

Εκμάθηση Αναπαραστάσεων (Representation Learning)

Βαθιά Μηχανική Μάθηση

ΔΠΜΣ Επιστήμης Δεδομένων & Μηχανικής Μάθησης

Εθνικό Μετσόβιο Πολυτεχνείο

Γιώργος Αλεξανδρίδης (gealexan@mail.ntua.gr)

Αναπαραστάσεις

- Ένα **πρόβλημα** λύνεται **ευκολότερα** αν η αναπαράσταση του είναι η **κατάλληλη**
 - Λχ ευκολότερα υπολογίζουμε την πράξη 30×40 με τους αραβικούς αριθμούς παρά με τους ελληνικούς ($\Lambda \times M$)
- Στη *μηχανική μάθηση* μια **αναπαράσταση** θεωρείται καλή αν **διευκολύνει** τη διαδικασία μάθησης
 - **Καλός συμβιβασμός** μεταξύ
 1. Συγκέντρωσης **όσο το δυνατόν περισσότερων** χαρακτηριστικών σχετικών με την είσοδο
 2. **Διατήρησης** όσων εμφανίζουν **ενδιαφέρουσες ιδιότητες**
 - λχ ανεξαρτησία μεταξύ τους

Αναπαραστάσεις

- Εκμάθηση κατάλληλων αναπαραστάσεων ιδιαίτερα σημαντική για προβλήματα **ημι-επιβλεπόμενης** και **μη-επιβλεπόμενης** μάθησης
- Πολλές φορές έχουμε **μεγάλο όγκο** δεδομένων **χωρίς** ετικέτα και **πολύ μικρό** με ετικέτα
 - Η επιβλεπόμενη μάθηση θα περιορίζονταν στο μικρό υποσύνολο των δεδομένων με ετικέτες
 - Πιθανά προβλήματα *υπερπροσαρμογής*
 - Μέσω διαδικασίας **μη-επιβλεπόμενης μάθησης**
 - Μπορούμε **να μάθουμε χαρακτηριστικά** του **χώρου εισόδου**
 - Χρησιμοποιούμε τις αναπαραστάσεις που μαθαίνουμε στο πρόβλημα της επιβλεπόμενης μάθησης

Μη-επιβλεπόμενη προπαίδευση (unsupervised pretraining)

- Βασίζεται σε μοντέλα που μαθαίνουν **λανθάνουσες αναπαραστάσεις** της εισόδου τους (λχ ΑΚ)
- Κάθε επίπεδο προ-εκπαιδεύεται **χωριστά**, χρησιμοποιώντας την είσοδο από το αμέσως προηγούμενο
 - Παράγεται μια καινούργια αναπαράσταση των δεδομένων η οποία, ιδανικά, είναι απλούστερη
- Χρήση για την αρχικοποίηση δικτύων όπως βαθιοί ΑΚ, deep belief networks, βαθιές μηχανές Boltzmann
 - Αποτέλεσε έναυσμα για την ανανέωση του ενδιαφέροντος για τα βαθιά δίκτυα από το 2006 και μετά

Άπληστη μη-επιβλεπόμενη ανά επίπεδο προπαίδευση

- *Greedy layer-wise unsupervised pretraining*

1. Άπληστη

- Κάθε τμήμα της βελτιστοποιείται ανεξάρτητα από τα υπόλοιπα

2. Ανά-επίπεδο

- Δομικό στοιχείο της βελτιστοποίησης είναι το επίπεδο
- Όταν βελτιστοποιείται το k -οστό επίπεδο, τα υπόλοιπα παραμένουν σταθερά

3. Μη-επιβλεπόμενη

- Χρήση αλγορίθμων μη-επιβλεπόμενης μάθησης

4. Προπαίδευση

- Αποτελεί αρχικό βήμα πριν την εφαρμογή αλγορίθμου εκπαίδευσης που βελτιστοποιεί όλα τα επίπεδα μαζί

Χαρακτηριστικά

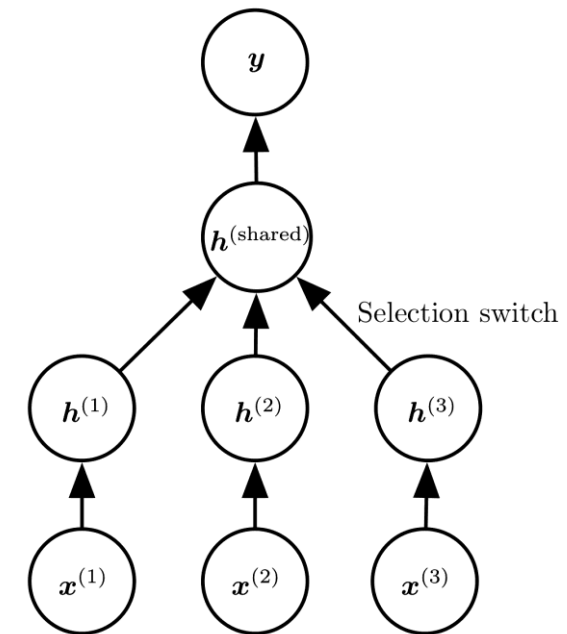
- Συνδυασμός **δύο** παρατηρήσεων
 1. Επιλογή **αρχικών παραμέτρων** βαθιού μοντέλου συνήθως έχει **σημαντική επίπτωση** στο ίδιο το μοντέλο
 2. Η **εκμάθηση της κατανομής της εισόδου** διευκολύνει στην εκμάθηση της **αντιστοίχισης** μεταξύ εισόδου-εξόδου
- Επιλογή αρχικών παραμέτρων στα βαθιά μοντέλα
 - Με **κατάλληλη επιλογή**, το μοντέλο καλύπτει περιοχές που δε θα **«έφτανε»** διαφορετικά
 - Λχ εκεί που η εκτίμηση της κλίσης θα ήταν αδύνατη
 - **Δύσκολο** να κατανοηθεί ποιες **πλευρές** των **προπαιδευμένων χαρακτηριστικών** μπορούν να **διατηρηθούν** στο στάδιο της **επιβλεπόμενης μάθησης**
- **Διευκόλυνση** επιβλεπόμενης μάθησης
 - Λχ βοηθάει να μπορούμε να ξεχωρίσουμε το πλήθος των τροχών σε περίπτωση που φτιάχνουμε ταξινομητή που ξεχωρίζει εικόνες αυτοκινήτων από μοτοσυκλετών

Χαρακτηριστικά

- Η μη-επιβλεπόμενη προπαίδευση βοηθά όταν η **αρχική αναπαράσταση** των δεδομένων είναι «φτωχή»
 - Λχ επεξεργασία κειμένου και απεικόνιση σε διανυσματικό χώρο
- **Χρησιμότητα**
 - Όταν τα δεδομένα χωρίς ετικέτα είναι πολυπληθή
 - Όταν η σχέση εισόδου-εξόδου είναι περίπλοκη
- **Μειονεκτήματα**
 - Δύο φάσεις εκπαίδευσης
 - Κάθε μια με δικές της παραμέτρους
 - Δεν μπορεί να προβλεφθεί η απόδοση της δεύτερης πριν τελειώσει η πρώτη
 - **Δεν επιτρέπει** τον ορισμό **του βαθμού συμμετοχής** της
 - Όπως λχ στην ομαλοποίηση
- Σήμερα χρησιμοποιείται κυρίως σε προβλήματα **επεξεργασίας φυσικής γλώσσας**

Μάθηση μέσω μεταφοράς (transfer learning)

- Χρησιμοποίηση γνώσης που έχει αποκτηθεί σε **ένα πεδίο** (λχ μια κατανομή P_1) για την επίτευξη **καλύτερης γενίκευσης** σε ένα **άλλο** (που περιγράφεται από κατανομή P_2)
 - Προϋπόθεση ότι **οι παράγοντες** που εξηγούν τις **διακυμάνσεις** στην P_1 είναι **σχετικές** με τις διακυμάνσεις που αναμένεται να εντοπιστούν στην P_2
- Ιδιαίτερα χρήσιμο σε προβλήματα **οπτικής αναγνώρισης**
 - Πολλές κατηγορίες αντικειμένων μοιράζονται χαρακτηριστικά χαμηλού επιπέδου όπως ακμές, σχήματα κλπ
- Σε ορισμένες περιπτώσεις, ο **διαμοιρασμός χαρακτηριστικών** γίνεται σε **ανώτερα επίπεδα**
 - λχ σε προβλήματα αναγνώρισης φωνής



Περιπτώσεις μάθησης μέσω μεταφοράς

- **Προσαρμογή πεδίου** (*domain adaptation*)

- Ίδιο ζητούμενο, **ελαφρά μεταβολή** εισόδου
 - Λχ σύστημα ανάλυσης συναισθήματος που έχει εκπαιδευτεί σε κριτικές ταινιών χρησιμοποιείται για την ανάλυση κριτικών μουσικής
 - Η ορολογία μεταβάλλεται μεταξύ των πεδίων, ωστόσο υπάρχει κοινή συνισταμένη για το ποιο σχόλιο είναι θετικό ή αρνητικό

- **Εννοιολογική ολίσθηση** (*concept drift*)

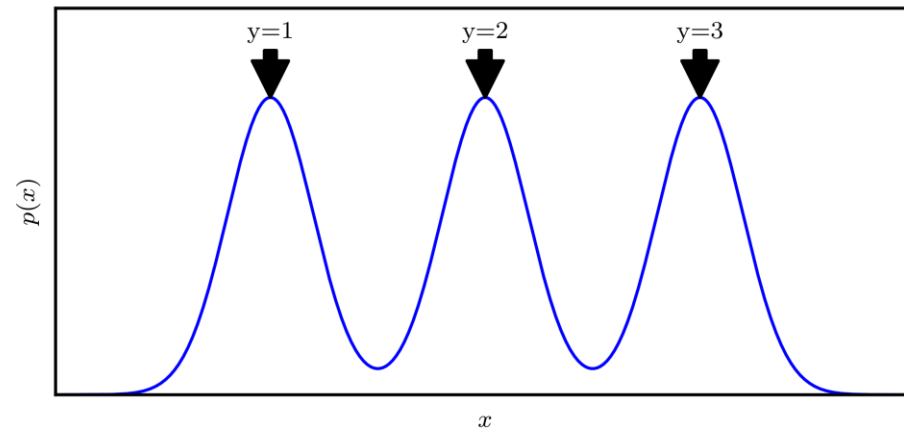
- **Βαθμιαία αλλαγή** στην κατανομή των δεδομένων εισόδου συναρτήσει του χρόνου
 - Λχ σύστημα εντοπισμού ανεπιθύμητης αλληλογραφίας ανταποκρίνεται στις αλλαγές του λεξιλογίου που χρησιμοποιούν οι spammers

Ποιότητα των αναπαραστάσεων

- Τι καθιστά μια αναπαράσταση **καλύτερη** από μια άλλη;
 1. Ιδανικά, τα χαρακτηριστικά της αναπαράστασης **ανταποκρίνονται** στις **υποκείμενες αιτίες** που παράγουν τα δεδομένα
 - Μια καλή αναπαράσταση για το $p(x)$ βοηθά στον υπολογισμό του $p(y|x)$, αν το y συμπεριλαμβάνεται στις κυρίαρχες αιτίες παραγωγής του x
 2. Η αναπαράσταση είναι **ευκολότερη** στη **μοντελοποίηση**
 - Αν μια αναπαράσταση h αναπαριστά τις περισσότερες από τις αιτίες που δημιουργούν τα x και η έξοδος y σχετίζεται με τα κυρίαρχα χαρακτηριστικά, τότε είναι πιο εύκολο να προβλεφθεί το y από το x

Ποιότητα των αναπαραστάσεων

- Πρόβλημα ημι-επιβλεπόμενης μάθησης
 - Κακή περίπτωση: Αν $p(x) \in \mathcal{U}$ τότε δεν μπορούμε να μάθουμε την $p(y|x)$
 - Καλή περίπτωση: Μίξη καλώς διαχωριζόμενων κατανομών, όπως στο δίπλα σχήμα

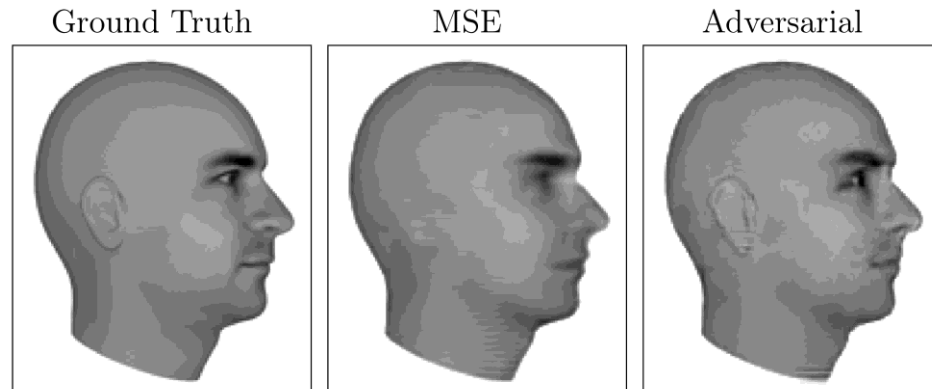
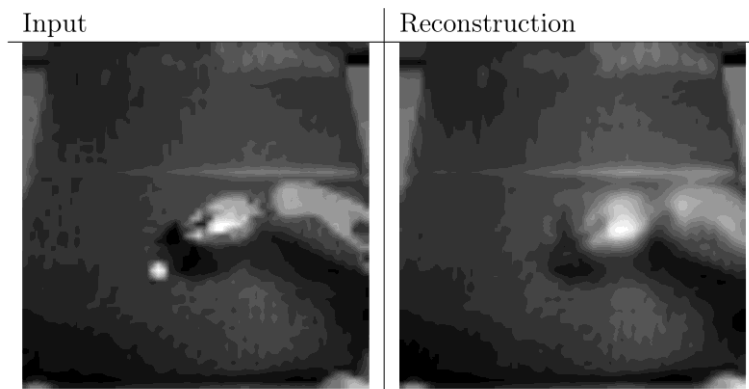


Παράγοντες συνάφειας (casual factors)

- Πως μπορούν να συνδεθούν $p(x)$ και $p(y|x)$;
 - Αν y στενά συνδεδεμένο με έναν από τους **παράγοντες συνάφειας** του x , τότε $p(x)$ και $p(y|x)$ στενά συνδεδεμένα μεταξύ τους
 - **h** : σύνολο παραγόντων συνάφειας
 - Υποκείμενες αιτίες που καθορίζουν το **h πολύ μεγάλες** σε πλήθος
 - **Δύσκολο** να βρεθούν με εξαντλητική αναζήτηση
 - Πχ σε πρόβλημα αναγνώρισης αντικειμένων σε εικόνα, πόσοι είναι οι πιθανοί παράγοντες συνάφειας;
 - Στρατηγικές αντιμετώπισης
 1. **Ταυτόχρονη** χρήση *επιβλεπόμενων* και *μη-επιβλεπόμενων* τεχνικών, για να «ανακαλύψει» το μοντέλο τους σημαντικότερους παράγοντες διακύμανσης στα δεδομένα
 2. Απεικόνιση σε **μεγαλύτερους χώρους** αναπαραστάσεων, όταν χρησιμοποιούνται **μόνο μη-επιβλεπόμενες** τεχνικές

Καθορισμός προεξεχόντων χαρακτηριστικών

- **Τεχνικές εκμάθησης αναπαραστάσεων** συνήθως βελτιστοποιούν ένα **συγκεκριμένο κριτήριο** για να εντοπίσουν τις προεξέχουσες αιτίες παραγωγής των δεδομένων
 - λχ μέσο τετραγωνικό σφάλμα
 - Δεν είναι πάντα αρκετό, μιας και παρακάτω «χάνονται» η μπάλα και το αυτί στις αναπαραστάσεις
- **Εναλλακτικοί τρόποι** καθορισμού προεξεχόντων χαρακτηριστικών
 - Αν ομάδα pixel περιέχει συχνά εμφανιζόμενο πρότυπο, τότε αυτό αναγνωρίζεται ως προεξέχων
 - Μοντέλο λειτουργίας των *Generative Adversarial Networks* (GANs), που αναγνωρίζουν το αυτί στη δεξιά εικόνα

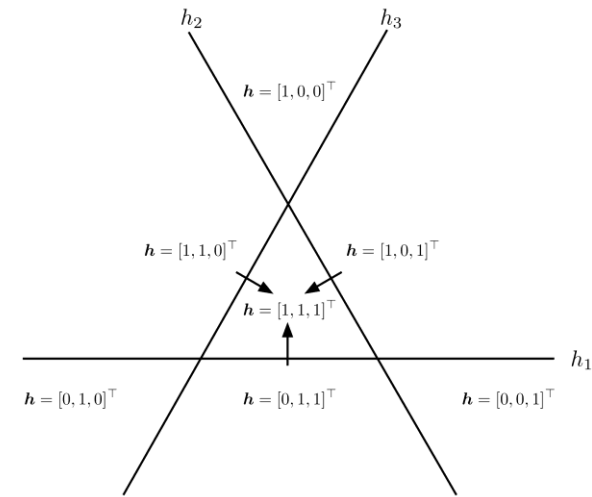


Κατανεμημένες και μη-κατανεμημένες αναπαραστάσεις

- Κατανεμημένες αναπαραστάσεις (*distributed representations*)
 - Χρησιμοποιούν n **χαρακτηριστικά** που παίρνουν k τιμές για να περιγράψουν k^n **διαφορετικές** έννοιες
- Μη-κατανεμημένες ή συμβολικές αναπαραστάσεις
 - Κάθε είσοδος αντιστοιχίζεται σε **συγκεκριμένη κατηγορία** ή **σύμβολο**
 - Δεν είναι σαφείς οι σχέσεις συνάφειας μεταξύ των συμβόλων
- Κατανεμημένες αναπαραστάσεις επιτυγχάνουν **γενίκευση**
 - πχ σε συμβολική αναπαράσταση η «γάτα» και ο «σκύλος» απέχουν όσο οποιαδήποτε άλλα σύμβολα
 - Σε κατανεμημένη αναπαράσταση μπορεί να ενώνονται μέσω χαρακτηριστικών όπως λχ το ότι είναι θηλαστικά, τετράποδα κλπ
- **Χώρος χαρακτηριστικών** κατανεμημένων αναπαραστάσεων έχει την ιδιότητα της **ομοιότητας**
 - Σημασιολογικά κοντινές έννοιες απεικονίζονται σε «**κοντινή**» απόσταση στο χώρο
 - Αυτό **δεν ισχύει** στις συμβολικές αναπαραστάσεις

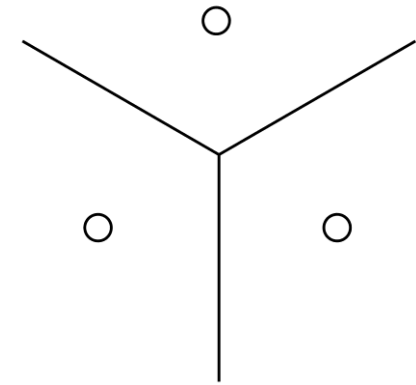
Παράδειγμα κατανεμημένης αναπαράστασης

- Διάνυσμα $n = 3$ δυαδικών χαρακτηριστικών
 - Λαμβάνει $2^3 = 8$ τιμές, που αντιστοιχούν σε **διαφορετικές** περιοχές του **χώρου εισόδου**
- **Διαχωρισμός** του χώρου εισόδου σε 3 περιοχές
 - Κάθε χαρακτηριστικό ορίζεται ως το **κατώφλι εξόδου** ενός γραμμικού μετασχηματισμού
 - Αν h_i^+ αντιπροσωπεύει τα σημεία για τα οποία $h_i = 1$ και h_i^- τα σημεία για τα οποία $h_i = 0$ τότε η αναπαράσταση $[1, 1, 1]^T$ αντιστοιχεί στην περιοχή $h_1^+ \cap h_2^+ \cap h_3^+$
- Στη γενική περίπτωση, μια κατανεμημένη αναπαράσταση n **χαρακτηριστικών** μαθαίνει n^d **διαφορετικές** περιοχές
 - **Εκθετικά περισσότερες** σε πλήθος σε σχέση λχ με τον αλγόριθμο των k -μέσων που μαθαίνει μόλις n



Παράδειγμα μη-κατανεμημένης αναπαράστασης

- Διαχωρισμός χώρου εισόδου σε 3 περιοχές
 - Αλγόριθμος εύρεσης πλησιέστερου γείτονα
 - **Διαφορετικοί** αλγόριθμοι μη-κατανεμημένης αναπαράστασης επιτυγχάνουν **διαφορετικές** γεωμετρίες
 - Όλοι όμως τεμαχίζουν τον χώρο σε περιοχές με διαφορετικό σύνολο παραμέτρων
- **Πλεονέκτημα**
 - Αλγόριθμος «**μαθαίνει**» την κατανομή των δεδομένων **χωρίς** να λύνει ένα **δύσκολο** πρόβλημα βελτιστοποίησης
 - Δεδομένου ότι υπάρχει ικανός αριθμός παραμέτρων για να περιγραφεί το πρόβλημα
- **Μειονεκτήματα**
 - Τοπική γενίκευση
 - Δεν μπορούν να «μάθουν» πιο περίπλοκους χώρους



Χαρακτηριστικά Συμβολικών Αναπαραστάσεων

- **Υπόθεση ομαλότητας** (*smoothness assumption*)
 - Αν $u \approx v$ για τη συνάρτηση f που θέλουμε να μάθουμε ισχύει εν γένει $f(u) \approx f(v)$
 - Βάση αλγορίθμων που μαθαίνουν συμβολικά χαρακτηριστικά
- **«Κατάρα» της διαστατικότητας** (*curse of dimensionality*)
 - Αν f μεταβάλλεται σε πολλές περιοχές, τότε χρειαζόμαστε **τουλάχιστον τόσα** δείγματα **όσες** και οι διακριτές περιοχές
- Κάθε σύμβολο αποτελεί **διαφορετικό βαθμό ελευθερίας** για κάθε περιοχή
 - Μπορούμε να μάθουμε «**αυθαίρετες**» απεικονίσεις μεταξύ περιοχών
 - **Δεν μπορούμε** να **γενικεύσουμε** για νέες περιοχές

Χαρακτηριστικά Κατανεμημένων Αναπαραστάσεων

- Αλγόριθμος εκμάθησης κατανεμημένων χαρακτηριστικών που εξάγει **δυσιαδικά χαρακτηριστικά** μέσω ορισμού κατώφλιου σε **γραμμικές συναρτήσεις**
 - Κάθε δυαδικό χαρακτηριστικό τεμαχίζει το χώρο \mathbb{R}^d σε 2 υπο-περιοχές
 - Από n υπερεπίπεδα του χώρου \mathbb{R}^d δημιουργούνται $\mathcal{O}(n^d)$ περιοχές
 - **Εκθετικές** ως προς τον χώρο της εισόδου
 - **Πολυωνυμικές** ως προς το εύρος των λανθανουσών χαρακτηριστικών
- Με $\mathcal{O}(nd)$ **δείγματα** προσεγγίζουμε $\mathcal{O}(n^d)$ **περιοχές** στο χώρο εισόδου
 - Στις συμβολικές αναπαραστάσεις απαιτούνται $\mathcal{O}(n^d)$ **δείγματα**
- Η παραπάνω συλλογιστική επεκτείνεται και στις περιπτώσεις που ο χώρος χωρίζεται από *μη-γραμμικά* κατώφλια

Βιβλιογραφία

- Ian Goodfellow, Yoshua Bengio, Aaron Courville “Deep Learning” – MIT Press (<https://www.deeplearningbook.org/>)
 - Εισαγωγή (§15.1)
 - Μη-επιβλεπόμενη προπαίδευση (§15.1)
 - Μάθηση μέσω μεταφοράς (§15.2)
 - Παράγοντες συνάφειας (§15.3)
 - Κατανεμημένες και μη-κατανεμημένες αναπαραστάσεις (§15.4)