Time Series and Sequence Learning

Temporal Convolutional Networks

Fredrik Lindsten, Linköping University

Limitations of AR models

Auto-regressive model, AR(p):

$$y_t = a_1 y_{t-1} + \cdots + a_p y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2).$$

The AR model is linear in the parameters:

- ▲ Learning of parameters easy ⇔ linear regression
- ▼ Flexibility/ability to model complex temporal dependencies is limited
- ▼ Receptive field is just p time steps

From NAR to TCN

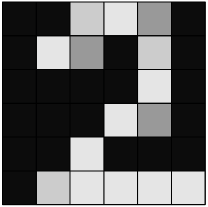
So far: Nonlinear AR = AR + MLP

- ▲ Flexibility increased.
- ▼ Receptive field is *still* just p steps

Alternative view: We can address this issue by using a convolutional network architecture!

Convolutional Neural Network

Image



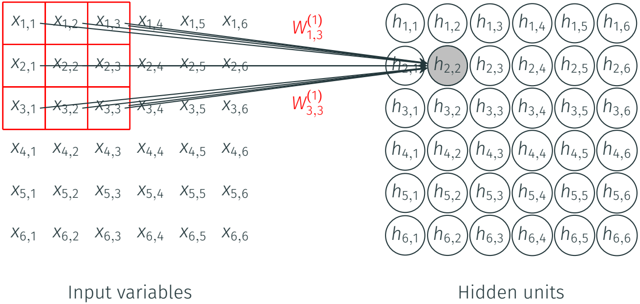
Data representation

0.0	0.0	0.8	0.9	0.6	0.0
0.0	0.9	0.6	0.0	0.8	0.0
0.0	0.0	0.0	0.0	0.9	0.0
0.0	0.0	0.0	0.9	0.6	0.0
0.0	0.0	0.9	0.0	0.0	0.0
0.0	0.8	0.9	0.9	0.9	0.9

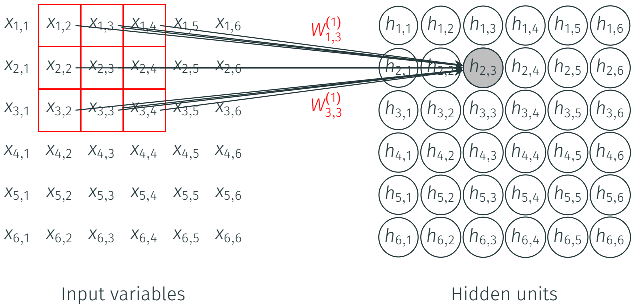
Input variables

$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,4}$	$x_{1,5}$	$x_{1,6}$
$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,4}$	$x_{2,5}$	$x_{2,6}$
$x_{3,1}$	$x_{3,2}$	$x_{3,3}$	$x_{3,4}$	$x_{3,5}$	$x_{3,6}$
$x_{4,1}$	$x_{4,2}$	$x_{4,3}$	$x_{4,4}$	$x_{4,5}$	$x_{4,6}$
$x_{5,1}$	$x_{5,2}$	$x_{5,3}$	$x_{5,4}$	$x_{5,5}$	$x_{5,6}$
$x_{6,1}$	$x_{6,2}$	$x_{6,3}$	$x_{6,4}$	$x_{6,5}$	$x_{6,6}$

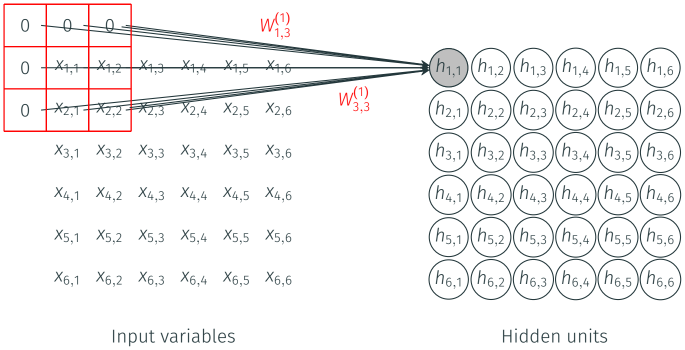
Convolutional Neural Network



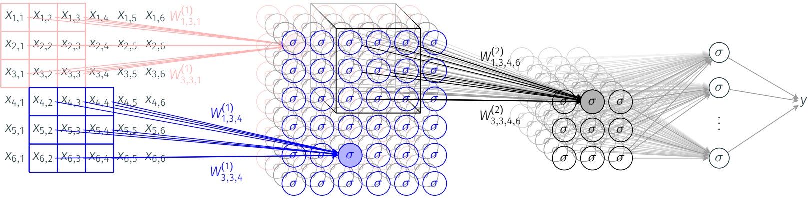
Convolutional Neural Network



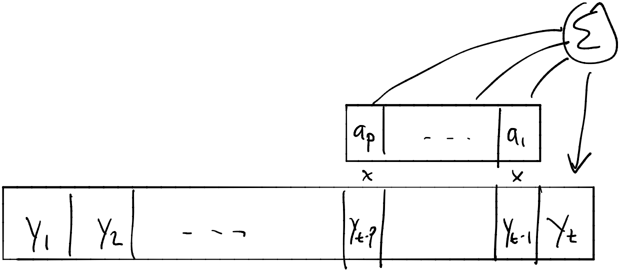
Convolutional Neural Network



Convolutional Neural Network



AR as convolution



$$Y_t = \sum_{j=1}^{p} a_j Y_{t-j} + \varepsilon_t$$

AR as convolution

This is a causal convolution between the process $y_{1:n}$ and the convolutional filter,

$$W = \begin{bmatrix} a_p & \dots & a_2 & a_1 \end{bmatrix}.$$

$\theta = (a_1 \;\; \dots \;\; a_p)^T$

Formally, we can write this as $y_t = \hat{y}_t + \varepsilon_t$ where:

$$\hat{y}_t = \color{red}{W}\color{black}{H_{t-1}}$$

$$H_{t-1} := \begin{bmatrix} y_{t-p} & \cdots & y_{t-2} & y_{t-1} \end{bmatrix}^{\top}$$

$\hat{y}_t = \theta^T \varphi_t$

$\varphi_t = [y_{t-1} \;\; \cdots \;\; y_{t-p}]^T$

$$y_{1:n} \longrightarrow \boxed{\text{Causal conv } \color{red}{W}} \longrightarrow \hat{y}_{1:n}$$

Temporal Convolutional Network

We can augment the AR model with a **second (hidden) convolutional layer.**

Let $h_t^{(0)} := y_t$ and introduce the "hidden signal"

nonlinear activation

"bias term"

$$h_t^{(1)} := \sigma\Big(\underbrace{W^{(1)} H_t^{(0)}}_{\text{linear AR}} + b^{(1)}\Big)$$

$$H_t^{(0)} := \left[h_{t-p-1}^{(0)} \; \cdots \; h_t^{(0)}\right]$$

Next, the output is given by

$$\hat{y}_{t+1} = W^{(2)} H_t^{(1)} + b^{(2)}$$

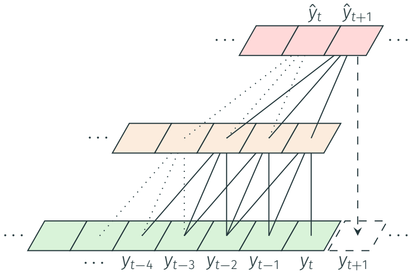
$$H_t^{(1)} = \left[h_{t-p-1}^{(1)} \; \cdots \; h_t^{(1)}\right]$$

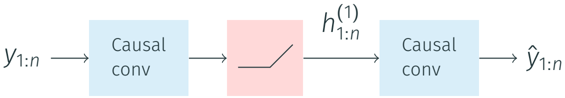
Temporal Convolutional Network

2-layer TCN:

$$h_t^{(0)} = y_t,$$
$$h_t^{(1)} = \sigma(W^{(1)}H_t^{(0)} + b^{(1)}),$$
$$y_{t+1} = W^{(2)}H_t^{(1)} + b^{(2)} + \varepsilon_{t+1},$$

with $\varepsilon_{t+1} \sim N(0, \sigma_\varepsilon^2)$.



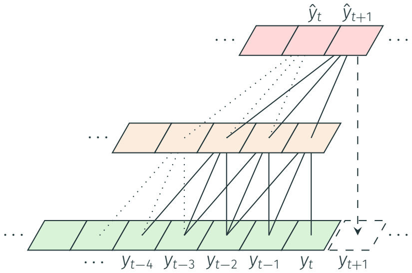


Temporal Convolutional Network

2-layer TCN:

$$h_t^{(0)} = y_t,$$
$$h_t^{(1)} = \sigma(W^{(1)} H_t^{(0)} + b^{(1)}),$$
$$y_{t+1} = W^{(2)} H_t^{(1)} + b^{(2)} + \varepsilon_{t+1},$$

with $\varepsilon_{t+1} \sim N(0, \sigma_\varepsilon^2)$.



Can extend to multiple layers of hidden signals, $h_t^{(2)}$, $h_t^{(3)}$, ...

- ▲ Multiple layers $\Rightarrow$ very flexible models
- ▲ Receptive field increases with depth...
- ▼ ...but only linearly