

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΕΡΓΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Αναδρομικά Νευρωνικά Δίκτυα - Recurrent Neural Networks (RNN)

Μοντέλα Συσχετισμένης Μνήμης, Δίκτυα Hopfield
Ακολουθιακά Μοντέλα Δεδομένων & RNN
Δίκτυα Long Short-Term Memory (LSTM)

καθ. Βασίλης Μάγκλαρης
maglaris@netmode.ntua.gr
www.netmode.ntua.gr

Αίθουσα 002, Νέα Κτίρια ΣΗΜΜΥ
Τρίτη 30/5/2023

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Associative Memory ή Content Addressable Memory (CAM)

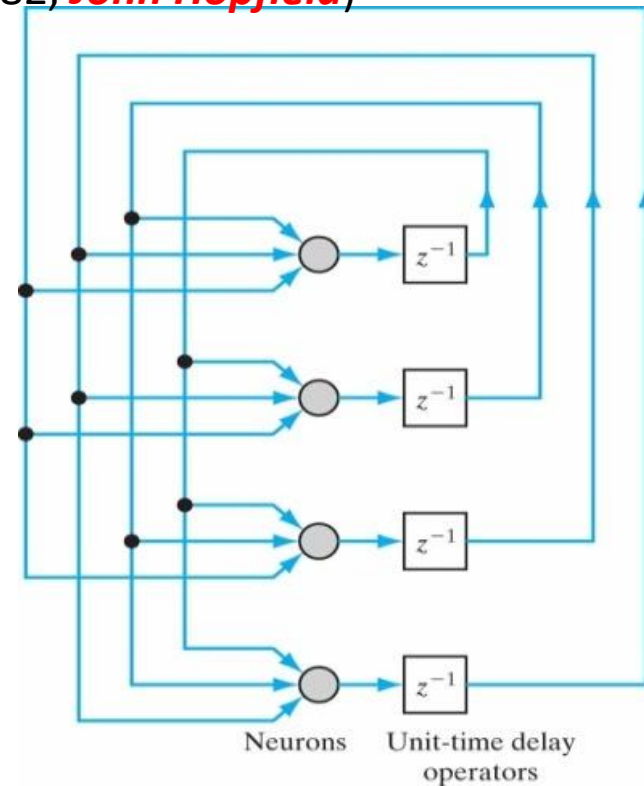
<https://www.doc.ic.ac.uk/~ae/papers/Hopfield-networks-15.pdf>

- Η οργάνωση ηλεκτρονικής μνήμης:
 - Παραδοσιακό χωρικό μοντέλο (**address based**)
 - Εξέλιξη μοντέλου αποθήκευσης προτύπων: Συσχετισμένη Μνήμη (**Associative Memory**) ή Μνήμη Διεύθυνσης Περιεχομένου (**Content Addressable Memory – CAM**)
- Η οργάνωση της ανθρώπινης μνήμης: Αποθήκευση προτύπων με βάση συσχετίσεις χαρακτηριστικών (**associative memory**) και όχι σε συγκεκριμένα σημεία του εγκεφάλου
- Νευροφυσιολογικός τρόπος μάθησης: Ρύθμιση συσχετίσεων (συνάψεις) νευρώνων και αποθήκευση συναπτικών βαρών, κατά προσέγγιση σύμφωνα με τον κανόνα του **Hebb**: Ενισχυμένα συναπτικά βάρη μεταξύ ταυτόχρονα ενεργών νευρώνων, μηδενικές σχέσεις μεταξύ μη συγχρονισμένων νευρώνων
- Αντίστοιχο μοντέλο μηχανικής μάθησης: Νευρωνικό δίκτυο **Hopfield** (**John Hopfield**, 1982: <https://www.pnas.org/doi/pdf/10.1073/pnas.79.8.2554>) – λειτουργεί σαν **attractor** που αντιστοιχεί δειγματικά στοιχεία σε αποθηκευμένα πρότυπα με ελαχιστοποίηση της «ενέργειας». Μαθαίνει και αποθηκεύει πρότυπα με επιβλεπόμενη μάθηση (**supervised learning**) με κανόνες **Hebb**

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Ντετερμινιστικό Νευρωνικό Δίκτυο Hopfield (1982, John Hopfield)

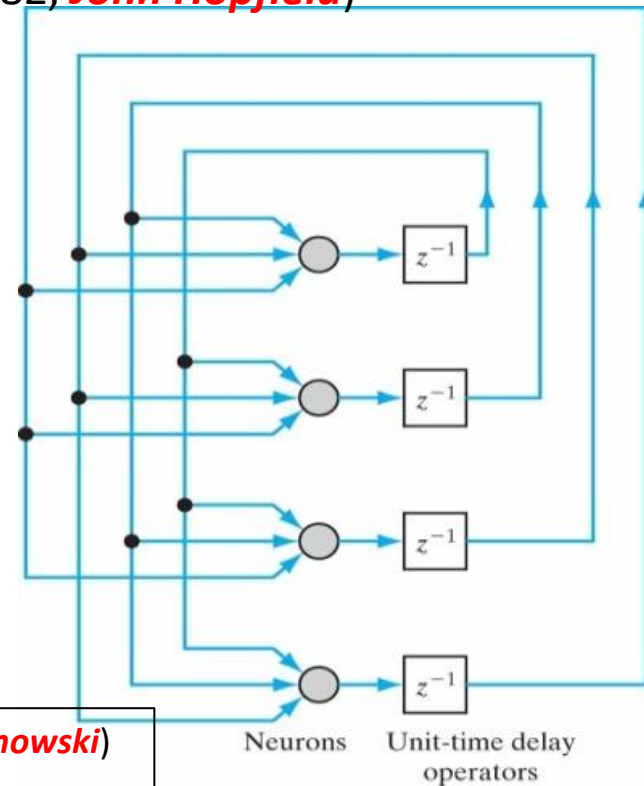
- Νευρώνες **deterministic** καταστάσεων ± 1 , **αναδρομικές** συμμετρικές συνάψεις $i \rightarrow j: w_{ji} = w_{ij}, w_{ii} = 0$, **threshold activation** ± 1 (sgn) & ισορροπία σύμφωνη με κανόνα **Hebb**
- Ρύθμιση w_{ji} μέσω ελαχιστοποίησης μέσης τετραγωνικής απόκλισης εξόδου από **labels** στοιχείων δείγματος μάθησης (**supervised learning**)
- Δειγματικό στοιχείο εισόδου συγκλίνει στην έξοδο σε σταθερά σημεία (**fixed points**), **target patterns** που έχουν αποθηκευτεί κατά τη μάθηση μέσω των w_{ji}
- Πρώτη εφαρμογή για αναγνώριση προτύπων (π.χ. διάκριση χειρόγραφων αριθμών από παραμορφωμένα στοιχεία εισόδου με βάση δείγμα μάθησης **MNIST**)



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Ντετερμινιστικό Νευρωνικό Δίκτυο Hopfield (1982, John Hopfield)

- Νευρώνες **deterministic** καταστάσεων ± 1 , **αναδρομικές** συμμετρικές συνάψεις $i \rightarrow j: w_{ji} = w_{ij}, w_{ii} = 0$, **threshold activation** ± 1 (sgn) & ισορροπία σύμφωνη με κανόνα **Hebb**
- Ρύθμιση w_{ji} μέσω ελαχιστοποίησης μέσης τετραγωνικής απόκλισης εξόδου από **labels** στοιχείων δείγματος μάθησης (**supervised learning**)
- Δειγματικό στοιχείο εισόδου συγκλίνει στην έξοδο σε σταθερά σημεία (**fixed points**), **target patterns** που έχουν αποθηκευτεί κατά τη μάθηση μέσω των w_{ji}
- Πρώτη εφαρμογή για αναγνώριση προτύπων (π.χ. διάκριση χειρόγραφων αριθμών από παραμορφωμένα στοιχεία εισόδου με βάση δείγμα μάθησης **MNIST**)



Στοχαστική Επέκταση: Boltzmann Machine (1985, Geoffrey Hinton & Terry Sejnowski)

Αποτελείται από K Visible και L Hidden Neurons με **συμμετρικές συνάψεις** $w_{ji} = w_{ij}$,

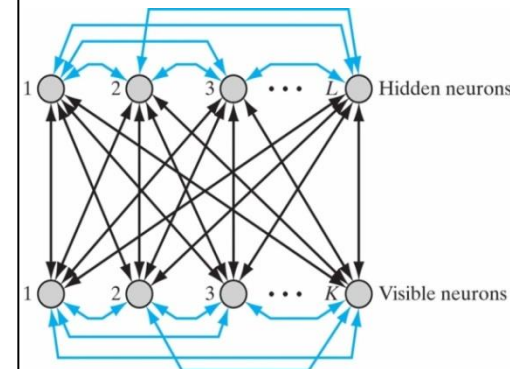
$w_{ii} = 0$, εν δυνάμει μεταξύ όλων των νευρώνων της **Boltzmann Machine** (BM)

Αποτελεί εξέλιξη του αναδρομικού δικτύου **Hopfield** με νευρώνες σε **δυναμικές**

καταστάσεις ± 1 σύμφωνα με ορισμένες **πιθανότητες** (**Stochastic Recurrent Network**)

Το δίκτυο συγκλίνει με **unsupervised learning** σε ισορροπία **Markov Random Field**:

- Δυναμικά παραδείγματα **μάθησης** εισάγονται στα **Visible Nodes** και με **gradient ascent** ρυθμίζονται τα συναπτικά βάρη & οι **τελικές καταστάσεις όλων** των νευρώνων
- Δυναμικά στοιχεία **test** εισάγονται στα **Visible Nodes** και η **BM** αναπαράγει παρεμφερή δειγματικά στοιχεία (πάλι στα **Visible Nodes**) συμβατά με τις στατιστικές ιδιότητες του **δείγματος μάθησης** (**sampling, generative model**)
- Απλοποίηση της **BM** σε **Restricted Boltzmann Machines** (RBM) με μηδενικές συνάψεις μεταξύ νευρώνων του ίδιου επιπέδου (**visible & hidden**)



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Δίκτυο Hopfield & Νευρωνική Μνήμη (1/2)

<https://towardsdatascience.com/hopfield-networks-neural-memory-machines-4c94be821073>

- Νευρώνες **deterministic** δυαδικών καταστάσεων s_i , **αναδρομικές** συμμετρικές συνάψεις $i \rightarrow j$: $w_{ji} = w_{ij}$, $w_{ii} = 0$, **binary threshold activation** ± 1 (sgn)
- Έξοδος νευρώνα i : $y_i = \text{sgn}\{\sum_j y_j w_{ij} + b_j\}$, b_j βάρος πόλωσης (**bias**)
- Κατάσταση νευρώνα i : $s_i = \begin{cases} +1, & y_i > 0 \\ -1, & y_i < 0 \end{cases}$

1^η Φάση: Επιβλεπόμενη Μάθηση – Αποθήκευση (**Storage**) Προτύπων

- Αποθήκευση ενός προτύπου (**target pattern**) σύμφωνα με τον κανόνα **Hebb** $w_{ij} = s_i s_j$
- Για ταυτόχρονη αποθήκευση M **target patterns** $\mu = 1, 2, \dots, M$ οι καταστάσεις των νευρώνων είναι διανύσματα με στοιχεία $s_i^{(\mu)}$ και τα συναπτικά βάρη προκύπτουν από τον γενικευμένο κανόνα του **Hebb** (**generalized Hebbian rule**) ως

$$w_{ij} = \frac{1}{M} \sum_{\mu=1}^M s_i^{(\mu)} s_j^{(\mu)}$$

- Εναλλακτικά, υπολογισμός των w_{ij} με **back-propagation supervised learning** απαιτεί μεγάλα δείγματα μάθησης

2^η Φάση: Διαδικασία Αναζήτησης (**Retrieval**) σε είσοδο Νέου Στοιχείου (**Probe**)

Επαναλήψεις σε διακριτά χρονικά σημεία για ανανέωση (**update**) καταστάσεων ανά νευρώνα με τυχαία σειρά βάσει της σχέσης $y_i = \text{sgn}\{\sum_j y_j w_{ij} + b_j\}$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Δίκτυο Hopfield & Νευρωνική Μνήμη (2/2)

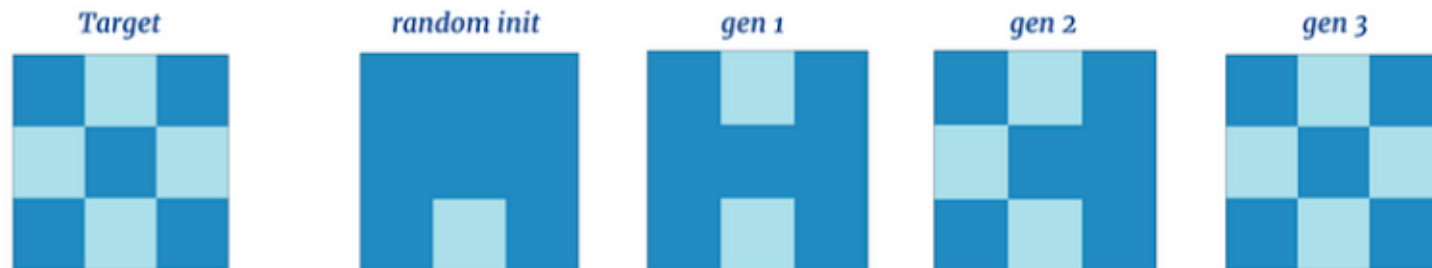
<https://towardsdatascience.com/hopfield-networks-neural-memory-machines-4c94be821073>

Παράδειγμα Εφαρμογής Δικτύου Hopfield για Αναγνώριση Προτύπου

- Πρότυπο $[0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0]^T$, διάνυσμα 9 δυαδικών ψηφίων $\{0,1\}$
- Μετασχηματισμός προτύπου σε δισδιάστατη εικόνα:



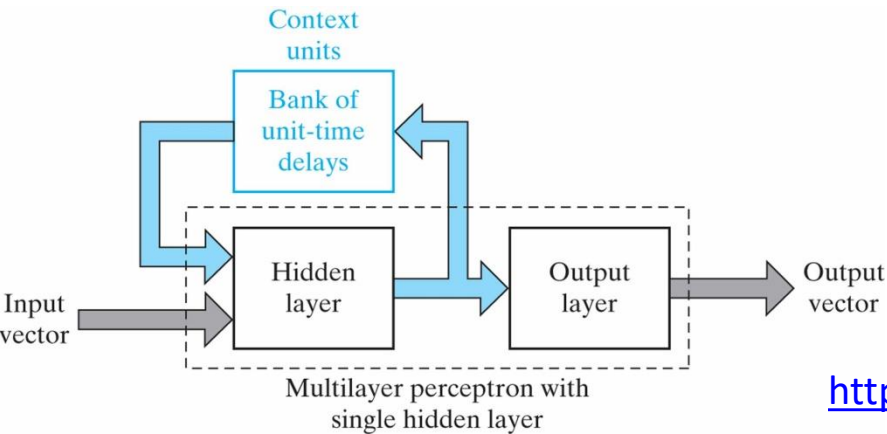
- Αποθήκευση σε δίκτυο **Hopfield** 9 νευρώνων με δυαδικές καταστάσεις $\{0,1\}$ αντί $\{-1, +1\}$
- Τα δίκτυα **Hopfield** αναφέρονται και σαν **attractor networks** γιατί νέα δειγματικά στοιχεία **test** έλκονται προς καταστάσεις ισορροπίας που αντιστοιχούν σε αποθηκευμένο πρότυπο. Η διαδικασία σύγκλισης προς αποθηκευμένο πρότυπο (**target pattern**) ξεκινά από τυχαία αρχική κατάσταση και συγκλίνει σε μόλις 3 **updates**:



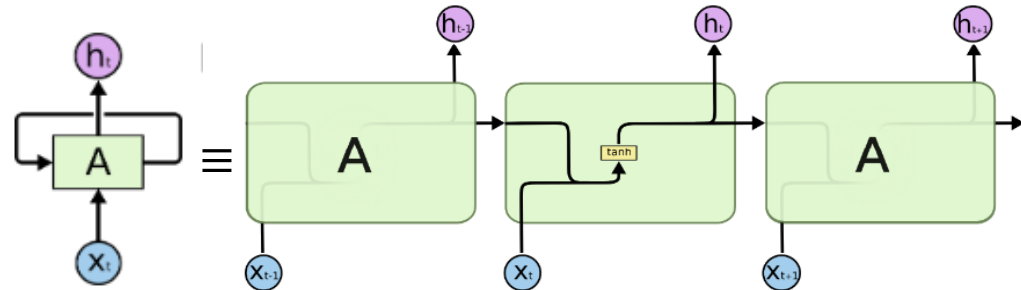
ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Αναδρομικά Νευρωνικά Δίκτυα - Recurrent Neural Networks (RNN) (1/2)

Simple Recurrent Network (SRN)



Ισοδύναμο Ανάπτυγμα SRN σε Μονοστρωματικά Νευρωνικά Δίκτυα



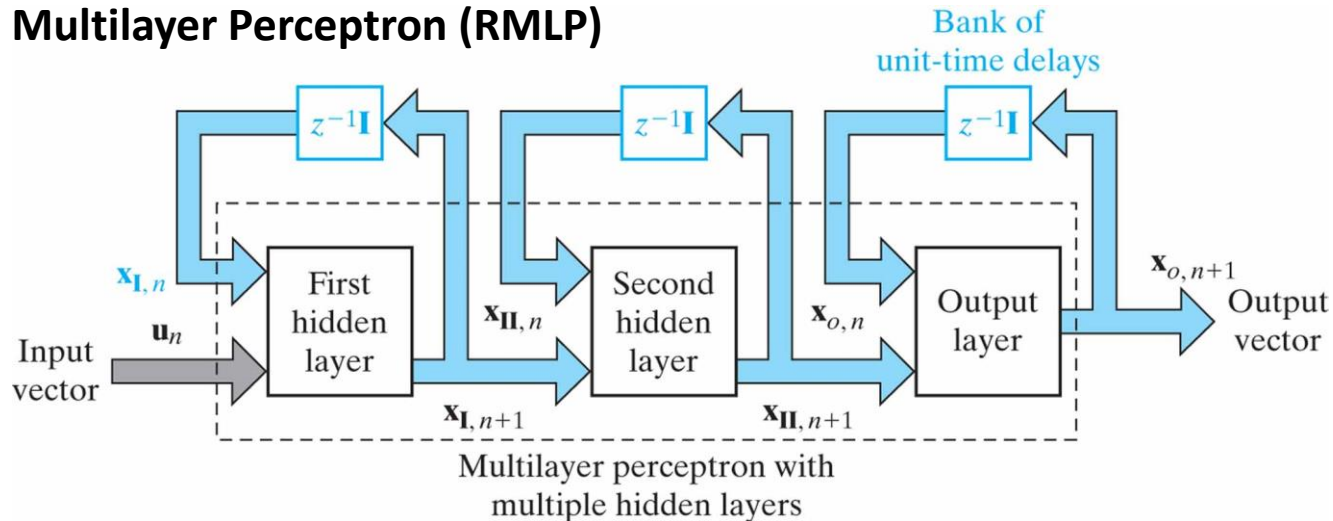
<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

A : Κρυφό στρώμα Νευρωνικού Δικτύου *Perceptron* με *Activation Function* \tanh και έξοδο $\in [-1, +1]$

x_t : Είσοδος στοιχείου στο A τη χρονική στιγμή t

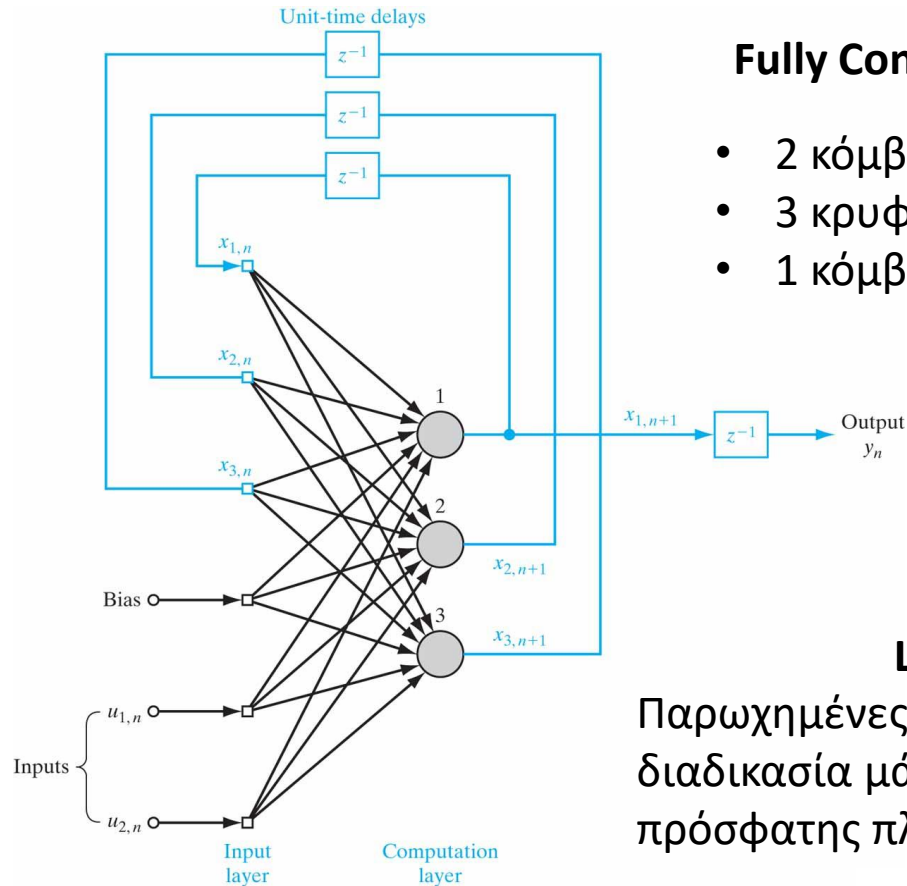
h_t : Κρυφή έξοδος του A τη χρονική στιγμή t

Recurrent Multilayer Perceptron (RMLP)



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Αναδρομικά Νευρωνικά Δίκτυα - Recurrent Neural Networks RNN (2/2)

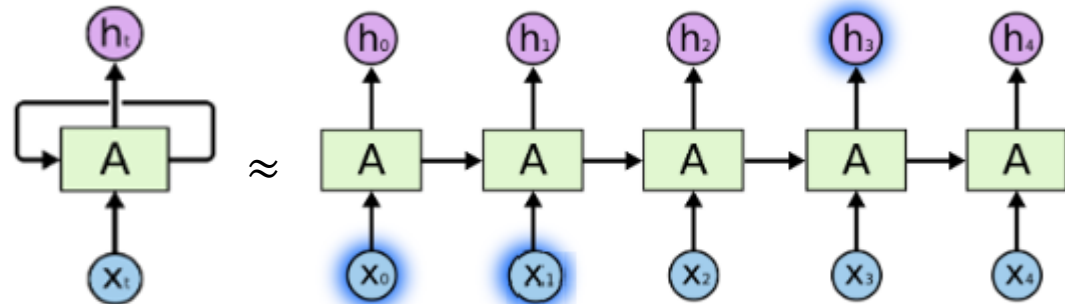


Fully Connected SRN

- 2 κόμβοι εισόδων
- 3 κρυφοί νευρώνες
- 1 κόμβος εξόδου

Long-Term Dependencies

Παρωχημένες παλιές εισοδοι δυσκολεύουν τη διαδικασία μάθησης – ανάγκη για διατήρηση **μόνο** πρόσφατης πληροφορίας (π.χ. 5 χρονικές περιοδοι)

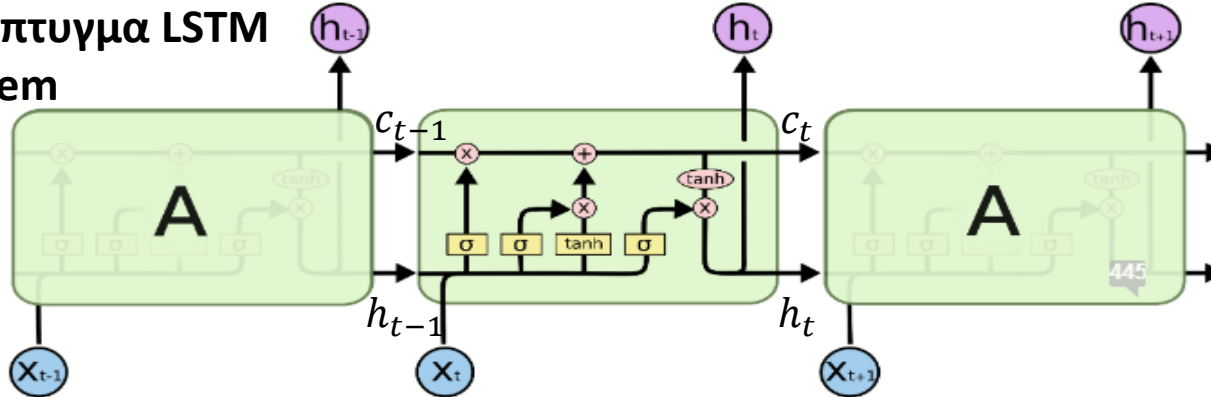


ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Long Short-Term Memory (LSTM) (1/3)

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

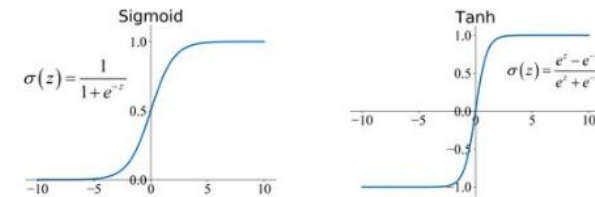
Ισοδύναμο Ανάπτυγμα LSTM με Cells in Tandem



The repeating module in an LSTM contains four interacting layers.

- x_t : Input vector of cell t
- c_t : **Cell State** vector ($c_{t-1} \rightarrow c_t$ αναδρομικά, ανάλογα με τα h_{t-1}, x_t και αν το επιτρέπουν οι θύρες ελέγχου - **Gates**)
- h_t : Hidden output vector of cell t
- σ : Sigmoid activation function $\sim \{0,1\}$ σε 3 **Gates** (**forget, input, output**) που ελέγχουν τη ροή πληροφορίας (π.χ. αποκόπτον παρωχημένα στοιχεία)
- \tanh : Hyperbolic tan activation function για ρύθμιση **Cell State** σε τιμές $\in [-1,1]$

Βασική Θύρα Ελέγχου: **Forget Gate** για αποκοπή παρωχημένων στοιχείων



Δομή Control Gate **Hadamard** (point-wise) product

$$v = [v_1 \dots v_k]^T$$

$$v \circ b = [v_1 \times b_1 \dots v_k \times b_k]^T$$

$$b = [b_1 \dots b_k]^T = [\sigma(a_1) \dots \sigma(a_k)]^T$$

$$a = [a_1 \dots a_k]^T$$

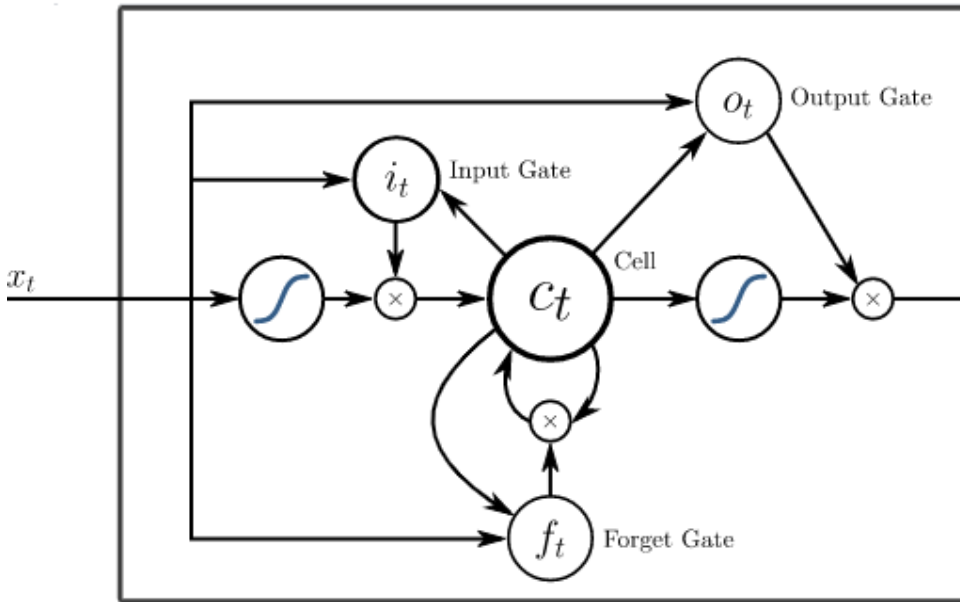
Sigmoid Function of vector $\sigma(a) = b$

$$a = [a_1 \dots a_k]^T \rightarrow b = \left[\frac{1}{1 + e^{-a_1}} \dots \frac{1}{1 + e^{-a_k}} \right]^T$$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Long Short-Term Memory (LSTM) (2/3)

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>



Αρχιτεκτονική Μονάδων LSTM

- Ρυθμιστικές λειτουργίες: **Input, Output, Forget Gates** (συνήθως νευρώνες με σιγμοειδή διέγερση)
- Memory **Cell** (αποθήκευση χρονικών εξαρτήσεων κατάστασης μέχρι ορίου λήθης, ρυθμιζόμενου από την **Forget Gate**)

Λειτουργίες Εισόδου/Εξόδου Μονάδων

$$f_t = \sigma_g(W_f x_t + U_f c_{t-1} + b_f)$$

$$i_t = \sigma_g(W_i x_t + U_i c_{t-1} + b_i)$$

$$o_t = \sigma_g(W_o x_t + U_o c_{t-1} + b_o)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + b_c)$$

$$h_t = \sigma_h(o_t \circ c_t)$$

όπου

$$\sigma_g(a) = \frac{1}{1+e^{-a}}, \text{ sigmoid function, οριακά } \{0,1\}$$

$$\sigma_c(a) = \tanh(a), \text{ hyperbolic tangent, οριακά } \pm 1$$

$$\sigma_c(a) = \tanh(a), \text{ hyperbolic tangent ή } \sigma_h(a) = a$$

◦ : **Hadamard** (point-wise) product

Ορισμοί Μεταβλητών Εισόδου/Εξόδου Μονάδων

d : number of input features

h : number of hidden units

$x_t \in \mathbb{R}^d$: input vector to the LSTM unit

$f_t \in \mathbb{R}^h$: forget gate's activation vector

$i_t \in \mathbb{R}^h$: input/update gate's activation vector

$o_t \in \mathbb{R}^h$: output gate's activation vector

$h_t \in \mathbb{R}^h$: output vector of the LSTM unit (hidden state vector)

$c_t \in \mathbb{R}^h$: cell state vector

$W \in \mathbb{R}^{h \times d}$, $U \in \mathbb{R}^{h \times h}$, $b \in \mathbb{R}^h$: weights & bias parameters, ρυθμίζονται με **supervised learning**

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Long Short-Term Memory (LSTM) (3/3)

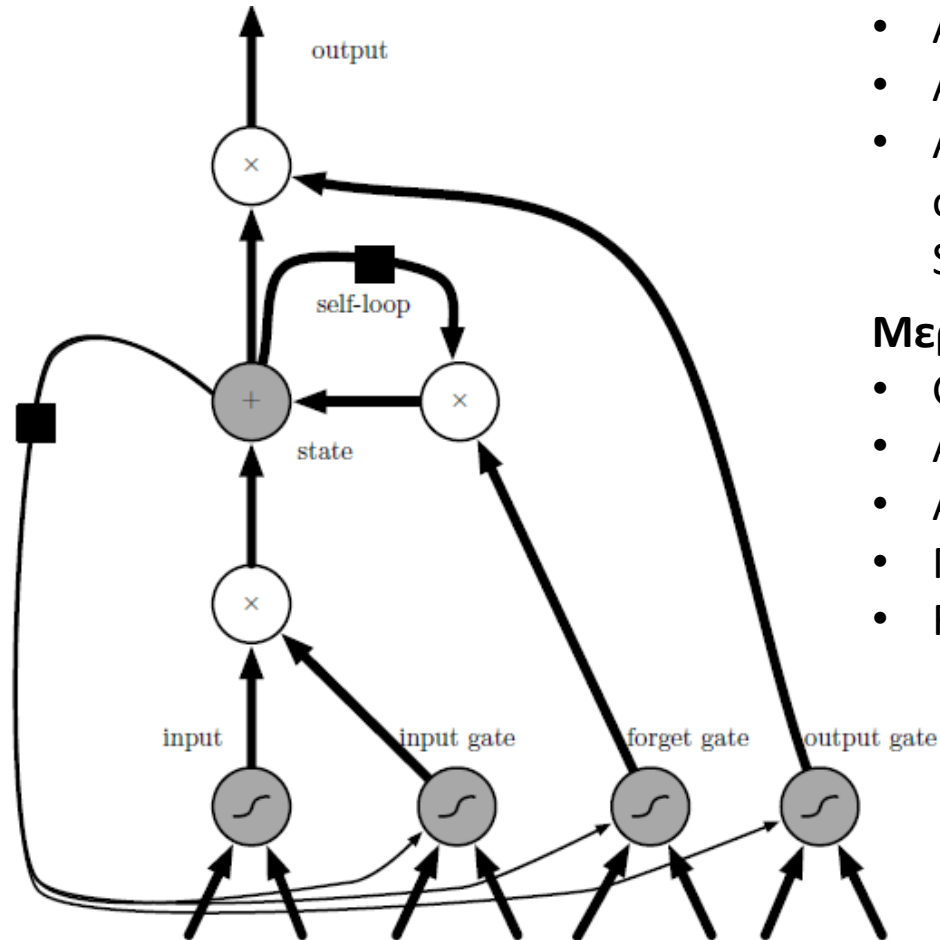
I. Goodfellow, Y. Bengio, A. Courville, "Deep Learning", Ch.10

Μερικά Πεδία Εφαρμογής LSTM

- Αναγνώριση χειρογράφων κειμένων
- Αναγνώριση φωνής
- Ανίχνευση ανωμαλιών σε υπολογιστικά συστήματα και δίκτυα – Intrusion Detection Systems (IDS)

Μερικές Εμπορικές Εφαρμογές LSTM

- Google (*Smartphone, Translate,...*)
- Apple (*Quicktype, iPhone, Siri...*)
- Amazon (*Alexa,...*)
- Microsoft (*Switchboard,...*)
- Facebook (*Automatic translation*)



Αποθήκευση Κατάστασης (State)

- Δυναμικός προσδιορισμός χρονικού παραθύρου μέσω **Forget Gate**
- Δυνατότητα πρόσβασης στη κατάσταση από άλλα στοιχεία (**Input. Output Gates**): **Peephole LSTM**
- Επιβλεπόμενη Μάθηση από **Labeled Datasets**