Εκπαίδευση Νευρωνικών Δικτύων

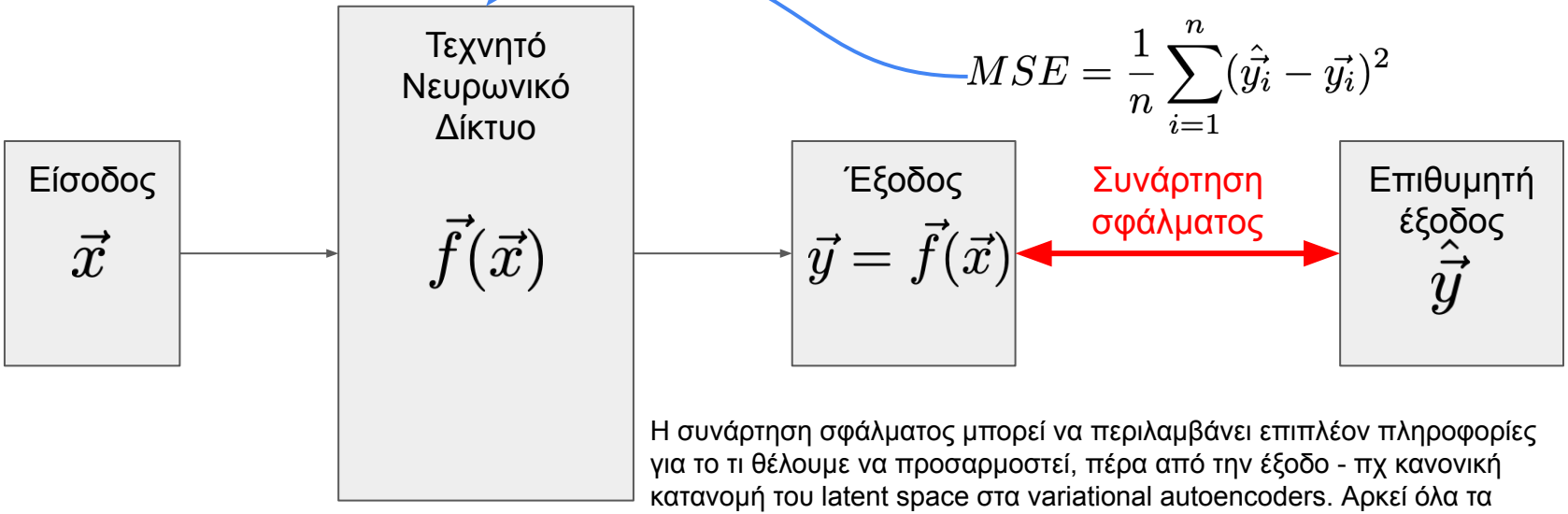
Συνάρτηση σφάλματος, παράγωγος, σύνολο επαλήθευσης και υπερπροσαρμογή

Γενική ιδέα της εκπαίδευσης

Back propagation: Μερική παράγωγος ως προς κάθε μεταβλητή παράμετρο (w, b) του δικτύου (πχ ακόμα και τιμές στα φίλτρα συνέλιξης)

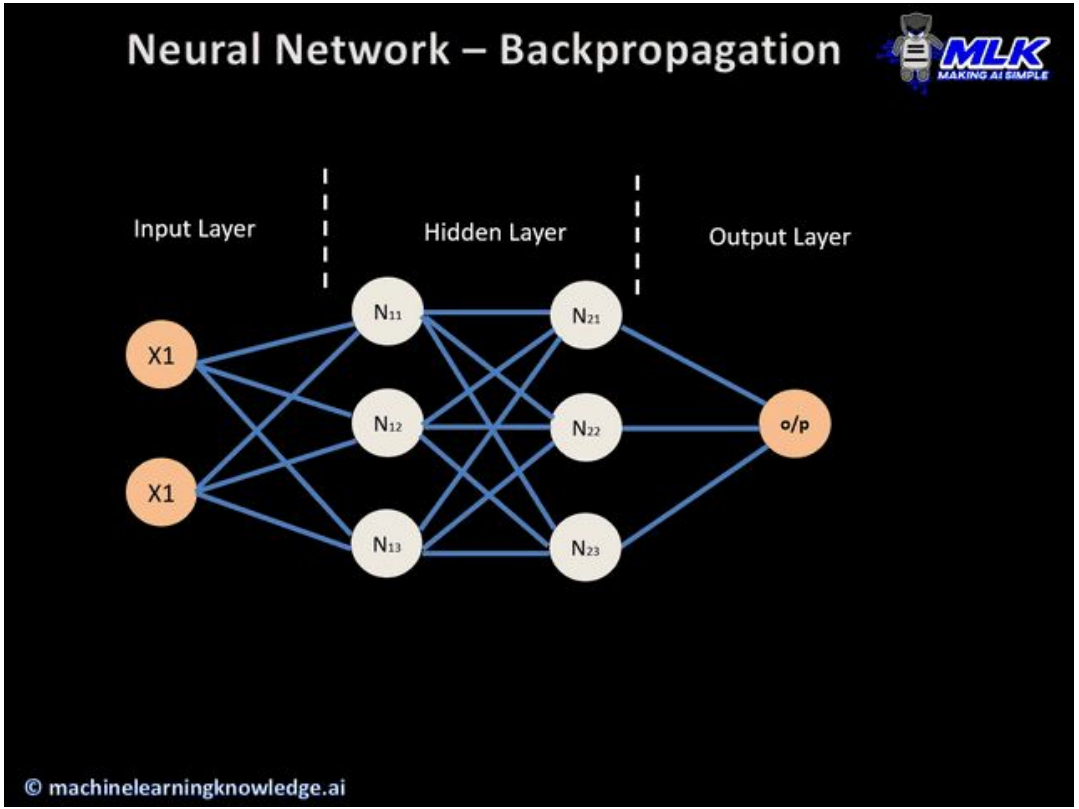
Παράδειγμα: μέσο τετραγωνικό σφάλμα (mean square error)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{\vec{y}_i} - \vec{y_i})^2$$



Η συνάρτηση σφάλματος μπορεί να περιλαμβάνει επιπλέον πληροφορίες για το τι θέλουμε να προσαρμοστεί, πέρα από την έξοδο - πχ κανονική κατανομή του latent space στα variational autoencoders. Αρκεί όλα τα επιπλέον στοιχεία να μπορούν να οριστούν ως διαφορίσιμες συναρτήσεις κόστους προς βελτιστοποίηση!

Backpropagation



https://machinelearningknowledge.ai/animated-explanation-of-feed-forward-neural-network-architecture/
https://www.youtube.com/watch?v=Ilg3gGewQ5U

Ρόλος της παραγώγου

Μας λέει προς ποιά κατεύθυνση η συνάρτηση μεγαλώνει σε καθε σημείο. Πχ

$$y = x^2 \Rightarrow y' = 2\ x$$

Για **x=3**, y(x)=9 και y'(x)=6 -> θετικό, άρα η συνάρτηση μεγαλώνει αν το x κινηθεί προς τα δεξιά (θετικά). Δηλ. για x=4, y(x)=16, που είναι μεγαλύτερο από το y(3).

Για **x=-2**, y(x)=9 και y'(x)=-6 -> αρνητικό, άρα η συνάρτηση μεγαλώνει αν το x κινηθεί προς τα αριστερά (αρνητικά). Δηλ. για x=-4, y(x)=16, που είναι μεγαλύτερο από το y(-3).

Πως βρίσκουμε το **ελάχιστο**; Κινούμαστε αντίθετα απ'ό,τι μας λέει η παράγωγος!

Gradient descent

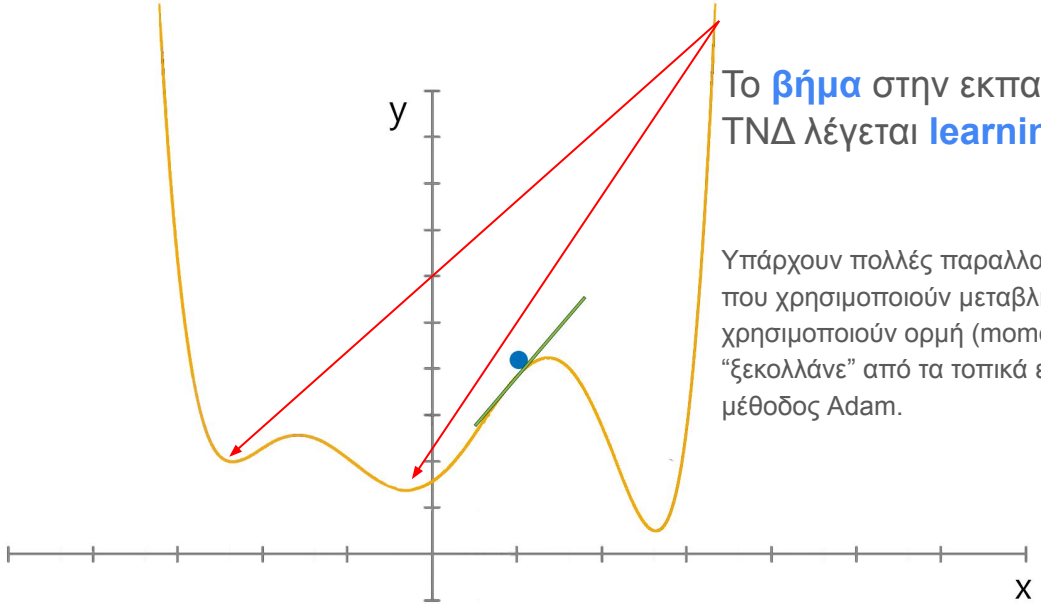
Είναι πιθανό να εγκλωβιστούμε σε τοπικά ελάχιστα. Εξαρτάται από το βήμα.

Ξεκινάμε από τυχαίο αρχικό σημείο x_0

Προχωράμε μέχρι να βρούμε ελάχιστο (πολύ μικρή τιμή παραγώγου):

$$x_{i+1} = x_i - h\ y'(x_i)$$

Όπου **h** είναι μια μικρή τιμή/**βήμα**. Το "-" σημαίνει ότι κινούμαστε αντίθετα.



Το **βήμα** στην εκπαίδευση ΤΝΔ λέγεται **learning rate**.

Υπάρχουν πολλές παραλλαγές μεθόδων που χρησιμοποιούν μεταβλητό βήμα, ή χρησιμοποιούν ορμή (momentum) για να "ξεκολλάνε" από τα τοπικά ελάχιστα. Πχ η μέθοδος Adam.

https://adatis.co.uk/introduction-to-artificial-neural-networks-part-two-gradient-descent-backpropagation-supervised-unsupervised-learning/

Gradient descent σε πολλές διαστάσεις

Π.χ. Στις 2 διαστάσεις:

Στο συγκεκριμένο παράδειγμα, ο αλγόριθμος βελτιστοποίησης τρέχει 3 φορές, από 3 διαφορετικά σημεία (βλέπουμε παράλληλα τις εκτελέσεις).

Παίζει σημαντικό ρόλο και το σημείο εκκίνησης - αρχικοποίηση.

Epochs and batch size

Το μέσο τετραγωνικό σφάλμα αφορά ένα δείγμα από τα πολλά, ενδεχομένως πολλές χιλιάδες, δείγματα (δεδομένα εκπαίδευσης)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{\vec{y}_i} - \vec{y_i})^2$$

Μια εποχή εκπαίδευσης ολοκληρώνεται όταν περάσουν όλα τα δεδομένα και υπολογιστεί το επόμενο βήμα. Δηλαδή: για όλα τα δεδομένα.

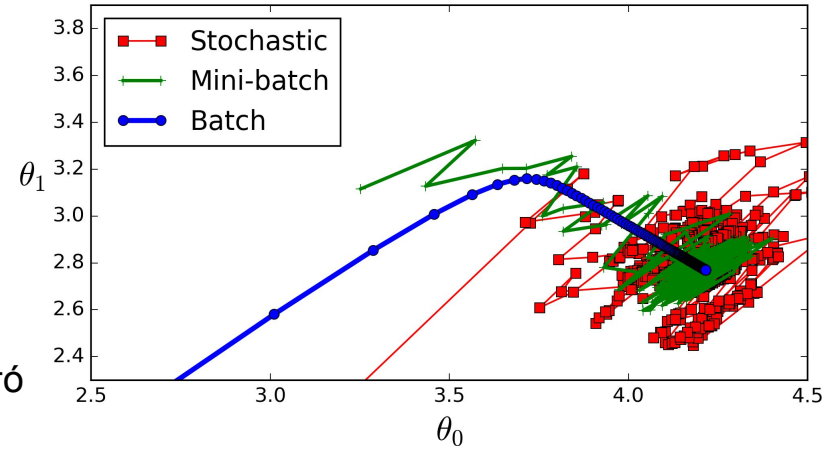
$$x_{i+1} = x_i - h\ y'(x_i)$$

Υπάρχουν τρεις τρόποι να περάσουν όλα τα δεδομένα εκπαίδευσης από τη διαδικασία:

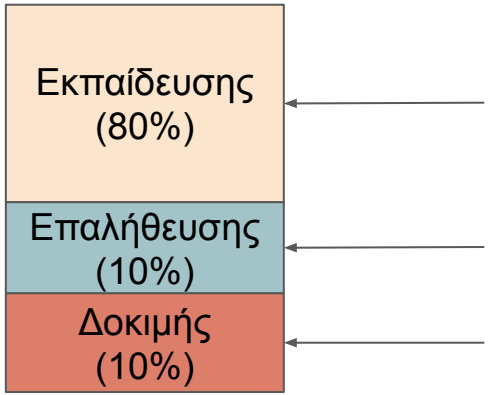
Stochastic: Οι τιμές των x_{i+1} ανανεώνονται σε κάθε δείγμα ξεχωριστά - τρέχει το backpropagation για κάθε δείγμα.

Batch: Περνούν όλα τα δεδομένα προς τα εμπρός, υπολογίζεται το άθροισμα του σφάλματος σε όλα και λειτουργεί το backpropagation μία φορά.

Mini-batch: η πιο διαδεδομένη, χωρίζεται το σύνολο εκπαίδευσης σε υποσύνολα, τα οποία περνούν με ξεχωριστό backpropagation στο άθροισμα των σφαλμάτων τους.



Σύνολο επαλήθευσης (validation set)

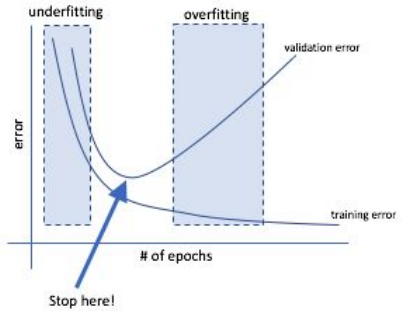


Backpropagation με σφάλματα στο σύνολο εκπαίδευσης

Μετά από κάθε εποχή, τρέχουμε το δίκτυο στο σύνολο επαλήθευσης, σαν μια ενδιάμεση δοκιμή σε δεδομένα που δεν έχει δει κατά την εκπαίδευση.

Αφού ολοκληρωθεί η εκπαίδευση, ελέγχουμε στο σύνολο δοκιμής.

Το σύνολο επαλήθεσης χρησιμεύει για να δούμε πόσο καλά γενικεύει σε κάθε εποχή - ίσως υπερπροσαρμόζεται και πρέπει να σταματήσουμε πρόωρα την εκπαίδευση…



Σύνοψη - σημαντικές έννοιες

- **Learning rate**: βήμα βελτιστοποίησης/ελαχιστοποίησης της συνάρτησης κόστους.
- **Epoch**: όταν ένα βήμα του αλγορίθμου βελτιστοποίησης έχει ολοκληρωθεί για όλα τα δεδομένα.
- **Batch size**: μέγεθος υποσυνόλου των δεδομένων στα οποία εφαρμόζεται ταυτόχρονα βελτιστοποίηση (στο άθροισμα του κόστους για κάθε δείγμα μέσα στο batch).
- **Validation set**: υποσύνολο των δεδομένων που εξετάζεται μετά την εκπαίδευση σε κάθε epoch για έλεγχο υπερμοντελοποίησης/υπερπροσαρμοστικότητας (overfitting).