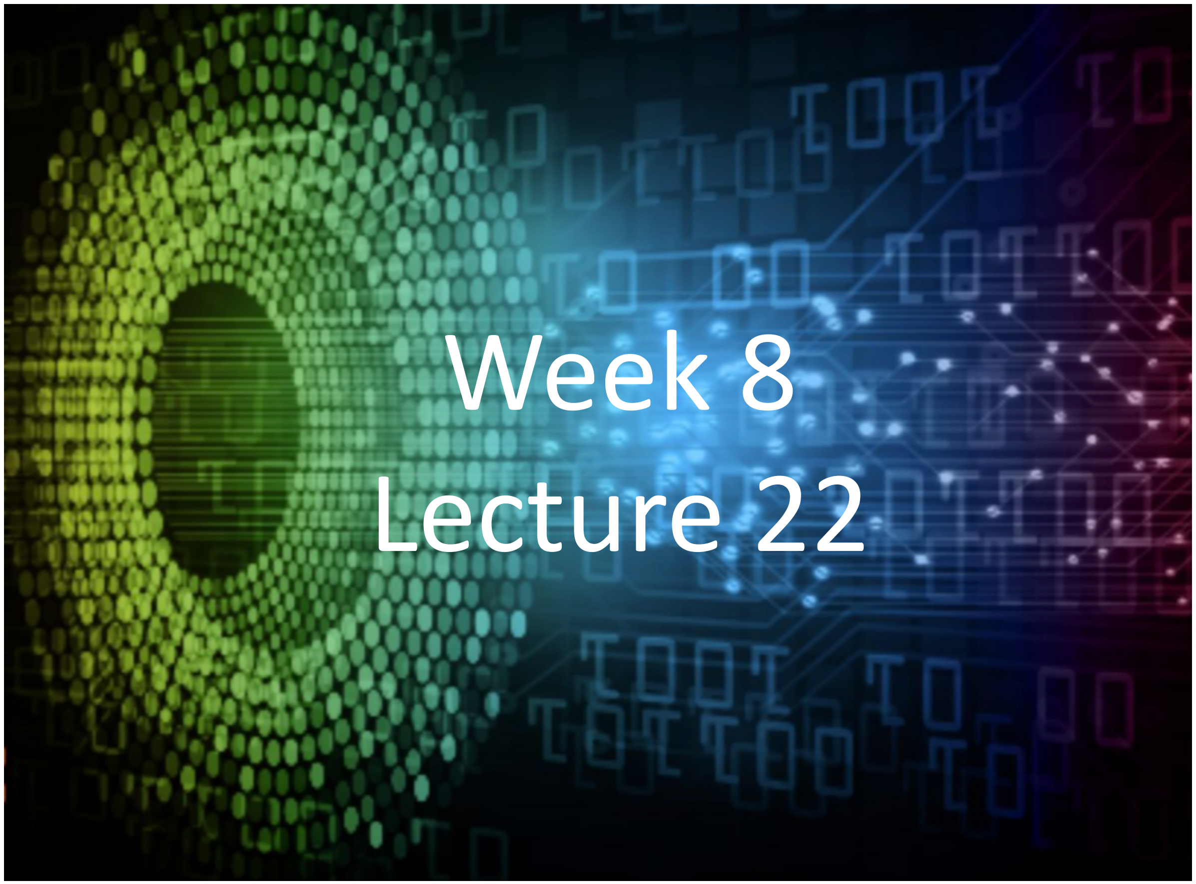
Introduction to Deep Learning Applications and Theory



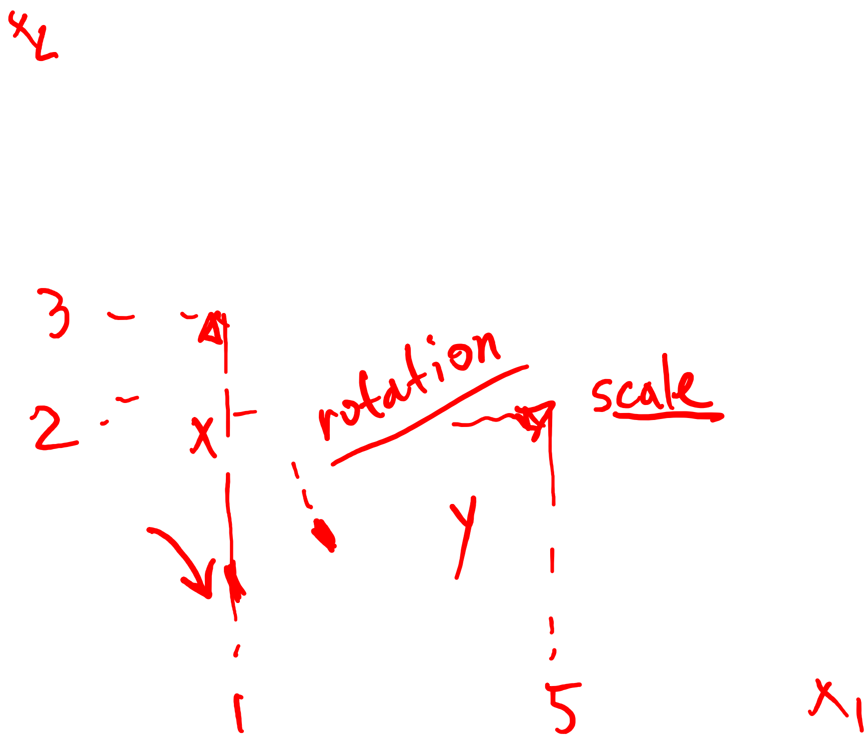
AMATH 563

This Week:
Data Organization: Decompositions, Embeddings and Clustering

- Singular Value Decomposition (SVD)
- Principal Component Analysis (PCA)
- Proper Orthogonal Decomposition (POD)
- Dynamic Mode Decomposition (DMD)
- Time-delayed Embeddings

Linear Transformations

$$\mathbf{x} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \ \underline{\mathbf{A}} = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}$$



Rotation

given θ

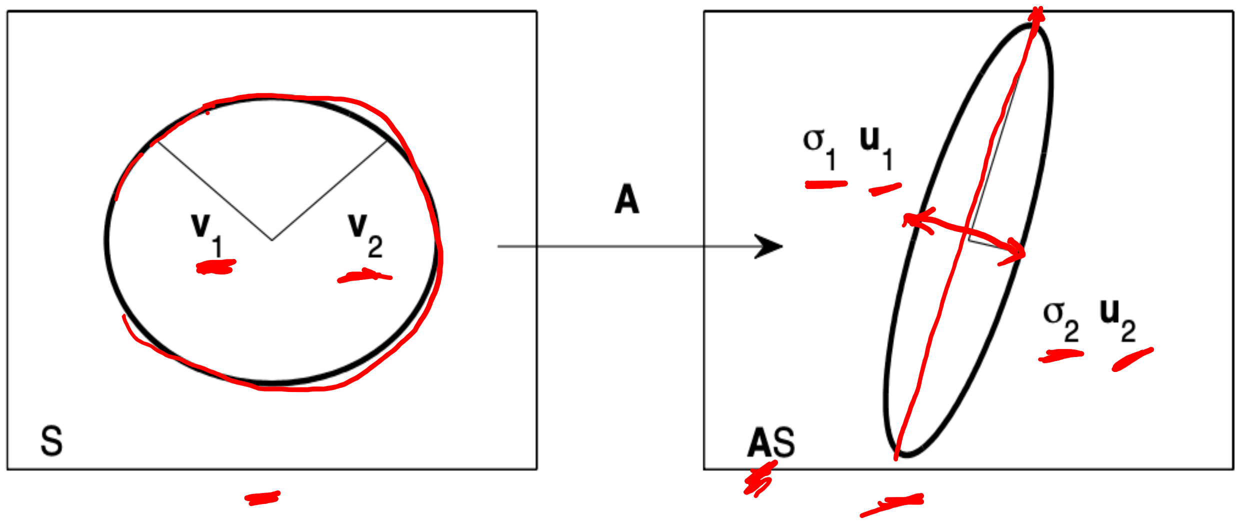
$$\mathbf{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Scaling

given α

$$\mathbf{A} = \begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix}$$

Unitary Vectors



Singular Value Decomposition

$$\boxed{\mathbf{A}\mathbf{v}_j = \sigma_j \mathbf{u}_j} \quad 1 \leq j \leq n\,.$$

$$\underbrace{\begin{bmatrix} \\ \\ \mathbf{A} \\ \\ \\ \end{bmatrix}}_{n\times n} \underbrace{\begin{bmatrix} \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ \\ \end{bmatrix}}_{n\times n} = \underbrace{\begin{bmatrix} \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \\ \\ \end{bmatrix}}_{n\times n} \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & 0 \\ & 0 & \ddots & \\ & & & \sigma_n \end{bmatrix}}_{n\times n}$$

$1\times n$

$V \cdot V^{-1}$ $\quad U \quad \Sigma \quad V^*$

$U^* = U^{-1}$, $V^* = V^{-1}$, $\|U\| = 1$, $\|V\| = 1$

SVD Properties

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^* \ .$$

m×n m×m m×n n×n

$\mathbf{U} \in \mathbb{C}^{m\times m}$ is unitary

$\mathbf{V} \in \mathbb{C}^{n\times n}$ is unitary

$\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$ is diagonal

singular values

$$\sigma_1 \ \geq \ \sigma_2 \ \geq \ \cdots \ \geq \ \sigma_p \ \geq \ 0$$

max

Computing SVD – Eigenvalue Problem

$$
\begin{aligned}
\mathbf{A}^T\mathbf{A}V &= (\mathbf{U\Sigma V}^*)^T(\mathbf{U\Sigma V}^*) \\
&= \mathbf{V\Sigma}\underbrace{\mathbf{U}^*\mathbf{U}}_{I}\mathbf{\Sigma V}^* \\
&= \mathbf{V\Sigma}^2\mathbf{V}^* \, V \quad V^*V = I
\end{aligned}
$$

$$\Rightarrow [A^TA]V = V\Sigma^2$$

$$\lambda = \sigma^2$$

$$\mathbf{A}\mathbf{A}^T U = (U\Sigma V^*)(U\Sigma V^*)^T$$

$$\ldots = U\Sigma^2$$

$$[AA^T]U = U\Sigma^2$$

$$\lambda = \sigma^2$$

Properties of SVD

Theorem: *The nonzero singular values of* $\mathbf{A}$ *are the square roots of the nonzero eigenvalues of* $\mathbf{A}^*\mathbf{A}$ *or* $\mathbf{A}\mathbf{A}^*$. *(These matrices have the same nonzero eigenvalues).*

Theorem: *The norm* $\|\mathbf{A}\|_2 = \sigma_1$ *and* $\|\mathbf{A}\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}$.

$E = \frac{\sigma_i}{\|A\|_F}$

Theorem: *For any* N *so that* $0 \le N \le r$, *we can define the partial sum*

$$\mathbf{A}_N = \sum_{j=1}^{N} \sigma_j \mathbf{u}_j \mathbf{v}_j^* \,. \qquad (14.2.103)$$

$$\|A - A_N\|_N = \sqrt{\sigma_{N+1}^2 + \sigma_{N+2}^2 + \cdots + \sigma_r^2}\,. \qquad (14.2.105)$$

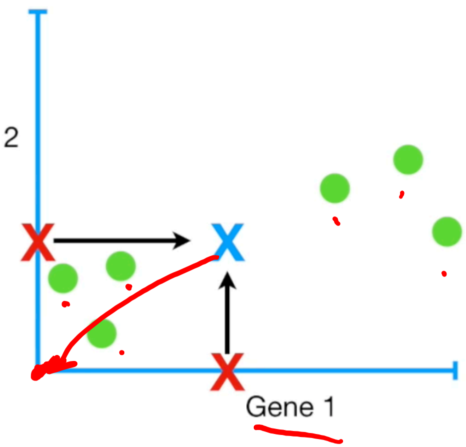
PCA

X

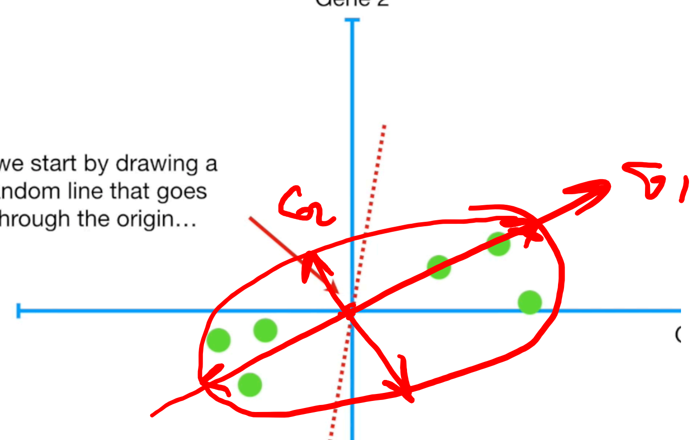
	Sample 1	Sample 2	Sample 3	Sample 4	...
Genes Variable 1	10	11	8	3	...
Variable 2	6	4	5	3	...

<Var1>

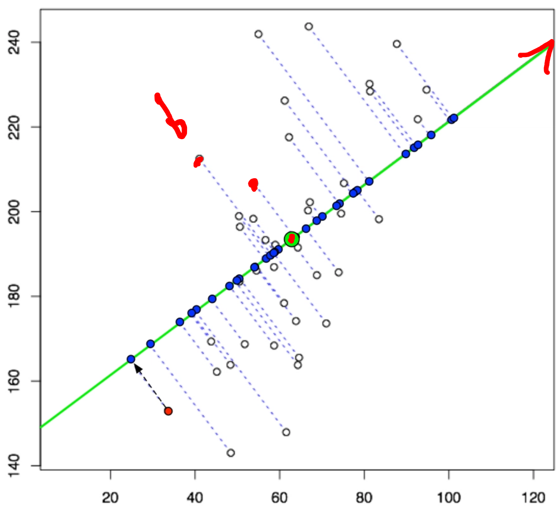
<Var2>

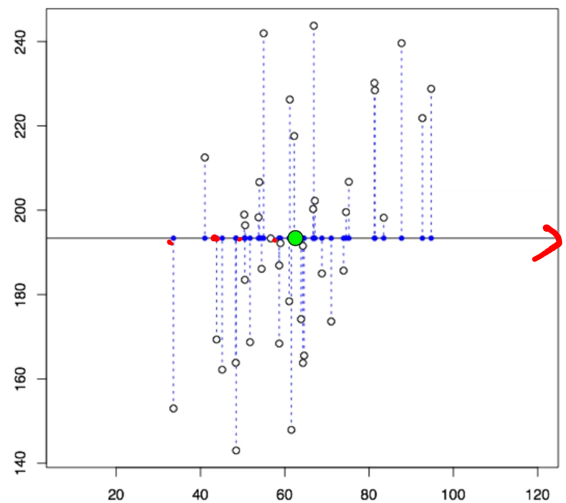


$X^T X$ Covariance Matrix



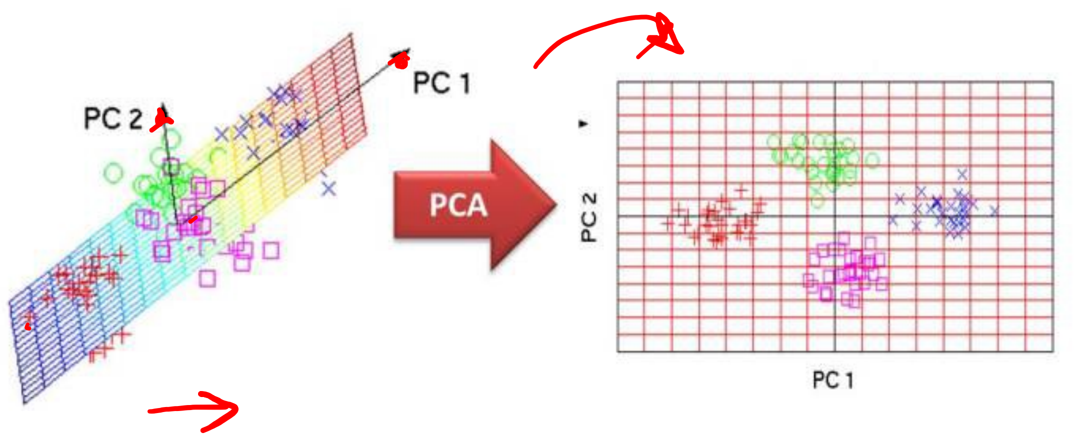
Embedding





$$X = U\Sigma V^T$$

$$XV_N = U_N\Sigma$$



(POD)

Proper Orthogonal Decomposition

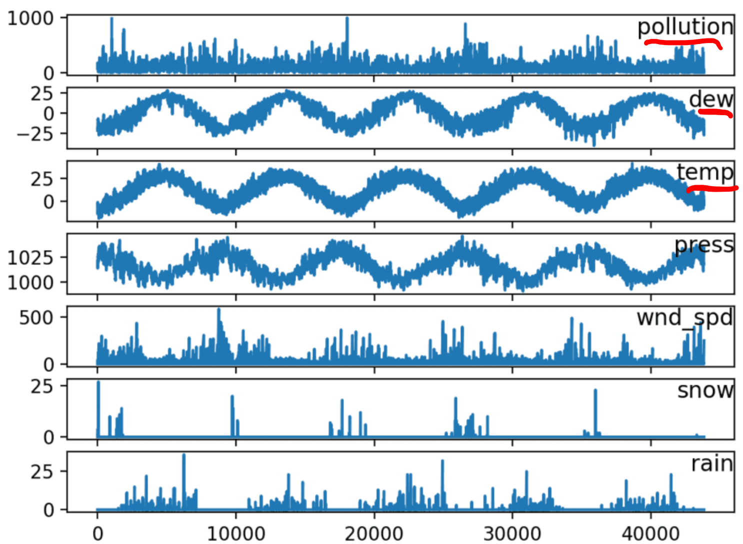
F

$$f(x,t) \approx \sum_{j=1}^{N} a_j(t)\phi_j(x)$$

space time

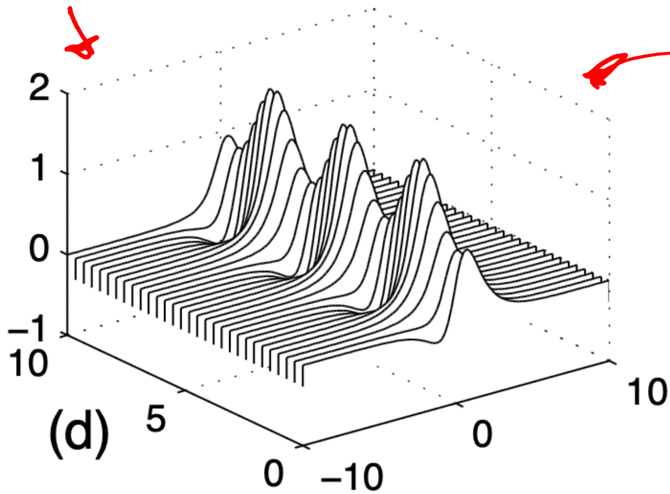
$$F(x,t) = U(x)\Sigma V(t)^T$$

sensor

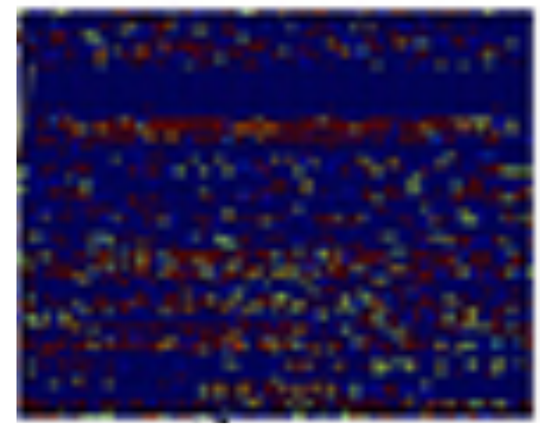


time

Proper Orthogonal Decomposition



Neurons

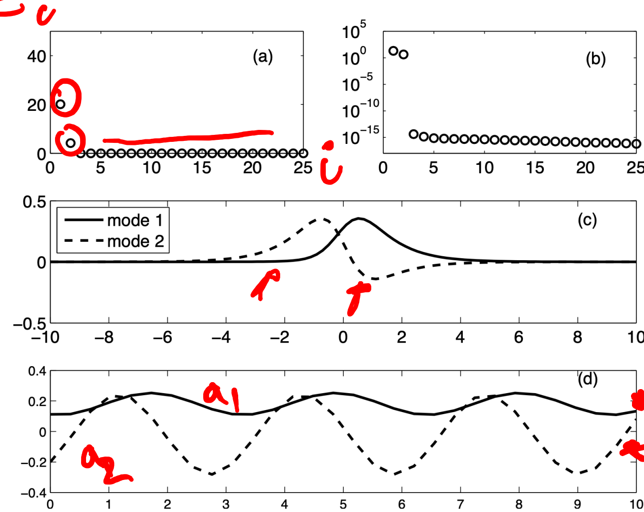


time

E_i

i

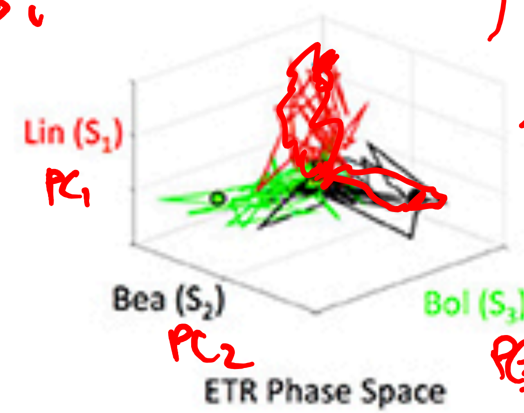
E_i



i

a_1

a_2



PC_1

PC_2

PC_3

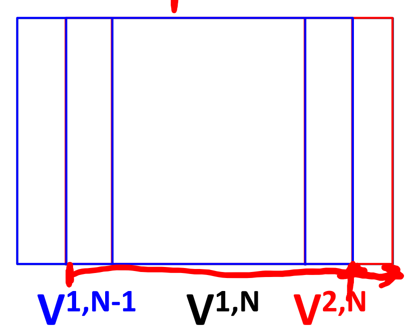
DMD

Dynamic Mode Decomposition

Dynamic mode decomposition (DMD): Computes a set of **modes** associated with fixed **oscillation** frequencies and **decay/growth** rates for given multi dimensional time series.

For data in form of **snapshots sequence**

snapshot matrix



$$V^{2,N} = A V^{1,N-1} + r e_{N-1}^T$$

DMD -- continued

Essentially, DMD is data driven (linear) method to predict the next snapshot from previous one:

$$v_{i+1} = Av_i$$

unknown

$$V_2^N = AV_1^{N-1} + re_{N-1}^T = AU\Sigma W^T + \cancel{re_{N-1}^T}$$

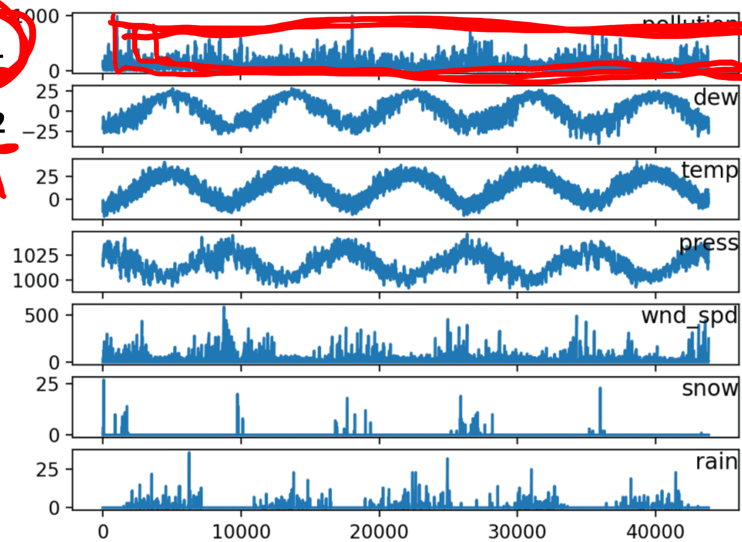
$$U^T AU = U^T V_2^N W\Sigma^{-1} \equiv \tilde{S}$$

Time Delay Embeddings

variable

$\mathbf{x_1}$

$\mathbf{x_2}$



unmeasured

time

2d+1

k+1

$$\boxed{H} = [x_1; x_1^{\delta}; \ldots, ; x_1^{(k+1)\delta}] \qquad H = \boxed{U\Sigma V^T}$$

Takens's Thm: Attractor with dim *d* can be embedded in Euclidean space with *k*>2*d* lags.

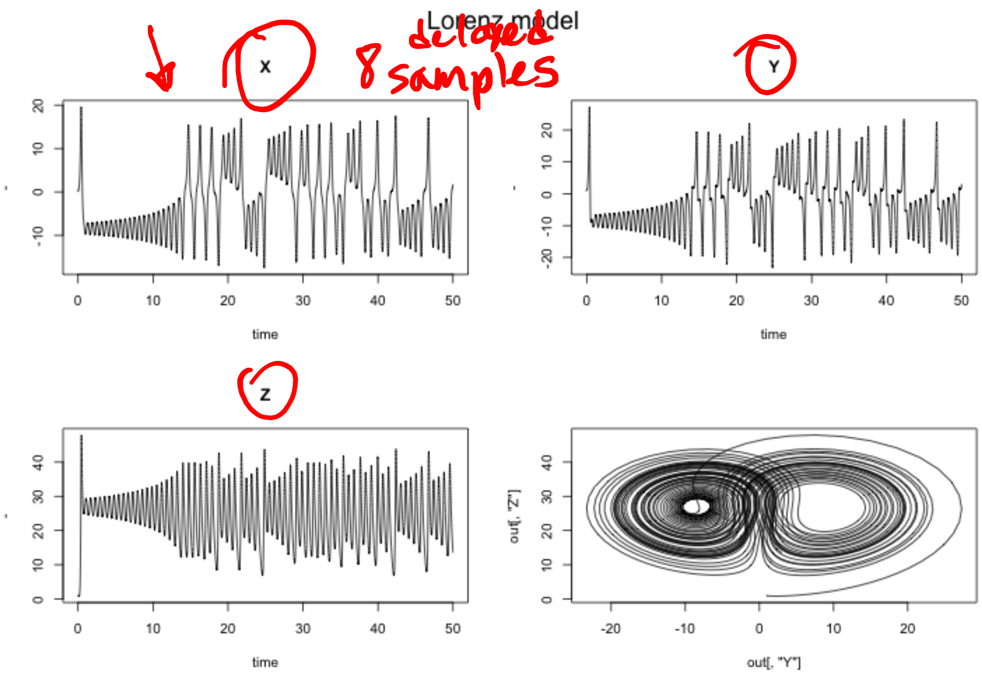
Time Delay Embeddings - Example

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x),$$

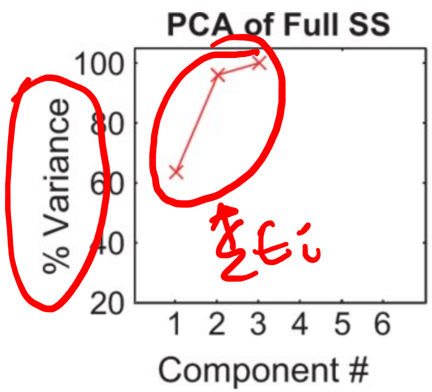
$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y,$$

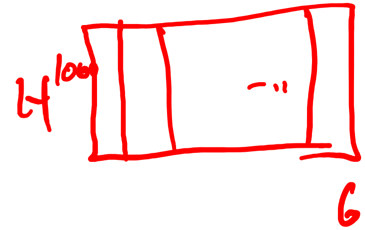
$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z.$$

delayed
8 samples



Time Delay Embeddings - Example





$U\Sigma V^*$