Deep Learning

< 13 : 20.06.2022

Semantic image segmentation

7-33

8. Recurrent Neural Network (RNN)

≅ Recursive neural network

Neural network with feedback.

Dense Network (Ch 4) } feed-forword
CNN (Ch 7) }

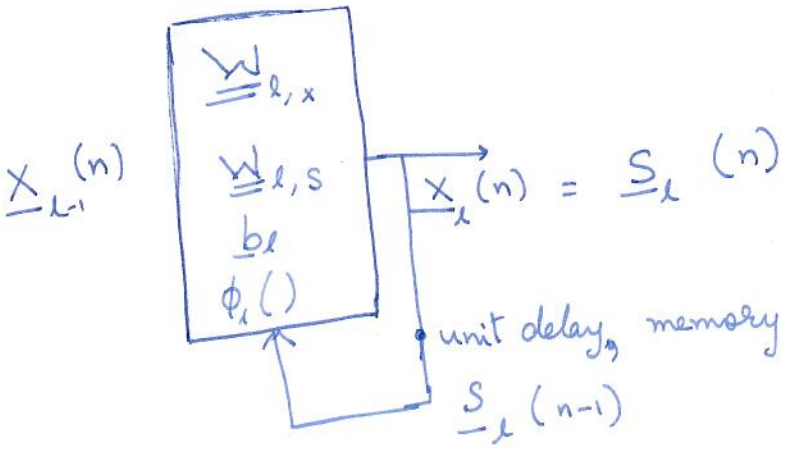
8-1, 8-2, 8-3, 8-4

9 temporal correlation- }

There is a relationship b/w the prev Sample and the next Sample. So we need the information of the prev samples to use and update weights etc in the next layers.

8.1 Recurrent layer and Recurrent neural network.

Recurrent layer l : M_l recurrent neurons



$n = 1, \ldots$ B : time index in 1 mini batch

↑ no of mini batches

$\underline{x}_{l-1}(n) \in \mathbb{R}^{M_{l-1}}$: input at time n

$$\underline{a}_l(n) = \underline{\underline{W}}_{l,x}\,\underline{x}_{l-1}(n) + \underbrace{\underline{\underline{W}}_{l,s}\,\underline{s}_l(n-1)}_{\text{feedback part}} + \underline{b}_l \in \mathbb{R}^{M_l}$$

$\underline{\underline{W}}_{l,x} \in \mathbb{R}^{M_l \times M_{l-1}}$

$\underline{\underline{W}}_{l,s} \in \mathbb{R}^{M_l \times M_l}$

} weight matrix

$\underline{b}_l \in \mathbb{R}^{M_l}$: bias

$\underline{x}_l(n) = \underline{s}_l(n) = \phi_l\left(\underline{a}_l(n)\right)$: output, memory state

$\phi_l(\,)$: activation function

$\underline{s}_l(0)$: initial value of one minibatch : often $\underline{0}$

$\hat{=}$ first order recursive non-linear filter for vector signals

8-5

Recurrent layer

$\underline{\underline{W}}_{1,s} = 0$

feedforward dense layer

$\phi_1(a) = a$

linear 1st order recursive filter

Often : •) $\underline{\underline{W}}_{1,x}$ full matrix $\hat{=}$ dense layer

•) $\underline{\underline{W}}_{1,s}$ diagonal $\hat{=}$ single neuron feedback

↓

There is no coupling between different feedbacks.

8-6, 8-7

Recurrent neural n/w (RNN): Contains atleast 1 recurrent layer.

8-8, 8-9

RNN can be widely used in time series analysis and audio, video etc processing. It all depends on the sliding window, delay that we want to give and weights on the basis of the problem formulation.

Training of RNN

- SGD : $\underline{\theta}^{t+1} = \underline{\theta}^t - \gamma^t \underline{\nabla} L(t; \underline{\theta})\Big|_{\underline{\theta}=\underline{\theta}^t}$ (minibatch t)
- Back propagation of gradients / derivatives through layers. old and known.

- for a parameter v in a recurrent layer → additional backpropagation through time (BPTT)

  $\hat{=}$ unfolded graph along the time axis

  $\hat{=}$ chain and product rule of a derivative

8-10, 8-11

We have gradient (vanishing) issues as there is a backpropagation. In RNN, we have vanishing gradient problem in b minibatches as well as across the time. So we have a ~~deeper~~ deeper problem to handle.