Feature Selection of Ordinal Partitions for Echo State Networks

by

Hector Morlet
BSc

Supervisors: Professor Michael Small
Doctor Eugene Tan

This thesis is presented for the partial requirements of the degree of Bachelor of Science with honours of the University of Western Australia.
March 21, 2025

Abstract

This thesis will present research into the use of the ordinal partitioning of time series data to inform the creation of an Echo State Network to predict this data.

Acknowledgements

I would like to thank ...

Contents

List of Figures

List of Tables

Chapter 1

Echo State Networks

- First introduced by Jaeger (2001).
- Type of recurrent neural network.
- Commonly studied form of reservoir computing.
- A form of supervised learning.
- Most weights are fixed and only some are fitted.
- Is computationally efficient compared to other neural networks.

State update equation:

$$\mathbf{s}(t+1) = f_{act}(\mathbf{W}_{in}x(t) + \mathbf{W}_{rec}\mathbf{s}(t) + \mathbf{W}_{bias})$$

- $\mathbf{W}_{in}$, $\mathbf{W}_{rec}$, $\mathbf{W}_{bias}$ and $\mathbf{s}(t)$ are randomly generated according to hyper-parameters.
- Echo state property
- $\mathbf{W}_{in}$, $\mathbf{W}_{rec}$ and $\mathbf{W}_{bias}$ are fixed.
- Only the readout vector weights is fitted.

Output equation:

$$y(t) = \mathbf{C}_{out}\mathbf{s}(t)$$

$$\mathbf{Y} = \mathbf{C}_{out}\mathbf{S}$$

Output regression with regularisation:

$$\mathbf{C}_{out} = \left(\mathbf{S}^T\mathbf{S} + \beta\mathbf{I}\right)\mathbf{S}^T\mathbf{Y}$$

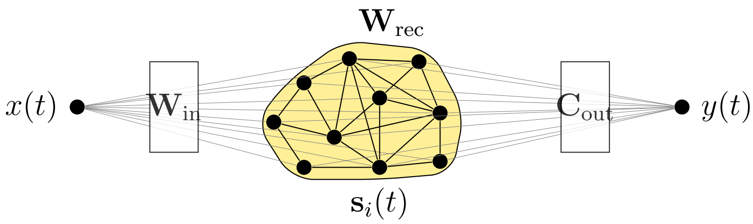

Figure 1.1: A diagram of an Echo State Network.

Chapter 2

Ordinal Partitions

'Ordinal analysis', first introduced by Bandt and Pompe (2002), proposed using the ordering of data points to partition a time series. Each data point in the timeseries is partitioned according to the ordering of its preceding points, and assigned a unique 'ordinal symbol', for example numbers from 1 to 6.

The probabilities of the timeseries transitioning from one partition to another can be calculated, giving the 'ordinal transition probabilities'.

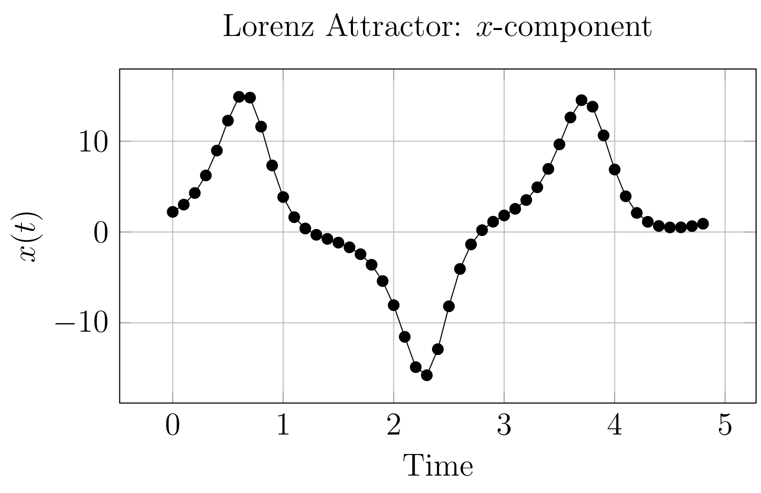

Figure 2.1: This is the caption

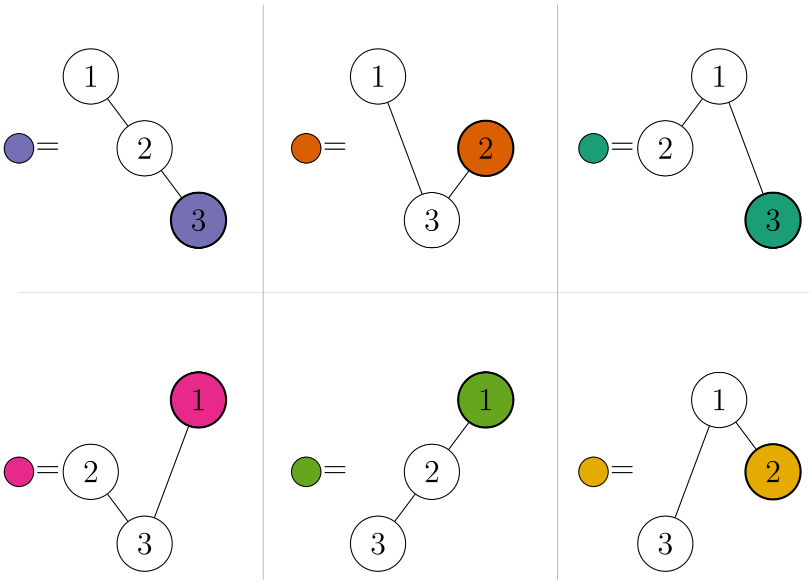

Figure 2.2: this is the caption

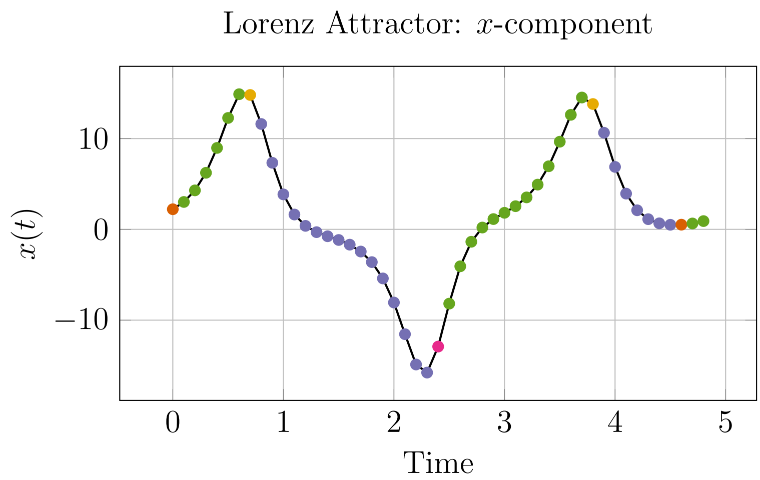

Figure 2.3: this is the caption

	0.7	0.1	0.1	0.05	0.05	0.0
	0.0	0.0	0.0	0.0	0.9	0.1
	0.7	0.1	0.1	0.05	0.05	0.0
	0.0	0.0	0.0	0.0	0.9	0.1
	0.0	0.0	0.0	0.0	0.9	0.1
	0.7	0.05	0.05	0.05	0.05	0.1

Table 2.1: Transition probabilities between ordinal partitions.

Chapter 3

Echo State Network with Ordinal Partition based readout switching

Switch the readout (C_{out}) vector based on the ordinal partition.

When fitting the readout vectors, for each partition p:

$$(\mathbf{C_{out}})_p = (\mathbf{S}_p^T \mathbf{S}_p + \beta \mathbf{I})\ \mathbf{S}_p^T \mathbf{Y}_p$$

Where

- $\mathbf{S}_p$ is $\mathbf{S}$ filtered to the states that results from data points with partition p.
- $\mathbf{Y}_p$ is $\mathbf{Y}$ filtered to data points that follow immediately from data points with partition p.

And when infering the prediction at partition p:

$$\mathbf{y}(t) = (\mathbf{C}_{out})_{P(t)} \mathbf{s}(t)$$

Where $P(t)$ is the partition of $x(t)$.

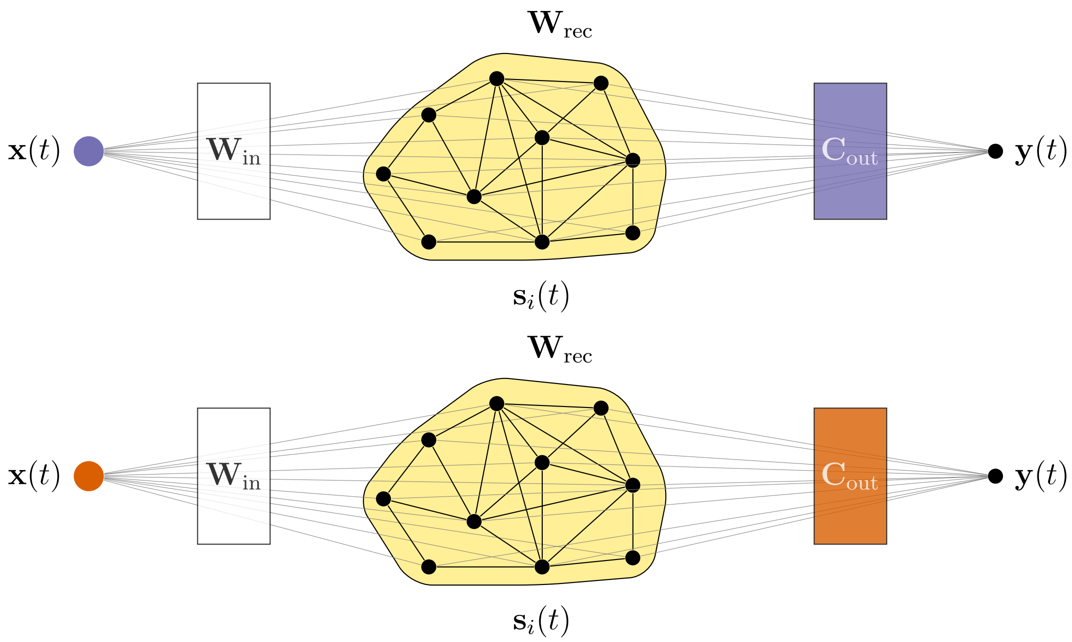

Figure 3.1: Echo State Network with readout switching.

Chapter 4

Echo State Network with Ordinal Partition based Sub-Reservoirs

- Restructure the reservoir into 'subreservoirs' for each partition.
- Feed the input based on the ordinal partition.
- Weight the connections between subreservoirs according to the ordinal transition probabilities.

For example, consider a partitioned Echo State Network with 3 partitions (not actually possible).

With 3 partitions, the time series has the following transition probabilities:

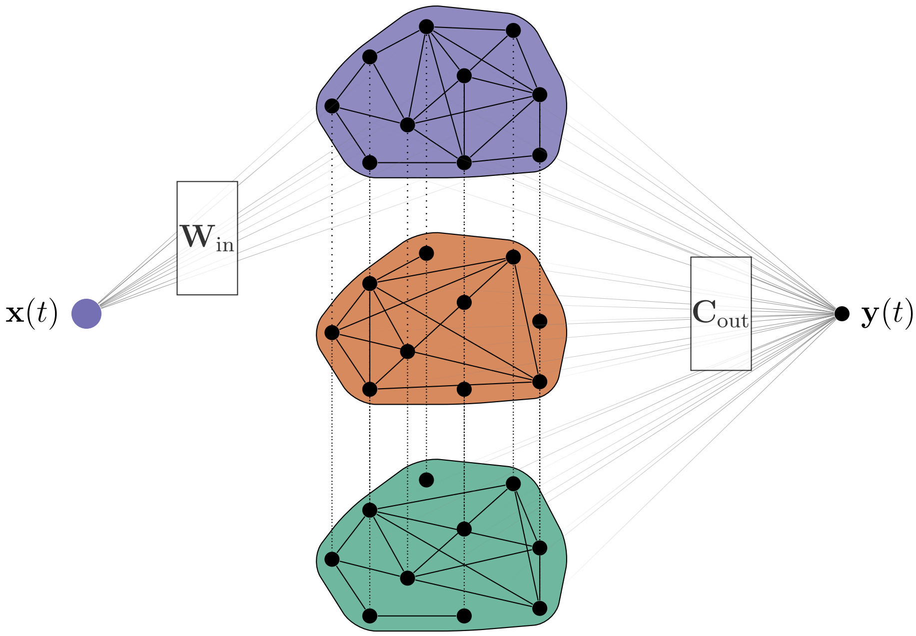

Figure 4.1: Ordinally Partitioned Echo State Network.

	🟣	🟠	🟢
🟣	0.7	0.1	0.2
🟠	0.9	0.1	0
🟢	0.1	0	0.8

Table 4.1: Example transition probabilities

Let k_{part} be the number of nodes in each partition's sub reservoir, and let $\mathbf{W}_{p,q}$ refer to the submatrix of $\mathbf{W}_{rec}$ with indices:

$$\begin{aligned} \text{Rows:} \quad & \{pk_{part}+1, pk_{part}+2, \ldots, (p+1)k_{part}\} \\ \text{Columns:} \quad & \{qk_{part}+1, qk_{part}+2, \ldots, (q+1)k_{part}\} \end{aligned}$$

Then $\mathbf{W}_{rec}$ is given by:

$$\mathbf{W}_{p,q} = \begin{cases} \mathbf{I}P(p,q), & p \neq q, \\ \mathbf{W}_{ER}, & p = q, \end{cases}$$

where

- $\mathbf{I}$ is the $k_{part} \times k_{part}$ identity matrix
- $P(p,q)$ is the probability of transitioning from partition p to partition q
- $\mathbf{W}_{ER}$ is an Erdos-Renyi randomly instantiated network.

Example:

For a k_{layer} of 4 and 3 partitions (not actually possible).

$$W_{rec} =$$

$$\begin{bmatrix}
0.12 & 0 & 0.34 & -0.67 & 0.78 & 0 & -0.23 & 0.11 & -0.56 & 0 & 0.47 & 0 \\
-0.44 & 0.23 & 0 & 0 & -0.76 & 0.54 & 0 & -0.34 & 0 & 0.56 & 0.89 & 0 \\
0.23 & 0.78 & 0 & 0 & 0.34 & 0.89 & 0 & 0.45 & 0.67 & 0 & -0.89 & -0.75 \\
0 & 0.34 & 0.89 & 0 & 0.23 & 0 & -0.67 & 0.12 & 0 & 0.56 & 0 & -0.78 \\
0.67 & 0.12 & 0 & 0.45 & 0 & 0.78 & 0 & 0.89 & 0.12 & 0 & 0.23 & 0 \\
0 & 0.56 & 0.89 & 0 & 0.23 & 0 & 0.67 & 0 & 0.34 & 0.78 & 0 & 0 \\
0.45 & 0 & 0.89 & 0.23 & 0 & 0.56 & 0.78 & 0 & 0.12 & 0 & 0.45 & 0.1 \\
0.78 & 0.12 & 0 & 0 & 0.34 & 0 & 0.45 & 0.56 & 0 & 0.89 & 0 & 0 \\
0.89 & 0.23 & 0 & 0.56 & 0 & 0.89 & 0 & 0 & 0.34 & 0 & 0.78 & 0.24 \\
0.34 & 0 & 0.78 & 0 & 0.12 & 0 & 0 & 0.67 & 0 & 0.12 & 0 & 0 \\
0.78 & 0.12 & 0 & 0.34 & 0 & 0.78 & 0.89 & 0 & 0.23 & 0.56 & 0 & -0.95 \\
0 & 0.89 & 0.23 & 0 & 0.56 & 0 & 0.89 & 0.12 & 0 & 0.34 & 0 & 0
\end{bmatrix}$$

Example:

For a k_{layer} of 4 and 3 partitions (not actually possible).

$$W_{rec} =$$

$$\begin{bmatrix}
0.12 & 0 & 0.34 & -0.67 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-0.44 & 0.23 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.23 & 0.78 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.34 & 0.89 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.78 & 0 & 0.89 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.23 & 0 & 0.67 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.56 & 0.78 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.34 & 0 & 0.45 & 0.56 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.34 & 0 & 0.78 & 0.24 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.12 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.23 & 0.56 & 0 & -0.95 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.34 & 0 & 0
\end{bmatrix}$$

Example:
For a k_{layer} of 4 and 3 partitions (not actually possible).

$$W_{rec} =$$

$$\begin{bmatrix}
0.12 & 0 & 0.34 & -0.67 & 0.1 & 0 & 0 & 0 & 0.2 & 0 & 0 & 0 \\
-0.44 & 0.23 & 0 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0.2 & 0 & 0 \\
0.23 & 0.78 & 0 & 0 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0.2 & 0 \\
0 & 0.34 & 0.89 & 0 & 0 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0.2 \\
0.9 & 0 & 0 & 0 & 0 & 0.78 & 0 & 0.89 & 0 & 0 & 0 & 0 \\
0 & 0.9 & 0 & 0 & 0.23 & 0 & 0.67 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.9 & 0 & 0 & 0.56 & 0.78 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.9 & 0.34 & 0 & 0.45 & 0.56 & 0 & 0 & 0 & 0 \\
0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.34 & 0 & 0.78 & 0.24 \\
0 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.12 & 0 & 0 \\
0 & 0 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0.23 & 0.56 & 0 & -0.95 \\
0 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0.34 & 0 & 0
\end{bmatrix}$$

To only feed the input into the relevant layer, we mask all values of $\mathbf{W}_{in}$ except those in the relevant partition.

Given a randomly instantiated input vector $\mathbf{W}_{all}$ of size $k_{part}O$ where O is the number of partitions, we can define $\mathbf{W}_{in}$ as a function of p_t, the partition of the input at time t:

$$\mathbf{W}_{in}(p_t)_i = \begin{cases} (\mathbf{W}_{all})_i, & k_{part}(p_t - 1) < i \leqslant k_{part}p_t \\ 0, & otherwise. \end{cases}$$

Example:

$$\mathbf{W}_{all} = \begin{bmatrix} 0.12 & 0.5 & 0.34 & -0.67 & 0.1 & -0.43 & 0.98 & -0.64 & 0.2 & 0.45 & 0.2 & -0.45 \end{bmatrix}$$

$$\mathbf{W}_{in}(1) = \begin{bmatrix} 0.12 & 0.5 & 0.34 & -0.67 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{W}_{in}(2) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.1 & -0.43 & 0.98 & -0.64 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{W}_{in}(3) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.2 & 0.45 & 0.2 & -0.45 \end{bmatrix}$$

Chapter 5

Echo State Network with Ordinal Partition based Sub-Reservoirs and Stochastic Sub-Reservoir Connections

Appendix A

Appendix Title

Appendix A content

Bibliography