

Αυτοκωδικοποιητές (Autoencoders) και Εκμάθηση Αναπαραστάσεων (Representation Learning)

Βαθιά Μηχανική Μάθηση

ΔΠΜΣ Επιστήμης Δεδομένων & Μηχανικής Μάθησης

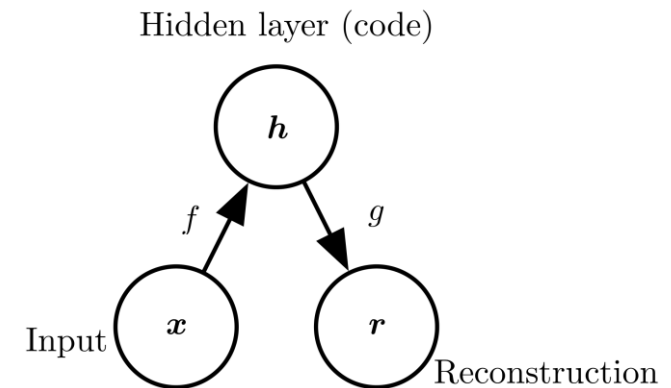
Εθνικό Μετσόβιο Πολυτεχνείο

Γιώργος Αλεξανδρίδης

Αυτοκωδικοποιητές

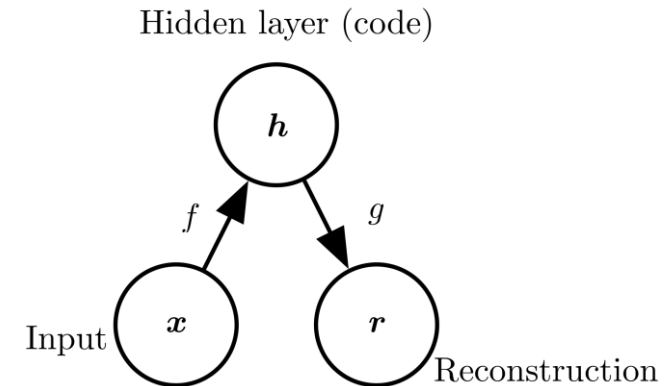
Αυτοκωδικοποιητές

- **Αυτοκωδικοποιητής** (*Autoencoder* ή ΑΚ)
 - Νευρωνικό δίκτυο που εκπαιδεύεται να αντιγράψει την είσοδό του στην έξοδό του
 - Εσωτερικά αποτελείται από κρυφό επίπεδο \mathbf{h} στο οποίο αναπαρίσταται **κωδικοποιημένη** (*coded*) η είσοδος
 - Δύο μέρη
 1. Συνάρτηση **κωδικοποίησης** (*encoder function*) $\mathbf{h} = f(\mathbf{x})$
 2. Συνάρτηση **αποκωδικοποίησης** (*decoder function*) $\mathbf{r} = g(\mathbf{h})$



Αυτοκωδικοποιητές

- Εκπαίδευση
 - Μέσω οπίσθιας διάδοσης του σφάλματος
 - Όπως στα ΤΝΔ πρόσθιας τροφοδότησης
 - Μέσω επανακυκλοφορίας (recirculation)
 - Συγκρίνεται η ενεργοποίηση των νευρώνων στην αρχική είσοδο και στην αναπαράσταση
- Τετριμμένη λύση: Μάθε την $g(f(x)) = x, \forall x$
 - Δεν έχει νόημα γι' αυτό οι ΑΚ σχεδιάζονται έτσι ώστε **να μην μπορούν** να αντιγράψουν τέλεια
- Παρουσιάστηκαν στα μέσα του 1980
- Χρήση σε προβλήματα **μείωσης διαστατικότητας** (*dimensionality reduction*) και **εξαγωγής χαρακτηριστικών** (*feature extraction*)



Διαδικασία Μάθησης

- **Δεν μας ενδιαφέρει** τόσο η διαδικασία **αντιγραφής**, όσο να **αποτυπωθούν** στο **h** **χρήσιμες ιδιότητες** του **x**
 - **Υποπλήρης** (*undercomplete*) ΑΚ: Περιορίζουμε το **h** ώστε να έχει μικρότερες διαστάσεις από το **x**
 - Μαθαίνει τα **προεξέχοντα** (*salient*) χαρακτηριστικά της εισόδου
- **Διαδικασία μάθησης**: Ελαχιστοποίηση συνάρτησης απώλειας $L(x, g(f(x)))$
 - Αν g γραμμική και L Μέσο Τετραγωνικό Σφάλμα (ΜΤΣ), τότε ο υποπλήρης ΑΚ καλύπτει τον ίδιο **υποχώρο** με την PCA (Γιατί; Άσκηση για το σπίτι!)
 - Αν f, g μη γραμμικές, ο υποπλήρης ΑΚ μαθαίνει μια πιο **ισχυρή μη-γραμμική γενίκευση** της PCA
 - Αν f, g έχουν **πολύ μεγάλη χωρητικότητα**, τότε απλά θα μάθουν να **αντιγράφουν** την είσοδο στην έξοδο χωρίς εξαγωγή χαρακτηριστικών!
- **Υπερπλήρης** (*overcomplete*) ΑΚ: Θέτουμε $\dim(\mathbf{h}) \geq \dim(\mathbf{x})$
- Η επιλογή της **διάστασης** του ΑΚ καθώς και της χωρητικότητας των f, g εξαρτάται από την **πολυπλοκότητα** της **υποκείμενης** (*underlying*) κατανομής των δεδομένων που μοντελοποιούνται

Ομαλοποιημένοι Αυτοκωδικοποιητές

- **Ομαλοποιημένοι** (*regularized*) ΑΚ
 - Περιορισμός της χωρητικότητας τους μέσω **συνάρτησης απώλειας**
 - Θυμηθείτε τη συζήτηση περί ομαλοποίησης σε προηγούμενη διάλεξη
- Συνάρτηση απώλειας επιβάλλει **πρόσθετες ιδιότητες** πέραν της αντιγραφής
 - **Αραιότητα** (*sparsity*) της αναπαράστασης
 - Μικρό **μέγεθος** της παραγώγου
 - **Ανοχή** σε **θόρυβο** ή σε απουσιάζουσες τιμές
- Ένας *ομαλοποιημένος* ΑΚ μπορεί να είναι *μη-γραμμικός* και *υπερπλήρης* αλλά παρόλα αυτά να **μαθαίνει χαρακτηριστικά** της κατανομής των δεδομένων

Αραιοί Αυτοκωδικοποιητές

- **Αραιός** (*sparse*) ΑΚ
 - Προσθήκη **όρου ποινής** αραιότητας $\Omega(\mathbf{h})$ του επιπέδου κωδικοποίησης \mathbf{h} στη διαδικασία μάθησης
 - $L(\mathbf{x}, g(f(\mathbf{x}))) + \Omega(\mathbf{h})$
- Προσθήκη **όρων ομαλοποίησης** στη διαδικασία μάθησης
 - Προσέγγιση της *Μεϋζιανής Συμπερασματολογίας* υπό τη μορφή της *Μέγιστης εκ των Υστέρων Πιθανότητας* (MAP)
 - Το συναντήσαμε και στην περίπτωση της ομαλοποίησης σε προηγούμενη διάλεξη
 - **Ποινή ομαλοποίησης**: η εκ των προτέρων κατανομή $p(\theta)$ των παραμέτρων θ του μοντέλου
 - $p(\theta|x) \approx p(x|\theta)p(\theta)$

Αραιοί Αυτοκωδικοποιητές

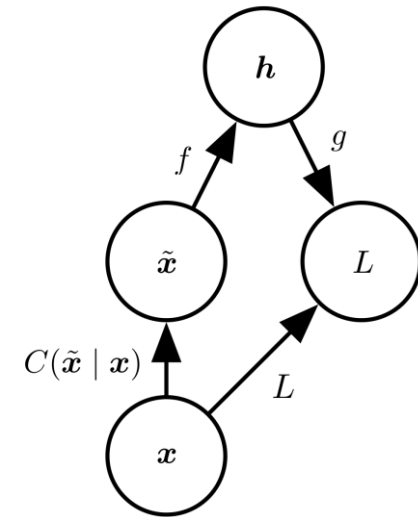
- Στους ΑΚ η ποινή ομαλοποίησης εφαρμόζεται στα δεδομένα (\mathbf{x}) και όχι στις παραμέτρους (θ)!
 - Ο αραιός ΑΚ δρα ως εκτιμητής μέγιστης πιθανοφάνειας ενός παραγωγικού (*generative*) μοντέλου με λανθάνουσες (*latent*) μεταβλητές \mathbf{h}
 - $p_{model}(\mathbf{x}, \mathbf{h}) = p_{model}(\mathbf{h})p_{model}(\mathbf{x}|\mathbf{h})$
 - $p_{model}(\mathbf{h})$: εκτίμηση μοντέλου για την εκ των προτέρων κατανομή των λανθανουσών μεταβλητών
- $\log p_{model}(\mathbf{x}) = \log \sum_{\mathbf{h}} p_{model}(\mathbf{h}, \mathbf{x})$
 - Στους ΑΚ το άθροισμα προσεγγίζεται από σημειακή εκτίμηση μιας πολύ πιθανής τιμής για το \mathbf{h}
 - $\max_{\mathbf{h}} \log p_{model}(\mathbf{h}, \mathbf{x}) = \max_{\mathbf{h}} \log p_{model}(\mathbf{h}) + \max_{\mathbf{h}} \log p_{model}(\mathbf{x}|\mathbf{h})$

Εκ των προτέρων κατανομή λανθανουσών μεταβλητών

- **Εισαγωγή αραιότητας:** κατάλληλη επιλογή $p_{model}(\mathbf{h})$
 - πχ Laplace, Student-t
- Κατανομή Laplace: **Ισοδυναμεί** με κανονικοποίηση L_1
 - $p_{model}(h_i) = \frac{\lambda}{2} e^{-\lambda|h_i|} \Rightarrow -\log p_{model}(\mathbf{h}) = \sum_i (\lambda|h_i| - \log \frac{\lambda}{2}) = \Omega(\mathbf{h}) + c$
 - Όρος c **σταθερά** (υπερπαράμετρος του μοντέλου)
 - Δεν εξαρτάται από το \mathbf{h} αλλά μόνο από το λ
- Η ποινή αραιότητας **δεν είναι** όρος *κανονικοποίησης*
 - Γιατί **δεν εφαρμόζεται** στις **παραμέτρους** του μοντέλου αλλά στις **λανθάνουσες μεταβλητές**
 - Αποτελεί **συνέπεια** της θεώρησης για την **κατανομή** του μοντέλου όσον αφορά τις **λανθάνουσες μεταβλητές**
- Συνεπώς, η **εκπαίδευση** ενός **ΑΚ** είναι αντίστοιχη της εκπαίδευσης ενός **παραγωγικού μοντέλου**
 - Τα χαρακτηριστικά που μαθαίνει ο **ΑΚ** είναι επί της ουσίας οι **λανθάνουσες** μεταβλητές που χαρακτηρίζουν την είσοδο

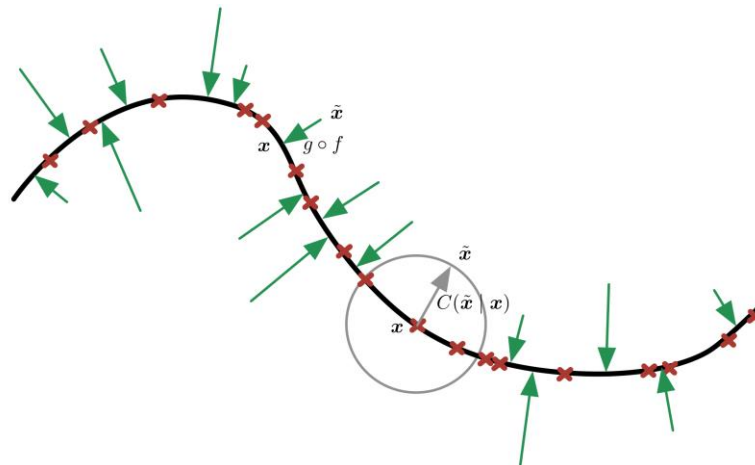
Αυτοκωδικοποιητές απαλοιφής θορύβου

- **ΑΚ απαλοιφής θορύβου** (*Denoising Autoencoders* ή DAE)
 - Ελαχιστοποίηση της $L(x, g(f(\tilde{x})))$
 - \tilde{x} αλλοίωση του x μέσω της προσθήκης **θορύβου**
 - $C(x|\tilde{x})$: Υπό **συνθήκη κατανομής** παραγωγής αλλοιωμένων \tilde{x} δεδομένου x
- Ο ΑΚ μαθαίνει την **κατανομή αποκατάστασης** (*reconstruction distribution*)
 - Λήψη δείγματος x από τα δεδομένα εκπαίδευσης και **αλλοιωμένης του μορφής** \tilde{x} από την $C(\tilde{x}|x = x)$
 - Χρήση (x, \tilde{x}) ως **δείγμα εκπαίδευσης** για την εκτίμηση $p_r(\tilde{x}|x) = p_d(x|h)$
 - p_d ορίζεται από $g(h)$
- Αν θέσουμε f ντετερμινιστική, ο DAE **συμπεριφέρεται** ως ένα ΤΝΔ πρόσθιας τροφοδότησης
 - Συνεπώς μπορούμε να χρησιμοποιήσουμε αντίστοιχές τεχνικές μάθησης (λχ κατάβαση κλίσης)



Συνταίριασμα τιμών (score matching)

- Εναλλακτική μέθοδος στην εκτίμηση μέγιστης πιθανοφάνειας
 - Παραγωγή εκτιμήσεων για το μοντέλο, οι οποίες «ενθαρρύνονται» να εμφανίζουν παρόμοιες τιμές (scores) με την κατανομή των δεδομένων σε κάθε δείγμα x
 - Οι τιμές προκύπτουν ως η αποτίμηση ενός πεδίου κλίσεων (gradient field): $\nabla_x \log p(x)$
- Σημαντική ιδιότητα των DAE
 - Το κριτήριο εκπαίδευσής τους κάνει τον ΑΚ να μάθει ένα διανυσματικό πεδίο $g(f(x)) - x$ που αποτελεί εκτίμηση της τιμής της κατανομής των δεδομένων



Εκμάθηση Πολλαπλότητας

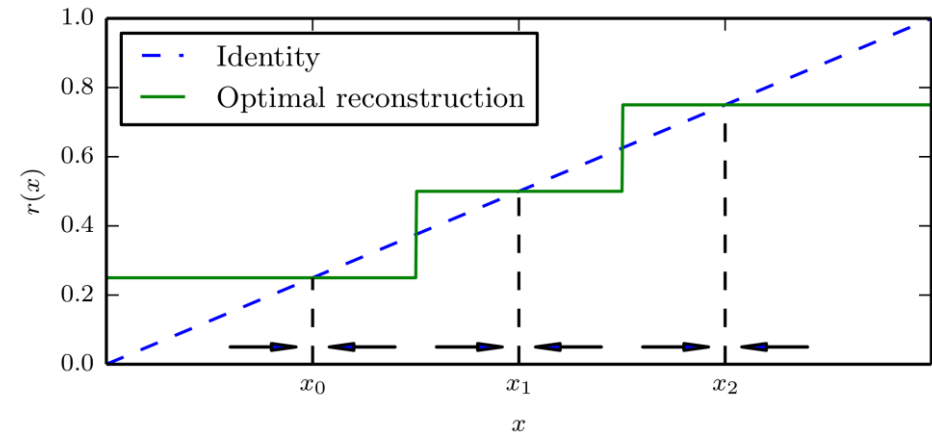
- Γενική υπόθεση ΑΚ
 - Τα δεδομένα είναι συγκεντρωμένα γύρω από **πολλαπλότητες** (*manifolds*) **χαμηλότερων διαστάσεων**
- Πολλαπλότητες χαρακτηρίζονται από τις **εφαπτόμενες πλευρές** τους (*tangent planes*)
 - Πολλαπλότητα **διάστασης** d ορίζεται από d **διανύσματα βάσης**
 - Διανύσματα βάσης ορίζουν τις επιτρεπόμενες **αποκλίσεις** εντός της πολλαπλότητας
 - Πόσο μπορεί να «αλλάξει» το x παραμένοντας ενός της πολλαπλότητας

Εκμάθηση Πολλαπλότητας

- Συμβιβασμός μεταξύ **δύο αντίρροπων** δυνάμεων
 1. **Εκμάθησης κωδικοποίησης h** από τα δεδομένα x
 - έτσι ώστε το x να μπορεί να ανακτηθεί προσεγγιστικά από το h μέσω του ΑΚ
 2. **Ικανοποίησης περιορισμών** που περιορίζουν την χωρητικότητα του ΑΚ
 - λχ ποινές ομαλοποίησης
 - ΑΚ μαθαίνει αναπαραστάσεις που είναι **λιγότερο «ευαίσθητες»** στη μεταβολή της εισόδου
- Τελικά, ο ΑΚ **μαθαίνει μόνο** τις **διακυμάνσεις** που είναι **απαραίτητες** για την ανακατασκευή των δειγμάτων εκπαίδευσης

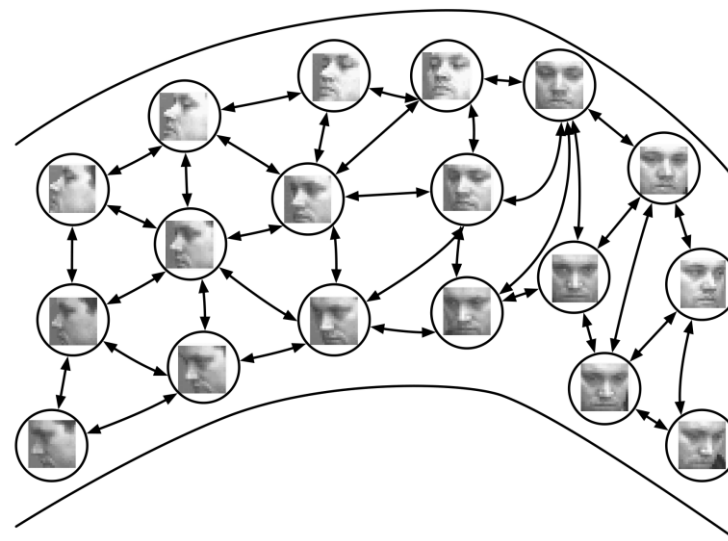
Παράδειγμα Εκμάθησης Πολλαπλότητας

- Χώρος δεδομένων μίας διάστασης
 - Δε μπορεί να μειωθεί άλλο η διάσταση
 - «Σημειακή» πολλαπλότητα
- Διακεκομμένη γραμμή
 - Ιδεατή συνάρτηση ταυτότητας
 - Αυτή επιθυμεί να κατασκευάσει ο ΑΚ
- Βέλη
 - Κατεύθυνση διανύσματος αναπαράστασης $r(x) - x$
 - Δείχνουν προς την **πλησιέστερη πολλαπλότητα** (εδώ, σημείο στο χώρο)
- Συνεχόμενη γραμμή
 - **Ιδεατή** συνάρτηση αναπαράστασης
 - **Τέμνει** την ιδεατή συνάρτηση ταυτότητας στα δείγματα των δεδομένων
 - **Μεγάλη** παράγωγος στον χώρο μεταξύ των πολλαπλοτήτων
 - Έτσι τα «αλλοιωμένα» σημεία απεικονίζονται πάνω στη «σημειακή» πολλαπλότητα



Μη-παραμετρικές μέθοδοι εκμάθησης πολλαπλότητας

- Τεχνικές μη-επιβλεπόμενης μάθησης
 - Γράφος πλησιέστερων γειτόνων
- Ένας κόμβος για κάθε δείγμα εκπαίδευσης
 - Ένωση με ακμές με τους γειτονικούς κόμβους
- Ανάκτηση εφραπτόμενων πλευρών κάθε κόμβου και της διανυσματικής του θέσης
 - **Embedding**
- Γενίκευση σε νέα δείγματα μέσω **παρεμβολής**
 - Αρκεί το αρχικό πλήθος των δειγμάτων να μπορεί να καλύψει πλήρως τα χαρακτηριστικά της πολλαπλότητας
 - Καμπυλότητες, πτυχώσεις κλπ



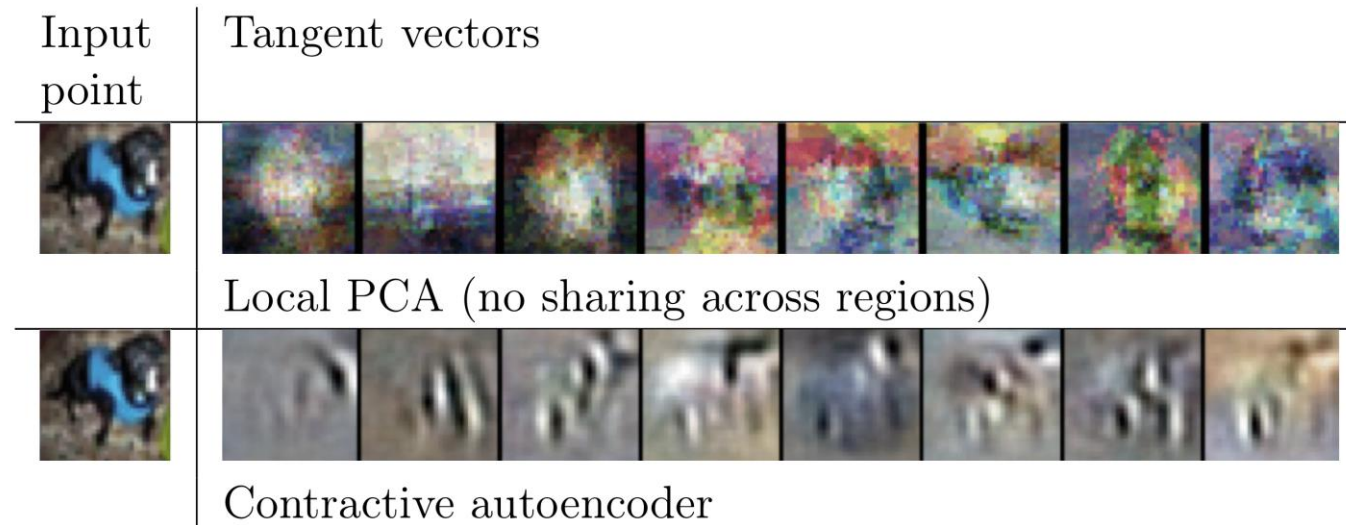
Συσταλτικοί Αυτοκωδικοποιητές

- *Contractive Autoencoders* ή CAE
 - Δεν πρέπει να μεταβάλλεται **πολύ** η αναπαράσταση όταν η είσοδος μεταβάλλεται **ελαφρά**
 - $L(x, g(f(x))) + \Omega(\mathbf{h}, x), \quad \Omega(\mathbf{h}, x) = \lambda \|\nabla_x \mathbf{h}\|^2$
 - **Όρος ποινής** $\Omega(\mathbf{h})$: Το τετράγωνο της νόρμας Frobenius του Ιακωβιανού πίνακα της συνάρτησης κωδικοποίησης
- Σχέση DAE και CAE
 - Το **σφάλμα αναπαράστασης** όταν έχει προστεθεί στην είσοδο μικρή ποσότητα γκαουσιανού θορύβου είναι ισοδύναμο με **μια συσταλτική ποινή** στη συνάρτηση αναπαράστασης $g \circ f$

Συσταλτικοί Αυτοκωδικοποιητές

- Ιδιότητα συστολής είναι **τοπική**
 - Ο Ιακωβιανός πίνακας *προσεγγίζει γραμμικά* γύρω από σημείο x μια μη-γραμμική συνάρτηση
- Αντίρροπες δυνάμεις
 1. Σφάλμα αναπαράστασης
 2. Ποινή συστολής $\Omega(\mathbf{h})$
- **Ισορροπία** στα σημεία που οι περισσότερες *μερικές* παράγωγοι είναι **μηδέν**
 - Μόνο ένα μικρό πλήθος χαρακτηριστικών (διαστάσεων) θα έχει μεγάλες τιμές
 - CAE απεικονίζει χαρακτηριστικά σε μια **πολλαπλότητα**
- Γενικά, CAE μαθαίνει χαρακτηριστικά **σταθερά** ως προς x

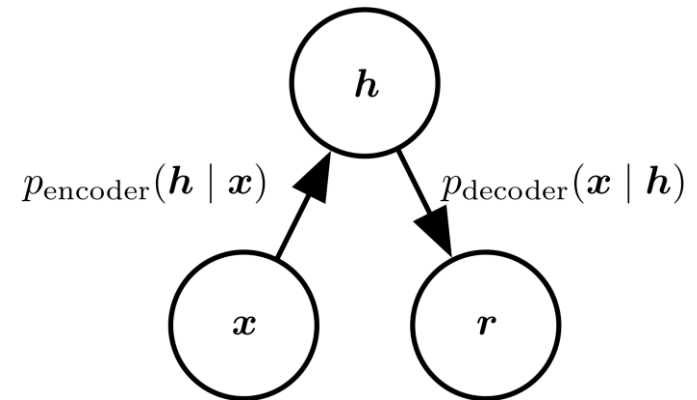
Συσταλτικοί Αυτοκωδικοποιητές: Παράδειγμα



- Εικόνα από το CIFAR-10 dataset
- Εφαπτόμενα διανύσματα υπολογίζονται από τα **κυρίαρχα ιδιάζοντα διανύσματα** Ιακωβιανής μήτρας $\frac{\partial h}{\partial x}$
- CAE **καλύτερη αναπαράσταση** από PCA
 - Εκμεταλλεύεται ιδιότητα **διαμοιρασμού παραμέτρων** μεταξύ διαφορετικών περιοχών

Στοχαστικοί αυτοκωδικοποιητές

- Διαδικασία κωδικοποίησης και αποκωδικοποίησης στοχαστική
 - Συνάρτηση Κωδικοποίησης $f \Rightarrow$ κατανομή κωδικοποίησης $p_{encoder}(\mathbf{h}|\mathbf{x})$
 - Αντίστοιχα $g \Rightarrow p_{decoder}(\mathbf{x}|\mathbf{h})$
- Κάθε μοντέλο λανθανουσών μεταβλητών $p_m(\mathbf{x}, \mathbf{h})$ μπορεί να ορίσει στοχαστικό κωδικοποιητή και αποκωδικοποιητή
 - $p_{encoder}(\mathbf{h}|\mathbf{x}) = p_m(\mathbf{h}|\mathbf{x})$
 - $p_{decoder}(\mathbf{x}|\mathbf{h}) = p_m(\mathbf{x}|\mathbf{h})$
- Στη γενική περίπτωση, ωστόσο, οι $p_{encoder}$ και $p_{decoder}$ δεν αποτελούν υπό συνθήκη κατανομές μιας ενιαίας κοινής κατανομής



Εκμάθηση αναπαραστάσεων

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*
- Άπληστη
 - Κάθε τμήμα της βελτιστοποιείται ανεξάρτητα από τα υπόλοιπα
- Ανά-επίπεδο
 - Δομικό στοιχείο της βελτιστοποίησης είναι το επίπεδο
 - Όταν βελτιστοποιείται το k -οστό επίπεδο, τα υπόλοιπα παραμένουν σταθερά
- Μη-επιβλεπόμενη
 - Χρήση αλγορίθμων μη-επιβλεπόμενης μάθησης
- Προπαίδευση
 - Αποτελεί αρχικό βήμα πριν την εφαρμογή αλγορίθμου εκπαίδευσης που βελτιστοποιεί όλα τα επίπεδα μαζί

Χαρακτηριστικά

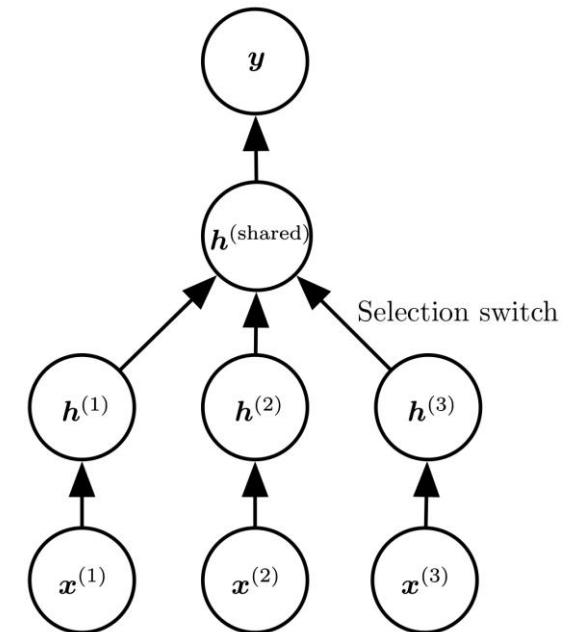
- Συνδυασμός **δύο** παρατηρήσεων
 1. Επιλογή **αρχικών παραμέτρων** βαθιού μοντέλου συνήθως έχει **σημαντική επίπτωση** στο ίδιο το μοντέλο
 2. Η **εκμάθηση της κατανομής της εισόδου** διευκολύνει στην εκμάθηση της **αντιστοίχισης** μεταξύ εισόδου-εξόδου
- Επιλογή αρχικών παραμέτρων στα βαθιά μοντέλα
 - 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

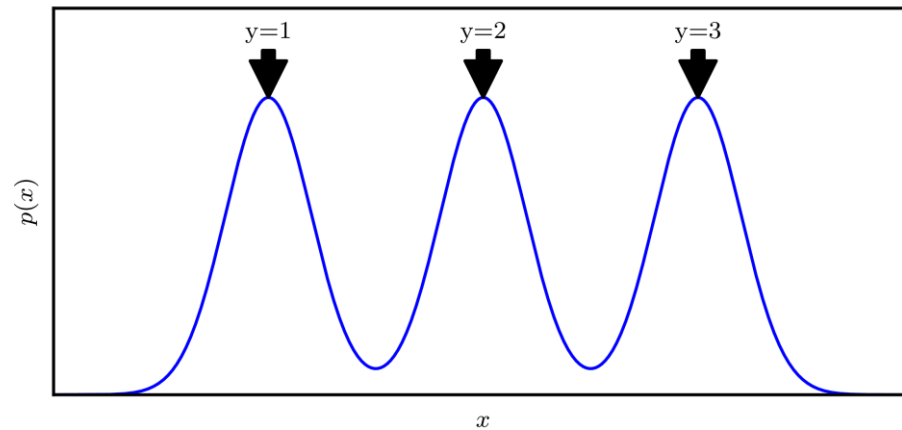
- **Προσαρμογή πεδίου** (*domain adaptation*)
 - Ίδιο ζητούμενο, **ελαφρά μεταβολή** εισόδου
 - Λχ σύστημα ανάλυσης συναισθήματος που έχει εκπαιδευτεί σε κριτικές ταινιών χρησιμοποιείται για την ανάλυση κριτικών μουσικής
 - Η ορολογία μεταβάλλεται μεταξύ των πεδίων, ωστόσο υπάρχει κοινή συνισταμένη για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

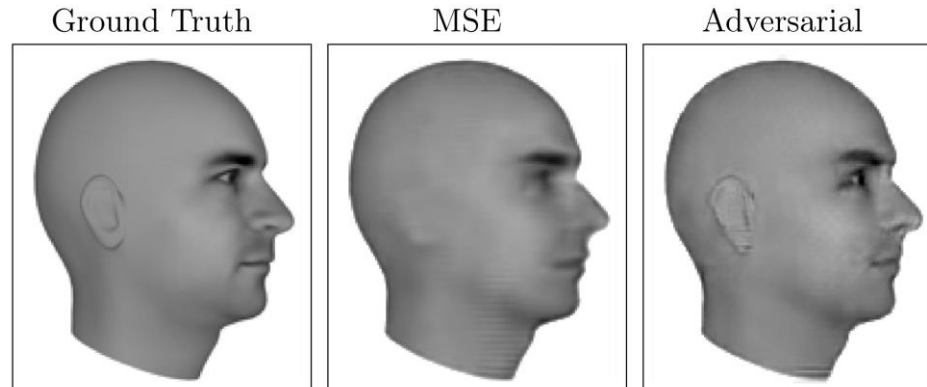
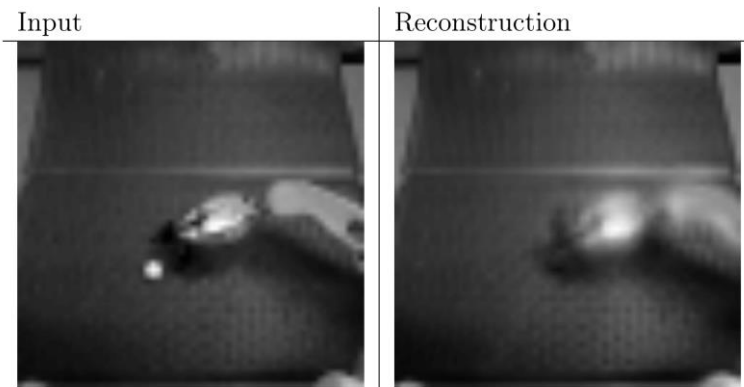


Παράγοντες συνάφειας (casual factors)

- 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 μέσω 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

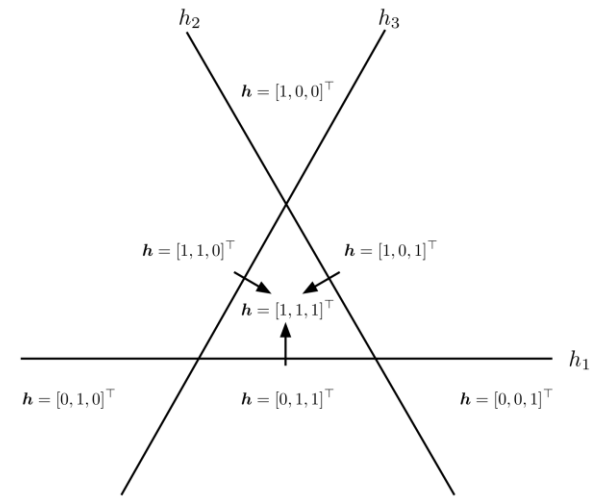


Κατανεμημένες και μη-κατανεμημένες αναπαραστάσεις

- 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**
 - **Σημσιολογικά** κοντινές έννοιες απεικονίζονται σε «**κοντινή**» απόσταση στο χώρο
 - Αυτό **δεν ισχύει** στις συμβολικές αναπαραστάσεις

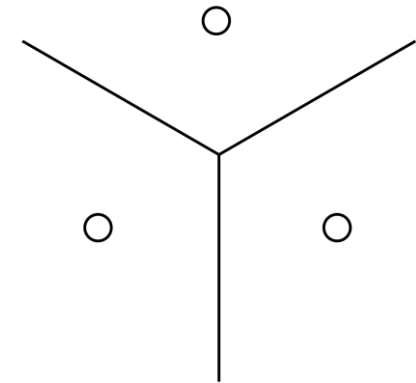
Παράδειγμα κατανεμημένης αναπαράστασης

- 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
- **«Κατάρρα» 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/>)
 - Αυτοκωδικοποιητές (Κεφ. 14)
 - Εισαγωγή (§14.1)
 - Ομαλοποιημένοι Αυτοκωδικοποιητές (§14.2, §14.5, §14.7)
 - Στοχαστικοί Αυτοκωδικοποιητές (§14.4)
 - Εκμάθηση Αναπαραστάσεων (Κεφ. 15)
 - Εισαγωγή (§15.1)
 - Μη-επιβλεπόμενη προπαίδευση (§15.1)
 - Μάθηση μέσω μεταφοράς (§15.2)
 - Παράγοντες συνάφειας (§15.3)
 - Κατανεμημένες και μη-κατανεμημένες αναπαραστάσεις (§15.4)