Summary of Chapter 3

Yuhang Zhang

Beijing Institute of Technology(BIT)

October 27, 2019

Outline

Outline

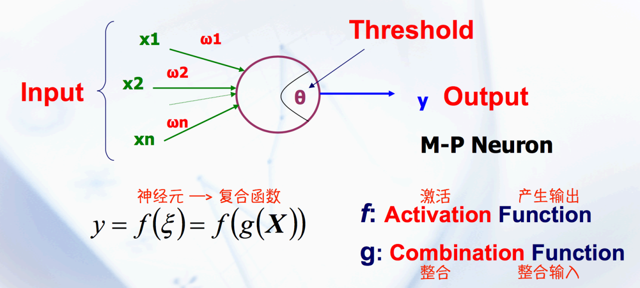
1 Overview

2 Feedforward Network
- Perceptron
- Deep Network
- Self-Organizing Feature Map

3 Feedback Network
- Hopfield Network
- Long Short-Term Memory

4 Review

Overview

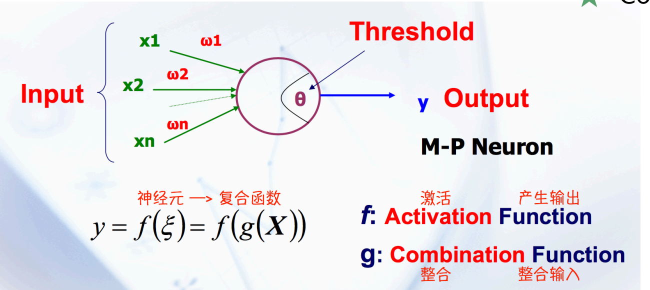


Neurons.pdf

Figure: Artificial Neuron

Overview

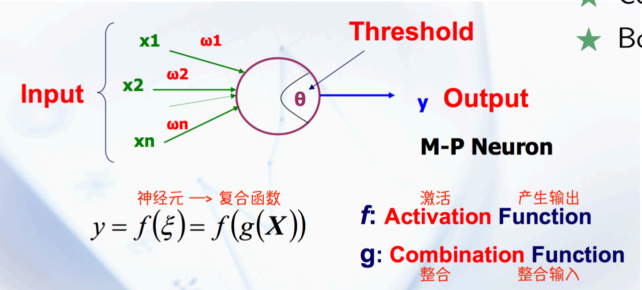
★ Connectionism



Neurons.pdf

Figure: Artificial Neuron

Overview

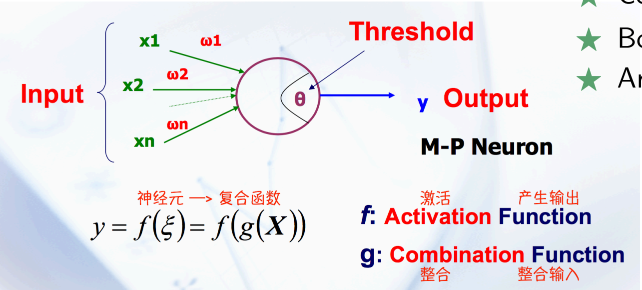


Neurons.pdf

Figure: Artificial Neuron

- ★ Connectionism
- ★ Bottom-Up AI

Overview

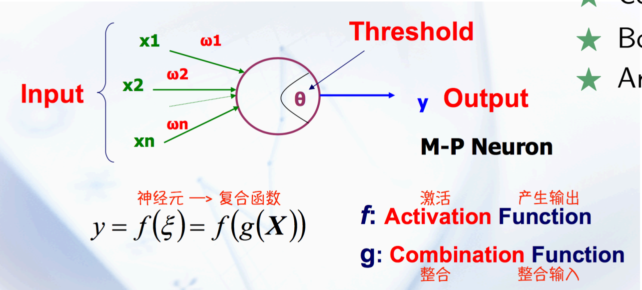


Neurons.pdf

Figure: Artificial Neuron

- ★ Connectionism
- ★ Bottom-Up AI
- ★ Artificial Neurons

Overview

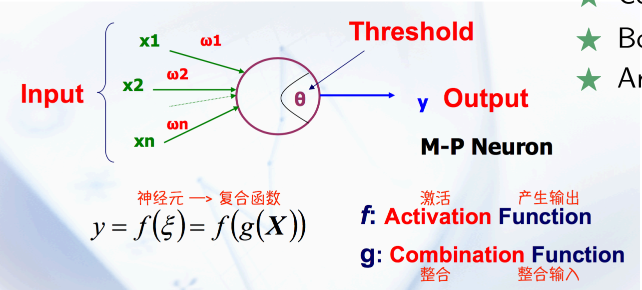


Neurons.pdf

Figure: Artificial Neuron

- ★ Connectionism
- ★ Bottom-Up AI
- ★ Artificial Neurons
  - ▶ Combination

Overview



Neurons.pdf

Figure: Artificial Neuron

- ★ Connectionism
- ★ Bottom-Up AI
- ★ Artificial Neurons
  - ▶ Combination
  - ▶ Activation

Overview

Main problems

- Architecture

Overview

Main problems

- Architecture
  - feedforward (static)

Overview

Main problems

- Architecture
  - feedforward (static)
  - feedback (dynamic)

Overview

Main problems

- Architecture
  - feedforward (static)
  - feedback (dynamic)
- Learning Approach

Overview

Main problems

- Architecture
  - feedforward (static)
  - feedback (dynamic)
- Learning Approach
  - Incremental vs. Batch

Overview

Main problems

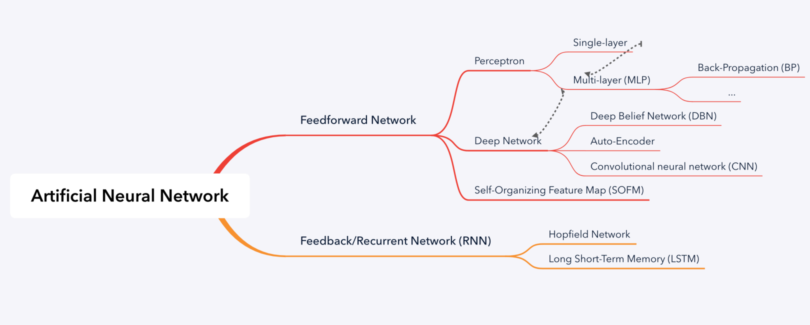
- Architecture
  - feedforward (static)
  - feedback (dynamic)
- Learning Approach
  - Incremental vs. Batch
  - Supervised vs. Unsupervised

Overview

Main problems

- Architecture
  - feedforward (static)
  - feedback (dynamic)
- Learning Approach
  - Incremental vs. Batch
  - Supervised vs. Unsupervised
  - Error Correction vs. Hebbian Learning vs. Competitive Learning

Overview


Artificial Neural Network
Feedforward Network
Perceptron
Single-layer
Multi-layer (MLP)
Back-Propagation (BP)
...
Deep Network
Deep Belief Network (DBN)
Auto-Encoder
Convolutional neural network (CNN)
Self-Organizing Feature Map (SOFM)
Feedback/Recurrent Network (RNN)
Hopfield Network
Long Short-Term Memory (LSTM)

Outline

Outline

1 Overview

2 Feedforward Network
- Perceptron
- Deep Network
- Self-Organizing Feature Map

3 Feedback Network
- Hopfield Network
- Long Short-Term Memory

4 Review

Single-layer vs. Multi-layer

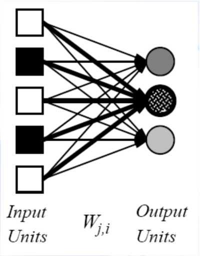


Figure: Single-layer Perceptron

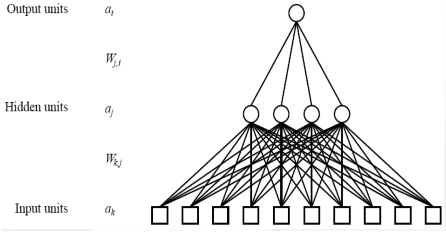


Figure: Multi-layer Perceptron

Single-layer vs. Multi-layer

- Single-layer : only linear functions

Single-layer vs. Multi-layer

- Single-layer : only linear functions
- Multi-layer : non-linear operations
  hidden layer added
  - BP : Use sigmoid activation function

Multi-layer ▶ BP Network

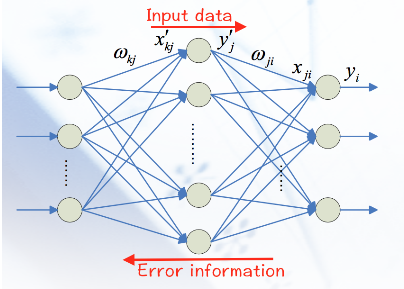


Figure: BP Structure

Multi-layer ▶ BP Network

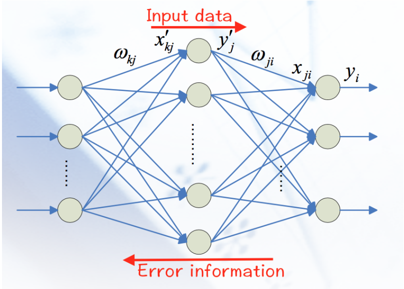


Figure: BP Structure

- objectiveness (measuring criterion)
  ▶ *MSE* :

$$e(\omega) = \frac{1}{2}\sum_{i=1}^{n}[d_i - y_i]^2$$

Multi-layer ▶ BP Network

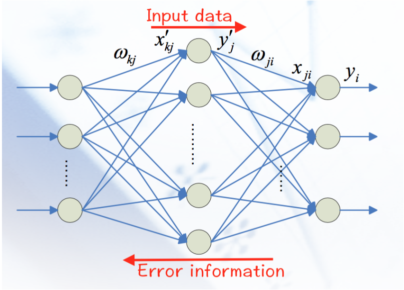


Figure: BP Structure

- objectiveness (measuring criterion)
  - ▶ *MSE* :

$$e(\omega) = \frac{1}{2}\sum_{i=1}^{n}[d_i - y_i]^2$$

- optimization
  - ▶ *Gradient Descent* :

$$\omega = \omega - \eta\frac{\partial e}{\partial \omega}$$

Multi-layer ▶ BP Network

Problems : (especially for deeper layers)

- Diffusion of Gradient → *early layers not trained well*

Multi-layer ▶ BP Network

Problems : (especially for deeper layers)

- Diffusion of Gradient → *early layers not trained well*
- Supervised learning → *not enough labeled data*

Outline

Deep Belifef Network (DBN)

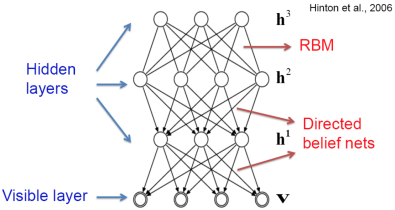


Figure: DBN Structure

Deep Belifef Network (DBN)

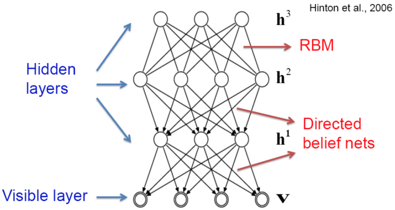


Figure: DBN Structure

- Key points :

Deep Belifef Network (DBN)

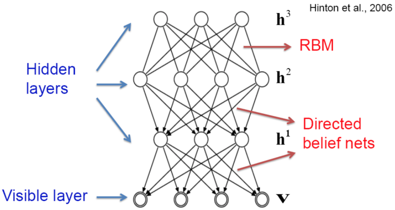


Figure: DBN Structure

- Key points :
  - Unsupervised pre-learning

Deep Belifef Network (DBN)

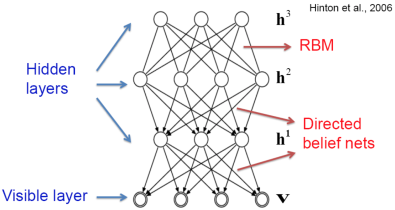


Figure: DBN Structure

- Key points :
  - Unsupervised pre-learning
  - Greedy layer-wise training

Deep Belifef Network (DBN)

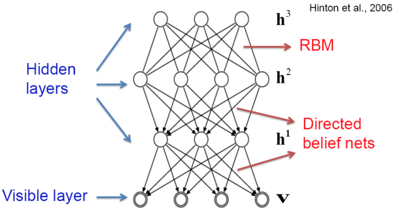


Figure: DBN Structure

- Key points :
  - Unsupervised pre-learning
  - Greedy layer-wise training
  - Use softmax for output layer

Deep Belifef Network (DBN)

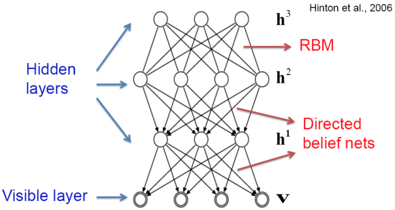


Figure: DBN Structure

- Key points :
  - Unsupervised pre-learning
  - Greedy layer-wise training
  - Use softmax for output layer
- objectiveness : MLE

Deep Belifef Network (DBN)

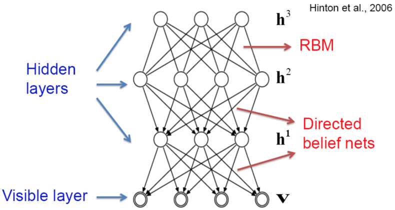


Figure: DBN Structure

- Key points :
  - Unsupervised pre-learning
  - Greedy layer-wise training
  - Use softmax for output layer
- objectiveness : MLE
- optimization : Gradient Ascent

DBN ▶ Bases

Statistical distribution

- random variable for each neuron

$$P(v, h^1, h^2, \ldots, h^l) = P(v|h^1)P(h^1|h^2)\ldots P(h^{l-2}|h^{l-1})P(h^{l-1}, h^l)$$

$$P(h^i|h^{i+1}) = \prod_{j=1}^{n^i} P(h^i_j|h^{i+1})$$

DBN ▶ Bases

Statistical distribution

- random variable for each neuron

$$P(v, h^1, h^2, \dots, h^l) = P(v|h^1)P(h^1|h^2)\dots P(h^{l-2}|h^{l-1})P(h^{l-1}, h^l)$$

$$P(h^i|h^{i+1}) = \prod_{j=1}^{n^i} P(h^i_j|h^{i+1})$$

- RBM for distribution between layers

DBN ▶ Training

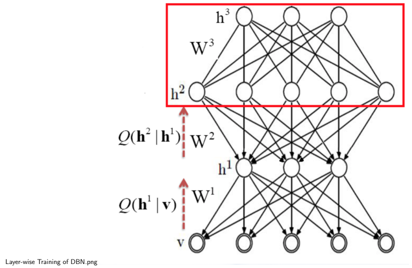


Figure: Greedy Layer-wise Training

DBN ▶ Training

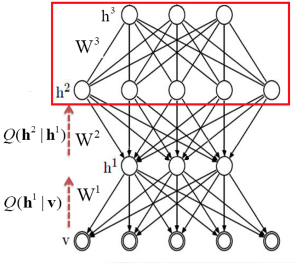
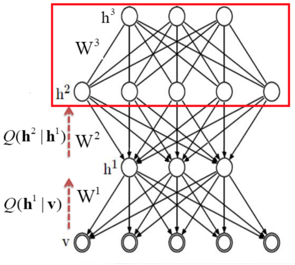


Figure: Greedy Layer-wise Training

- Greedy Layer-wise Training

DBN ▶ Training

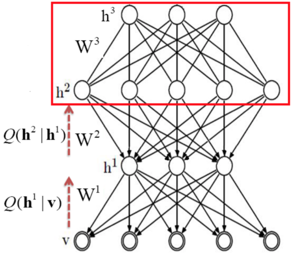


Layer-wise Training of DBN.png

Figure: Greedy Layer-wise Training

- Greedy Layer-wise Training
  - i construct a RBM with v and h^1

DBN ▶ Training

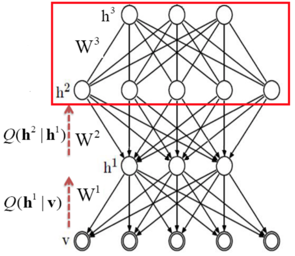


Layer-wise Training of DBN.png

Figure: Greedy Layer-wise Training

- Greedy Layer-wise Training
  - i construct a RBM with v and h^1
  - ii form a new RBM with h^1 and h^2

DBN ▶ Training

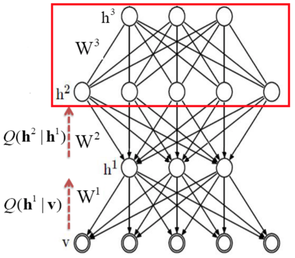


Layer-wise Training of DBN.png

Figure: Greedy Layer-wise Training

- Greedy Layer-wise Training
  - i construct a RBM with v and h^1
  - ii form a new RBM with h^1 and h^2
  - iii continued. . .

DBN ▶ Training



Layer-wise Training of DBN.png

Figure: Greedy Layer-wise Training

- Greedy Layer-wise Training
  - i construct a RBM with v and h^1
  - ii form a new RBM with h^1 and h^2
  - iii continued...
- Fine-tuning

DBN ▶ Training

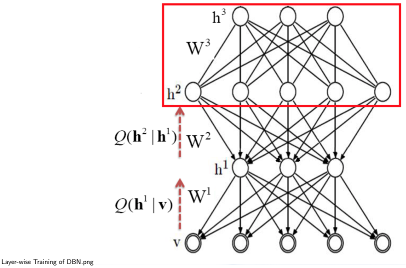
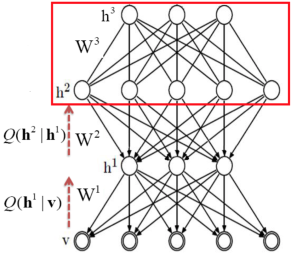


Figure: Greedy Layer-wise Training

- Greedy Layer-wise Training
  - i construct a RBM with v and h^1
  - ii form a new RBM with h^1 and h^2
  - iii continued...
- Fine-tuning
  - Generative (unsupervised): Up-down algorithm

DBN ▶ Training



Layer-wise Training of DBN.png

Figure: Greedy Layer-wise Training

- Greedy Layer-wise Training
  - i construct a RBM with v and h^1
  - ii form a new RBM with h^1 and h^2
  - iii continued. . .
- Fine-tuning
  - Generative (unsupervised): Up-down algorithm
  - Discriminative (supervised): Back propagation

Auto-encoder

- Bases

Auto-encoder

- Bases
  - Encoder

Auto-encoder

- Bases
    - Encoder
    - Decoder

Auto-encoder

- Bases
  - Encoder
  - Decoder
- AutoEncoder vs. RBM

Auto-encoder

- Bases
  - Encoder
  - Decoder
- AutoEncoder vs. RBM
  - AutoEncoder : reconstruct each training data

Auto-encoder

- Bases
  - Encoder
  - Decoder
- AutoEncoder vs. RBM
  - AutoEncoder : reconstruct each training data
  - RBM : reconstruct distribution of data

Outline

SOFM

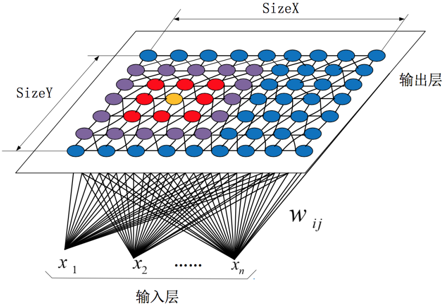


Figure: SOFM structure

SOFM

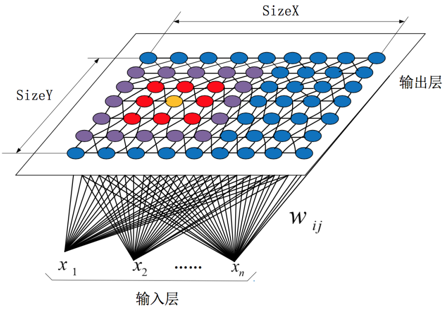


Figure: SOFM structure

- Fundamental idea
  Dimensionality reduction

SOFM

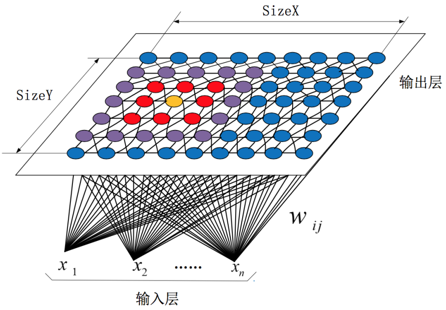


Figure: SOFM structure

- Fundamental idea
  - Dimensionality reduction
- Principles

SOFM

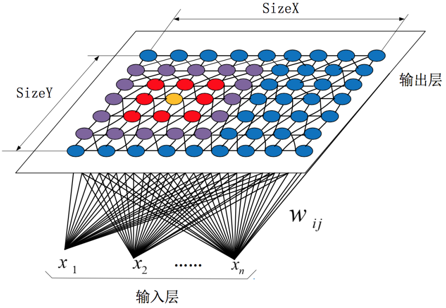


Figure: SOFM structure

- Fundamental idea
  - Dimensionality reduction
- Principles
  - Self-reinforcing

SOFM

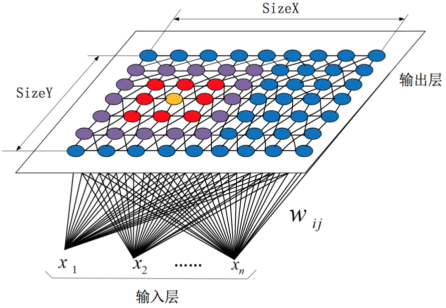


Figure: SOFM structure

- Fundamental idea
  - Dimensionality reduction
- Principles
  - Self-reinforcing
  - Competition

SOFM

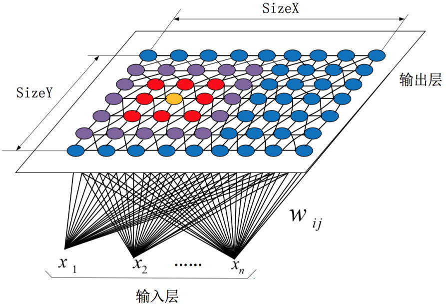


Figure: SOFM structure

- Fundamental idea
  - Dimensionality reduction
- Principles
  - Self-reinforcing
  - Competition
  - Cooperation

Outline

Feedback (Recurrent) Network (RNN)

(1) Stable feedback network

Feedback (Recurrent) Network (RNN)

(1) Stable feedback network

- Hopfield Network

Feedback (Recurrent) Network (RNN)

(1) Stable feedback network
- Hopfield Network

(2) Sequential feedback network $\leftrightsquigarrow$ Deep Network

Feedback (Recurrent) Network (RNN)

(1) Stable feedback network

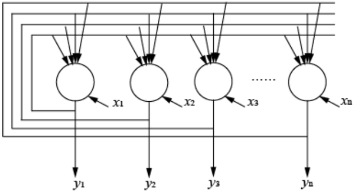
- Hopfield Network

(2) Sequential feedback network $\leftrightsquigarrow$ Deep Network

- LSTM

Outline

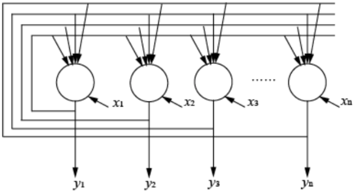
Bases



Network.png

Figure: Hopfield Network Structure

Bases

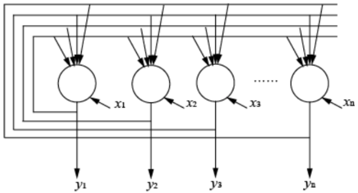


Network.png

Figure: Hopfield Network Structure

- Full link network

Bases

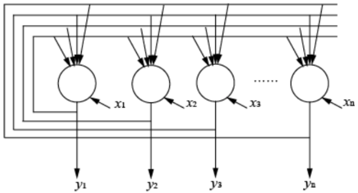


Network.png

Figure: Hopfield Network Structure

- Full link network
- Training method : change network state

Bases

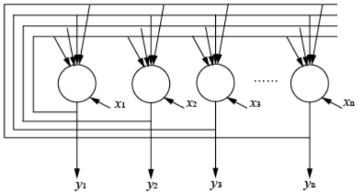


Network.png

Figure: Hopfield Network Structure

- Full link network
- Training method : change network state
  - asynchronous

Bases



Network.png

Figure: Hopfield Network Structure

- Full link network
- Training method : change network state
  - asynchronous
  - synchronous

Energy Function

- $$E = -\frac{1}{2}\sum_{i=0}^{n}\sum_{j=0}^{n}\omega_{ij}s_i s_j \quad (-\sum_{i=1}^{n} I_i s_i)$$

Energy Function

- $$E = -\frac{1}{2}\sum_{i=0}^{n}\sum_{j=0}^{n}\omega_{ij}s_i s_j \quad (-\sum_{i=1}^{n} I_i s_i)$$

- Stability condition

$$\omega_{ij} = \begin{cases} \omega_{ji} & i \neq j \\ 0 & i = j \end{cases}$$

Application

- Associative memory (CRM)
  - each local minimum represents one memorized data

Application

- Associative memory (CRM)
  - each local minimum represents one memorized data
- Combinatorial Optimization
  - TSP

Outline

Back-Propagation Through Time (BPTT)

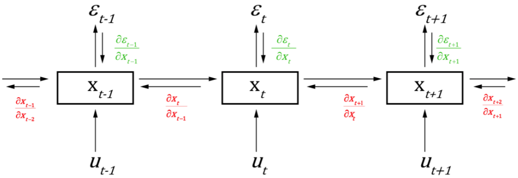


Figure: BPTT Structure

Back-Propagation Through Time (BPTT)

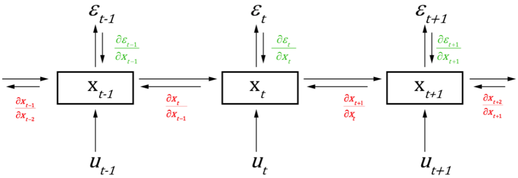


Figure: BPTT Structure

- Problems

Back-Propagation Through Time (BPTT)

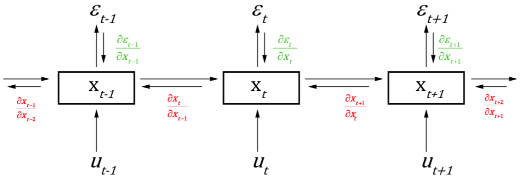


Figure: BPTT Structure

- Problems
  - Gradient Vanishing

Back-Propagation Through Time (BPTT)

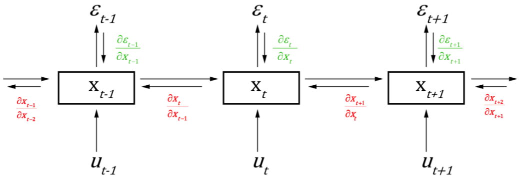
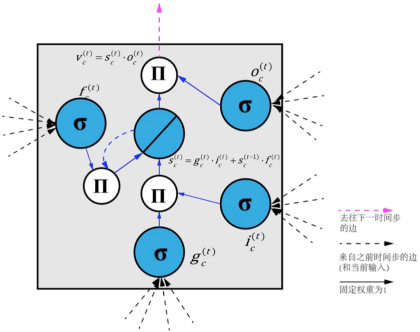


Figure: BPTT Structure

- Problems
  - Gradient Vanishing
  - Gradient Exploding

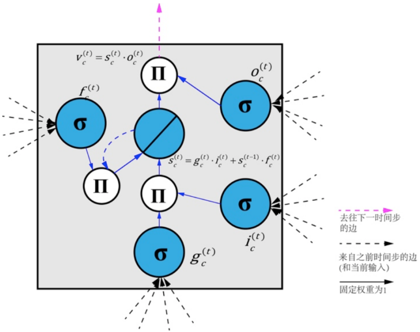
LSTM Memory Cell



cell.png

Figure: LSTM Memory Cell

LSTM Memory Cell

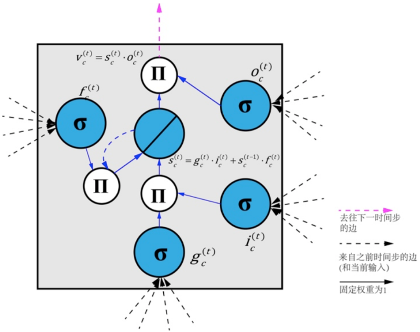


cell.png

Figure: LSTM Memory Cell

- Internal state (constant error carousel): kernel, linear activation function, self-cycling edge(weight fixed to 1)

LSTM Memory Cell

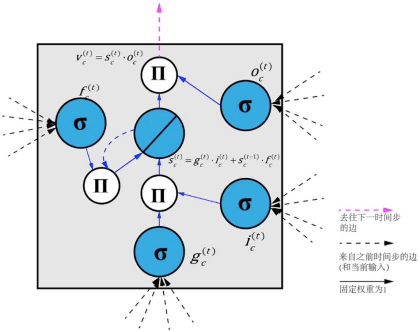


cell.png

Figure: LSTM Memory Cell

- Internal state (constant error carousel): kernel, linear activation function, self-cycling edge(weight fixed to 1)
- Input node : receive earlier signals

LSTM Memory Cell



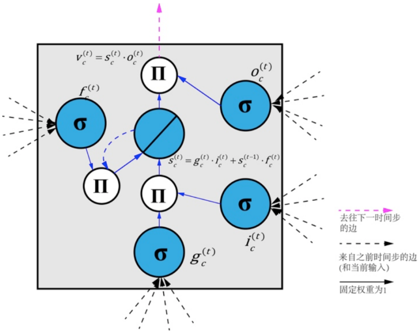
cell.png

Figure: LSTM Memory Cell

- Internal state (constant error carousel): kernel, linear activation function, self-cycling edge(weight fixed to 1)
- Input node : receive earlier signals
- Input gate : same as Input node, control result of input

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

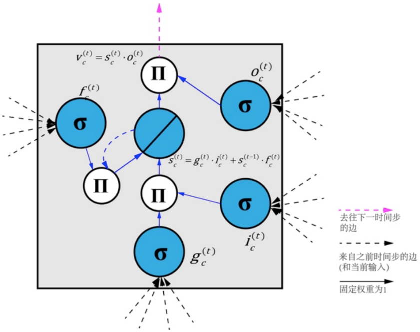
LSTM Memory Cell



cell.png

Figure: LSTM Memory Cell

LSTM Memory Cell

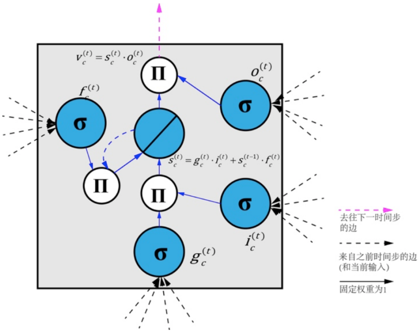


cell.png

Figure: LSTM Memory Cell

- Forget gate : eliminate previous internal state value

LSTM Memory Cell

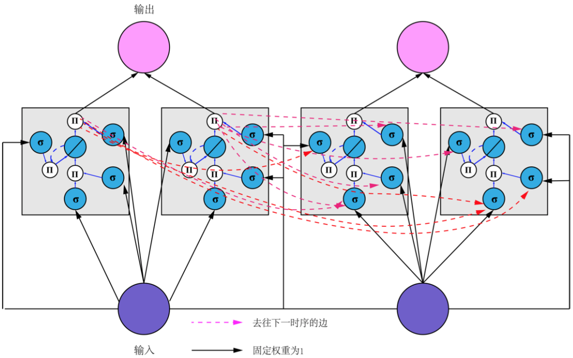


cell.png

Figure: LSTM Memory Cell

- Forget gate :
  eliminate previous internal state value
- Output gate :
  control result of output

LSTM Structure



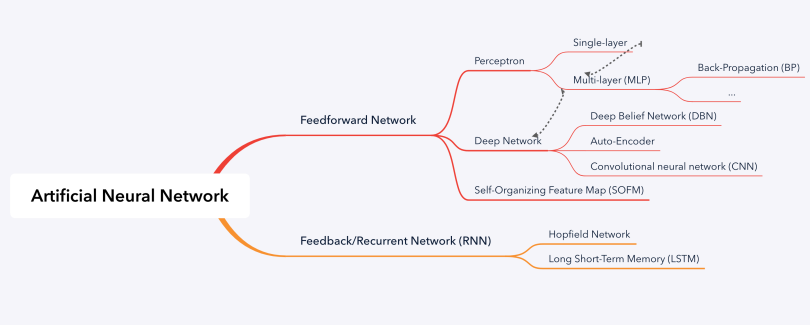
structure.png

Figure: LSTM structure

Outline

北京理工大學
BEIJING INSTITUTE OF TECHNOLOGY

Review


Artificial Neural Network
Feedforward Network
Perceptron
Single-layer
Multi-layer (MLP)
Back-Propagation (BP)
...
Deep Network
Deep Belief Network (DBN)
Auto-Encoder
Convolutional neural network (CNN)
Self-Organizing Feature Map (SOFM)
Feedback/Recurrent Network (RNN)
Hopfield Network
Long Short-Term Memory (LSTM)

Thank you

Thank you for listening!