ε.δε.μ^2

ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ & ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΕΡΓΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Παραγωγικά Μοντέλα μη Επιβλεπόμενης Μάθησης:

1. Μηχανή Boltzmann

2. Παραγωγικά (Generative) Στοχαστικά Νευρωνικά Δίκτυα

3. Generative Adversarial Networks (GAN)

καθ. Βασίλης Μάγκλαρης

maglaris@netmode.ntua.gr

www.netmode.ntua.gr

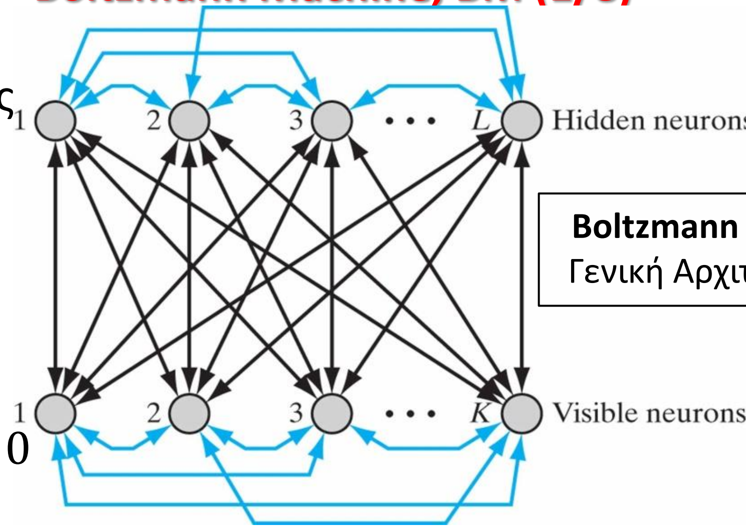
Αίθουσα 002, Νέα Κτίρια ΣΗΜΜΥ

Τρίτη 26/3/2024

Εφαρμογή Δειγματοληψίας Gibbs – Boltzmann Machine, BM (1/8)

(1985, Geoffrey Hinton & Terry Sejnowski)

Στόχος η προσέγγιση **ελλειμματικού** διανύσματος εισόδου (π.χ. ***pattern completion*** εικόνων) μέσω δημιουργίας διανύσματος εξόδου, στατιστικά συμβατού με ***unlabeled*** δείγμα μάθησης



Boltzmann Machine
Γενική Αρχιτεκτονική

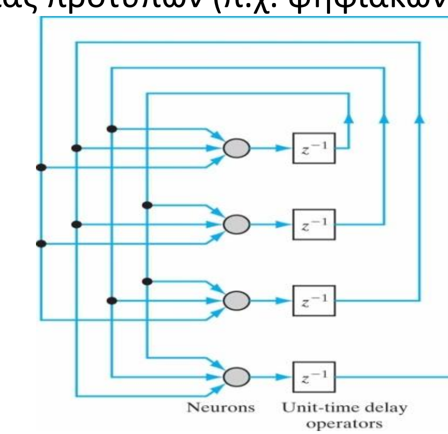
Μια ***Boltzmann Machine*** (**BM**) περιλαμβάνει:

- K **Visible** και L **Hidden Neurons**
- **Συμμετρικές Συνάψεις** $i \rightarrow j$: $w_{ji} = w_{ij}$, $w_{ii} = 0$ εν δυνάμει μεταξύ όλων των νευρώνων της **BM**

Αποτελεί εξέλιξη του αναδρομικού δικτύου ***Hopfield*** με νευρώνες σε **δυαδικές καταστάσεις** ± 1 σύμφωνα με ορισμένες **πιθανότητες** (***Stochastic Recurrent Network with Hidden Nodes***). Το δίκτυο συγκλίνει με **μη επιβλεπόμενη** μάθηση σε ισορροπία ***Markov Random Field***:

- Δυαδικά παραδείγματα **μάθησης** εισάγονται στα ***Visible Nodes*** και με **gradient ascent** προσδιορίζονται συναπτικά βάρη και **τελικές** καταστάσεις των νευρώνων (***Visible*** & ***Hidden***)
- Δυαδικά στοιχεία **test** εισάγονται στα ***Visible Nodes*** και η **BM** τα αναπαράγει (***generates, samples***) σύμφωνα με τις στατιστικές ιδιότητες του **δείγματος μάθησης**

<u>Νευρωνικό Δίκτυο Hopfield</u> (1982, ***John Hopfield***)
Δυαδικοί μη στοχαστικοί νευρώνες με αναδρομικές συμμετρικές συνάψεις, threshold activation και ***supervised learning*** για προσδιορισμό των $w_{ji} = w_{ij}$, $w_{ii} = 0$, συμβατών με το αξίωμα του ***Hebb*** σε κατάσταση ισορροπίας (τοπικό ελάχιστο της ***ενέργειας του συστήματος***). Εφαρμογές ταξινόμησης - επεξεργασίας προτύπων (π.χ. ψηφιακών εικόνων)



Εφαρμογή Δειγματοληψίας Gibbs – Boltzmann Machine, BM (2/8)

Δειγματοληψία Gibbs - Baysean Statistics

https://jwmi.github.io/BMS/chapter6-gibbs-sampling.pdf

- **Ορισμοί Sampling**:
  - Παραγωγή δειγματικών στοιχείων (***Sample Elements***) με στατιστικές ιδιότητες συμβατές με Δειγματικό Χώρο (***Sample Space***) → Δημιουργία ή/και εμπλουτισμός υποσυνόλου (***Δείγμα***, ***Sample***) στοιχείων
  - Επιλογή δειγματικών στοιχείων (***Sample Elements***) του Δειγματικού Χώρου σε υποσύνολο (***Δείγμα***, ***Sample***) για στατιστική ανάλυση που κατ' εκτίμηση γενικεύεται στον ευρύτερο χώρο (έρευνα προτιμήσεων πελατών, δημοσκοπήσεις…)
- **Δειγματοληψία Πολυδιάστατων Δειγματικών Χώρων**: ***Gibbs Sampling***
  - Δημιουργία (*συσχετισμένου*) Δείγματος μέσω ***Monte Carlo*** προσομοιώσεων πολυδιάστατης διακριτής κατάστασης ***Markov*** με χρονοσταθερή κατανομή ισορροπίας που ορίζεται από τις πιθανότητες μετάβασης σε κάθε βήμα εξέλιξης: Τροποποίηση αλγορίθμου ***Metropolis*** για ***διανυσματικές*** καταστάσεις
  - Παράδειγμα Αλγορίθμου για **2 Διαστάσεις**: Η Δειγματοληψία της $\mathbf{X}=[X\ Y]^{\mathrm{T}}$με ***Joint Probabilities*** $P(X,Y)$ δημιουργεί n στοιχεία $\mathbf{x}(n)=[x(n)\ y(n)]^{\mathrm{T}}$ σύμφωνα με πιο βατές υπό συνθήκη πιθανότητες (***Conditional Posterior Probabilities***, ***Bayes***):

$$x(n) \sim P(X|y(n-1) \text{ και } y(n) \sim P(Y|x(n))$$

  - Γενίκευση για $\boldsymbol{K}$ **Διαστάσεις**: Δειγματοληψία της $\mathbf{X}=[X_1\ X_2\ \cdots X_K\ ]^{\mathrm{T}}$

    $n=0$: Αυθαίρετη αρχικοποίηση $\mathbf{x}(0)=[x_1(0)\ x_2(0)\ \ldots\ x_K(0)]^{\mathrm{T}}$

    $n \to n+1$: Κατά σειρά για $i=1,\ldots,K$ δημιουργείται η $x_i(n+1)$ με πιθανότητα $P[X_i(n+1)|\{x_1(n+1)\ \ldots\ x_{i-1}(n+1)\ x_{i+1}(n)\ \ldots\ x_K(n)\}]$

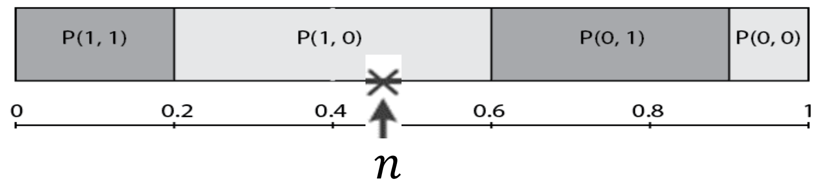
ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εφαρμογή Δειγματοληψίας Gibbs – Boltzmann Machine, BM (3/8)

Παράδειγμα Δισδιάστατης Δειγματοληψίας Gibbs

https://www.baeldung.com/cs/gibbs-sampling

- Δειγματικός Χώρος 2 Διαστάσεων $[X\ Y]^{\mathrm{T}}$ με **δυαδικές** Τυχαίες Μεταβλητές $x, y \in \{0,1\}$
- Οι συνδυασμένες πιθανότητες (***joint probabilities***) $P(X, Y)$ έχουν ως εξής:
  - $P(X=1, Y=1) = 0.2, \quad P(X=0, Y=1) = 0.3$
  - $P(X=1, Y=0) = 0.4, \quad P(X=0, Y=0) = 0.1$
- Ζητείται παραγωγή (***sampling***) n δειγματικών στοιχείων: $[x(n)\ y(n)]^{\mathrm{T}} \sim P(X, Y)$
- **Direct Sampling**: Επαναλαμβάνουμε πειράματα ***Monte Carlo*** παραγωγής ομοιόμορφα κατανεμημένων (ψευτο)τυχαίων αριθμών $0 < n \leq 1$ για προσδιορισμό ακολουθίας δειγματικών στοιχείων ανάλογα με την τιμή της n, π.χ. αν $0.2 < n \leq 0.6$ προκύπτει το δειγματικό στοιχείο $[1\ 0]^{\mathrm{T}}$



- **Gibbs Sampling:** Εναλλακτικά χρησιμοποιούμε τις ***conditional probabilities*** (***Bayes***)
  - $P(X=1|Y=1) = \frac{P(X=1,Y=1)}{P(Y=1)} = \frac{P(X=1,Y=1)}{P(X=1,Y=1)+P(X=0,Y=1)} = \frac{0.2}{0.2+0.3} = 0.4$
  - $P(X=1|Y=0) = 0.8$, $P(X=0|Y=0) = 0.2$, $P(Y=1|X=1) = 0.333$, $P(Y=0|X=1) = 0.666$, $P(Y=1|X=0) = 0.75$, $P(Y=0|X=0) = 0.25$

Εφαρμογή Αλγορίθμου Gibbs Sampling

- Αρχικοποίηση : $x(0) = 1$ $\quad y(0) = 0$
- Βήμα 1 : $x(1) \sim P(X|y(0))$ $\quad y(1) \sim P(Y|x(1))$
- Βήμα 2: $x(2) \sim P(X|y(1))$ $\quad y(2) \sim P(Y|x(2))$
- Βήμα n : $x(n) \sim P(X|y(n-1))$ $\quad y(n) \sim P(Y|x(n))$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εφαρμογή Δειγματοληψίας Gibbs – Boltzmann Machine, BM (4/8)

Φάσεις Μάθησης Μηχανής Boltzmann

- **Θετική Φάση Μάθησης**: Τα στοιχεία του ***δείγματος μάθησης*** $\mathcal{T}$ κλειδώνουν (***clamp***) σε δυαδικές καταστάσεις ± 1 των K ***ορατών νευρώνων*** με βάση τις τιμές γνωστών χαρακτηριστικών τους. Μέσω του προσδιορισμού των συναπτικών βαρών η **BM** κωδικοποιεί στους L ***κρυφούς νευρώνες*** στατιστικές ιδιότητες ανώτερης τάξεως (π.χ. συσχετίσεις) με οριακές πιθανότητες (***marginal distribution***) καταστάσεων ***Gibbs*** υπό τη συνθήκη κλειδωμένων καταστάσεων των K ορατών νευρώνων

- **Αρνητική Φάση Ελεύθερης Επεξεργασίας**: Σε δεύτερη φάση, οι νευρώνες (***ορατοί*** και ***κρυφοί***) αλληλεπιδρούν ελεύθερα χωρίς εξάρτηση από το δείγμα μάθησης $\mathcal{T}$ και ορίζουν συναπτικά βάρη που οδηγούν τη **BM** προς καταστάσεις *θερμικής ισορροπίας* (***Gibbs***). Οι τελικές καταστάσεις των ορατών νευρώνων ***παράγουν*** (στην έξοδο) ***νέα*** δειγματικά στοιχεία με οριακές πιθανότητες χαρακτηριστικών συμβατές με το $\mathcal{T}$

- **Πολυπλοκότητα Αλγορίθμου**: Συνήθως απαιτείται μεγάλος αριθμός κρυφών νευρώνων ***hyperparameter*** $L \gg K$ για κωδικοποίηση σύνθετων στατιστικών ιδιοτήτων χαρακτηριστικών πολυμόρφου δείγματος, καθώς και πολλές επαναλήψεις για ικανοποιητική σύγκλιση των συμμετρικών συνάψεων $w_{ij} = w_{ji}, w_{ii} = 0$ μεταξύ όλων των $L + K$ νευρώνων

- **Αναλογία με Φυσιολογικά Νευρολογικά Συστήματα**: Ενίσχυση συνάψεων μεταξύ ενεργών νευρώνων (αξίωμα ***Hebb***). *Θετική Φάση* $\sim$ Ενεργή Εγκεφαλική Λειτουργία, *Αρνητική Φάση* $\sim$ Επεξεργασία σε Κατάσταση Ύπνου (;)

Εφαρμογή Δειγματοληψίας Gibbs – Boltzmann Machine, BM (5/8)

Ορισμοί

- **Κατάσταση Δικτύου**: Τυχαίο Διάνυσμά $\mathbf{X} \rightarrow \mathbf{x} = [x_1\ x_2\ \dots\ x_K\ \dots\ x_m]^{\mathrm{T}}, m = L + K$
  $x_i \in \{-1,1\} \triangleq \{OFF, ON\}$ όπου x_i η κατάσταση του ***στοχαστικού*** νευρώνα i

- **Κατάσταση των K Ορατών** & L **Κρυφών Νευρώνων**: $\mathbf{X}_\alpha \rightarrow \mathbf{x}_\alpha,\ \mathbf{X}_\beta \rightarrow \mathbf{x}_\beta,\ \mathbf{x} = (\mathbf{x}_\alpha, \mathbf{x}_\beta)$

- **Συναπτικά Βάρη** $i \rightarrow j : w_{ji} = w_{ij}, w_{ii} = 0$ (πιθανή εξωτερική επίδραση ***bias*** στον κόμβο j θεωρείται ότι εισάγεται από κόμβο 0 σε κατάσταση ON με βάρος w_{j0})

- **Ενέργεια Κατάστασης BM**: $E(\mathbf{x}) \triangleq -\frac{1}{2}\sum_i \sum_{j \neq i} w_{ji} x_i x_j$ για $\mathbf{x}$ με στοιχεία $x_i \in \{-1,1\}$
  (αναλογία με θερμοδυναμική)

- **Πιθανότητες Θερμικής Ισορροπίας**: $P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z}\exp\left(-\frac{E(\mathbf{x})}{T}\right)$, κατανομή ***Gibbs/Boltzmann***

- **Κατάσταση των K Ορατών Νευρώνων**: $\mathbf{X}_\alpha \rightarrow \mathbf{x}_\alpha = [x_1\ x_2\ \dots\ x_K\ \dots\ x_K]^{\mathrm{T}}$
  Η κατάσταση του ορατού νευρώνα i αντιστοιχεί σε δυαδικό χαρακτηριστικό (***feature***) του στοιχείου εισόδου/εξόδου i με πιθανότητα να είναι ON ίση με $P(x_i = 1)$

https://www.cs.toronto.edu/~hinton/csc321/readings/boltz321.pdf
https://youtu.be/5jaBneYd5Ig

Εφαρμογή Δειγματοληψίας Gibbs – Boltzmann Machine, BM (6/8)

- **Συμβάντα (Events) για Διανυσματικό Δείγμα με m Διαστάσεις**:

  Για τυχαίο στοιχείο $\left[X_1 = x_1 \;\; X_2 = x_2 \;\; \dots \;\; X_j = x_j \;\; \dots \;\; X_m = x_m\right]^{\mathrm{T}}$ ορίζουμε τα ***events***

  A: $(X_j = x_j)$, B: $(X_1 = x_1, \dots, \; X_{j-1} = x_{j-1}, X_{j+1} = x_{j+1}, \dots, \; X_m = x_m)$ και

  C: $(X_1 = x_1, \; \dots, X_j = x_j, \; \dots, X_m = x_m)$ το joint event των A και B

  Σε θερμική ισορροπία και για X_j που προκύπτουν από τη δειγματοληψία ***Gibbs*** έχουμε:

$$P(C) = P(A, B) = \frac{1}{Z}\exp\left(\frac{1}{2T}\sum_i \sum_{j \neq i} w_{ji} x_i x_j\right)$$

$$P(B) = \sum_A P(A, B) = \frac{1}{Z}\sum_{x_j}\exp\left(\frac{1}{2T}\sum_{i \neq j}\sum_j w_{ji} x_i x_j\right)$$

- **Υπό Συνθήκη Πιθανότητες Μεταβάσεων**:

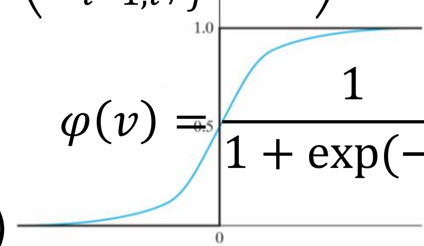
  Δεδομένου ότι x_i, x_j παίρνουν τις τιμές ± 1 η υπό συνθήκη πιθανότητα $P(A|B)$ απλοποιείται:

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{1}{1 + \exp\left(-\frac{x_j}{T}\sum_{i \neq j} w_{ji} x_i\right)}$$

$$P\left(X_j = x \middle| \{X_1 = x_1, \dots, \; X_{j-1} = x_{j-1}, \; X_{j+1} = x_{j+1}, \dots, X_m = x_m\}\right) = \varphi\left(\frac{x}{T}\sum_{i=1, i \neq j}^{m} w_{ji} x_i\right)$$

  όπου $\varphi(\cdot)$ η σιγμοειδής (***logistic***) συνάρτηση $\varphi(v)$

  Η $P(A, B)$ προκύπτει σαν αποτέλεσμα της δειγματοληψίας ***Gibbs*** από αρχική κατάσταση $\mathbf{x}(0)$ με διαδοχικές επισκέψεις $\mathbf{x}(n) \rightarrow \mathbf{x}(n+1)$ λαμβάνοντας υπόψη τις πιο πρόσφατες ανανεώσεις των $x_i(n)$ και διαδοχικά μειώνοντας την θερμοκρασία $T \rightarrow 0$ (***Simulated Annealing***)



$$\varphi(v) = \frac{1}{1 + \exp(-v)}$$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εφαρμογή Δειγματοληψίας Gibbs – Boltzmann Machine, BM (7/8)

Κανόνας Μάθησης Boltzmann

Εφαρμογή κριτηρίου Maximum Likelihood ή Log Likelihood

Το διάνυσμα της κατάστασης $\mathbf{x}$ αποτελείται από τη συρραφή δύο υποσυνόλων: Το υποσύνολο των καταστάσεων των ορατών νευρώνων $\mathbf{x}_\alpha$ και των κρυφών νευρώνων $\mathbf{x}_\beta$ με πιθανότητες που θεωρούμε πως συγκλίνουν σε οριακές πιθανότητες θερμικής ισορροπίας ***Gibbs***

Η λειτουργία της **BM** προχωρά σε δύο φάσεις:

- **Θετική Φάση** που καθορίζεται από τις συνθήκες κλειδώματος (***clamping***) καταστάσεων των ορατών νευρώνων στα παραδείγματα του δείγματος μάθησης $\mathcal{T}$
- **Αρνητική Φάση** όπου το δίκτυο λειτουργεί αυτόνομα χωρίς εισόδους από το περιβάλλον

Από τον ορισμό των συναπτικών βαρών w_{ji} (στοιχεία της μήτρας $\mathbf{w}$ όλου του δικτύου) προκύπτουν οι οριακές πιθανότητες ισορροπίας ***Gibbs*** των ***ορατών*** καταστάσεων $P(\mathbf{X}_\alpha = \mathbf{x}_\alpha)$ ώστε να προσεγγίζουν την κατανομή του $\mathcal{T}$. Αν έχουμε πολλά στοιχεία στο $\mathcal{T}$ μπορούμε να θεωρήσουμε ότι οι καταστάσεις $\mathbf{X}_\alpha$ είναι ***ανεξάρτητα*** τυχαία διανύσματα με συνολική πιθανότητα ισορροπίας το ***παραγοντικό γινόμενο*** $\prod_{\mathbf{x}_\alpha \in \mathcal{T}} P(\mathbf{X}_\alpha = \mathbf{x}_\alpha)$

Αν θεωρήσουμε τον λογάριθμο $L(\mathbf{w})$ του γινομένου έχουμε

$$L(\mathbf{w}) = \log \prod_{\mathbf{x}_\alpha \in \mathcal{T}} P(\mathbf{X}_\alpha = \mathbf{x}_\alpha) = \sum_{\mathbf{x}_\alpha \in \mathcal{T}} \log P(\mathbf{X}_\alpha = \mathbf{x}_\alpha)$$

Οι $P(\mathbf{X}_\alpha = \mathbf{x}_\alpha)$ συμπεριλαμβάνουν τις πιθανότητες των καταστάσεων $\mathbf{x} = (\mathbf{x}_\alpha, \mathbf{x}_\beta)$, $\forall\, \mathbf{x}_\beta$:

$$P(\mathbf{X}_\alpha = \mathbf{x}_\alpha) = \frac{1}{Z} \sum_{\mathbf{x}_\beta} \exp\left(-\frac{E(\mathbf{x})}{T}\right), \qquad Z = \sum_{\mathbf{x}} \exp\left(-\frac{E(\mathbf{x})}{T}\right)$$

Εφαρμογή Δειγματοληψία Gibbs – Boltzmann Machine, BM (8/8)

Κανόνας Μάθησης Boltzmann

Εφαρμογή κριτηρίου Maximum Likelihood ή Log Likelihood (συνέχεια)

Προκύπτει επομένως για τον λογάριθμο του παραγοντικού γινομένου

$$L(\mathbf{w}) = \sum_{\mathbf{x}_\alpha \in \boldsymbol{\mathcal{T}}} \left(\log \sum_{\mathbf{x}_\beta} \exp\left(-\frac{E(\mathbf{x})}{T}\right) - \log \sum_{\mathbf{x}} \exp\left(-\frac{E(\mathbf{x})}{T}\right)\right), \; E(\mathbf{x}) = -\frac{1}{2}\sum_i \sum_{j \neq i} w_{ji} x_i x_j$$

Παραγωγίζοντας ως προς τα συναπτικά βάρη w_{ji} έχουμε

$$\frac{\partial L(\mathbf{w})}{\partial w_{ji}} = \frac{1}{T} \sum_{\mathbf{x}_\alpha \in \boldsymbol{\mathcal{T}}} \left(\sum_{\mathbf{x}_\beta} P(\mathbf{X}_\beta = \mathbf{x}_\beta | \mathbf{X}_\alpha = \mathbf{x}_\alpha) x_j x_i - \sum_{\mathbf{x}} P(\mathbf{X} = \mathbf{x}) x_j x_i \right) \triangleq \frac{1}{T}\left(\rho_{ji}^+ - \rho_{ji}^-\right)$$

Το ρ_{ji}^+ υποδηλώνει τον μέσο ρυθμό ενεργοποίησης (***firing rate***) ή τη συσχέτιση (***correlation***) μεταξύ των καταστάσεων των νευρώνων $j \leftrightarrow i$ στη **Θετική Φάση** και το ρ_{ji}^- τη συσχέτιση (***correlation***) μεταξύ των καταστάσεων των νευρώνων $j \leftrightarrow i$ στη **Αρνητική Φάση**

Ο κανόνας μάθησης Boltzmann (***Boltzmann Learning Rule***) μεγιστοποιεί το $L(\mathbf{w})$ με τη μέθοδο του ***gradient ascent*** με σταθερό βήμα (***hyperparameter***) $\boldsymbol{\epsilon}$:

$$\Delta w_{ji} = \boldsymbol{\epsilon} \frac{\partial L(\mathbf{w})}{\partial w_{ji}} = \eta\left(\rho_{ji}^+ - \rho_{ji}^-\right)$$

Η ***learning rate*** $\eta = \frac{\boldsymbol{\epsilon}}{T}$ μεταβάλλεται σε διαδοχικές επαναλήψεις ***Simulated Annealing*** αντιστρόφως ανάλογα με τη μειούμενη T. Τα βάρη ανανεώνονται με βάση όλα τα στοιχεία του δείγματος μάθησης (***batch mode***) με μεγάλη πολυπλοκότητα και αργή σύγκλιση ⇒

ΑΝΑΓΚΗ ΑΠΛΟΠΟΙΗΣΗΣ ΔΙΚΤΥΟΥ → RESTRICTED BOLTZMANN MACHNE (RBM)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Στατιστική Ταξινόμηση: Generative & Discriminative Models

Παραδοσιακό Διακριτικό Μοντέλο (Discriminative Model) Στατιστικής Ταξινόμησης

Observable Input Data: x **Target Output Labels**: y

Απ’ ευθείας εκτίμηση $P(y|x)$ από δεδομένα του δείγματος μάθησης και ανάθεση της πιθανότερης y σε data x με βάση τις εμφανίσεις της y ***υπό συνθήκη*** x που μετρήθηκαν στη φάση της (***επιβλεπόμενης***) μάθησης, π.χ. ***Logistic Regression*** και ***Back-Propagation Algorithm***

Παραγωγικό Μοντέλο (Generative Model) Στατιστικής Ταξινόμησης

Observable Input Data: x **Target Output Labels**: y

Εκίμηση $P(x,y)$ με βάση ***συνδυασμένες*** στατιστικές παραδοχές εμφάνισης των x και y, υπολογισμός υπό συνθήκη πιθανοτήτων $P(y|x)=\frac{P(x,y)}{P(x)}$, $P(x)=\sum_y P(x,y)$ από κανόνα ***Bayes*** και ανάθεση της πιθανότερης y σε data x. Τα ζεύγη x,y ***δημιουργούνται*** σύμφωνα με τις εμπειρικές $P(x,y)$ όπως αυτές εκτιμήθηκαν από το δείγμα μάθησης ώστε να προσεγγίζουν τα στατιστικά χαρακτηριστικά συγκεκριμένων εφαρμογών ταξινόμησης δεδομένων

Παράδειγμα: $x\in\{1,2\}, y\in\{0,1\}$ (https://en.wikipedia.org/wiki/Generative_model)

$P(x,y)$	$y=0$	$y=1$
$x=1$	$1/2$	0
$x=2$	$1/6$	$2/6$

$\Rightarrow$

$P(y\|x)$	$y=0$	$y=1$
$x=1$	$\mathbf{1}$	0
$x=2$	$2/6$	$\mathbf{4/6}$

$P(x=1)=1/2,\ P(x=2)=3/6=1/2$

Προτιμάται για περιπτώσεις που τα δεδομένα παρουσιάζουν ελλείψεις (π.χ. κενά σε εικόνες ή δυσδιάκριτα σήματα φωνής) τις οποίες το σύστημα μάθησης καλείται να μαντέψει με βάση μοντέλα στατιστικών ***συσχετίσεων*** χαρακτηριστικών τους, π.χ. ***Boltzmann Machine***

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Στατιστική Προσέγγιση: Generative & Discriminative Models (1/2)

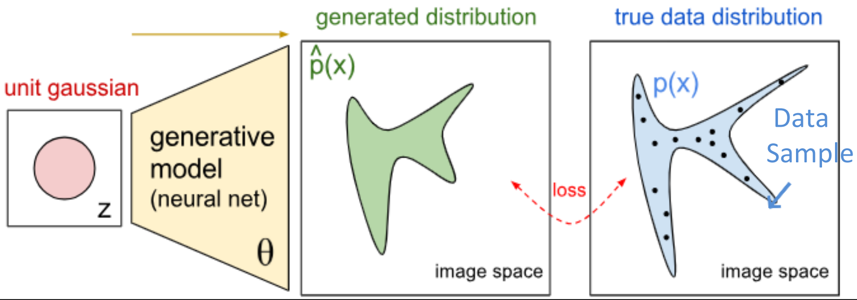
Γενίκευση Παραγωγικού Μοντέλου (Generative Model)

https://openai.com/blog/generative-models/

$p(x)$: Κατανομή των στοιχείων του δείγματος μάθησης (***Training Sample***) $\{x_1, x_2, \dots, x_n\}$

$\hat{p}_\theta(x)$: Κατανομή των εικονικών στοιχείων του παραγόμενου δείγματος (***Generated Sample***) στην **έξοδο** νευρωνικού δικτύου παραμέτρων θ με αυθαίρετο δείγμα **εισόδου**, π.χ. 100 τυχαίοι αριθμοί με κανονική κατανομή, ***Gaussian Sample*** Z

Διαδικασία Μάθησης: Ρύθμιση παραμέτρων θ νευρωνικού δικτύου με βάση δεδομένα μάθησης (***Training Sample***) ώστε $\hat{p}_\theta(x) \to p(x)$ (συνήθως κατά ***Kullback-Leibler***)



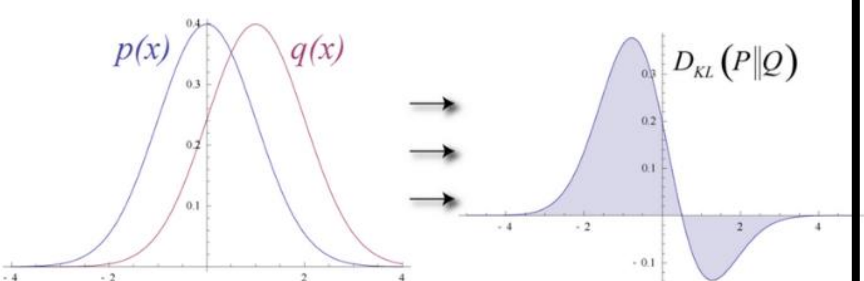
Μετρικές Ομοιότητας Κατανομών $p(x), q(x)$

- **Divergence Kullback-Leibler** (***KL***) (1951):

  $$D_{\mathrm{KL}}(P||Q) = \sum_{x\in\mathcal{T}} P(x)\log\frac{Q(x)}{P(x)}$$

  (*εφαρμόζεται σε Boltzmann Machine*)

  https://skymind.ai/wiki/restricted-boltzmann-machine



- **Expectation-Maximization** (***EM***) **Algorithm** :

  Επαναλήψεις δύο σταδίων για προσδιορισμό λανθανουσών (***latent***) παραμέτρων:

  (π.χ. προσδιορισμός ποσοστών μείξης τυχαίων μεταβλητών από 2 ανεξάρτητα δείγματα Gauss)

  https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Στατιστική Προσέγγιση: Generative & Discriminative Models (2/2)

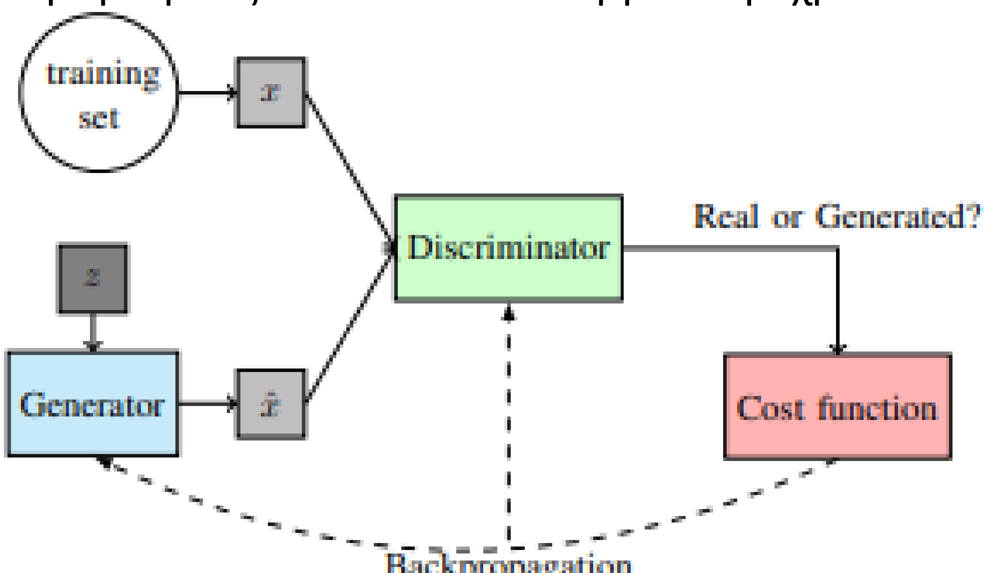
Generative Adversarial Networks - GAN (2014 Ian Goodfellow et.al.)

https://arxiv.org/pdf/1406.2661.pdf

Συνδυασμός ανεξάρτητης επεξεργασίας από δύο *παίκτες* σε ***zero-sum adversarial min-max game*** μεταξύ ***παραγόμενου εικονικού δείγματος*** και ***αληθινού δείγματος .*** Η ***μάθηση*** βασίζεται σε δυο βαθιά νευρωνικά δίκτυα τύπου Multilayer Perceptron - ***MLP***:

- **Generator** (**G**) που με είσοδο ***latent random variables*** z (π.χ. ***Gauss***) δημιουργεί στην έξοδο $G(z)$ εικονικό παραγόμενο (***generated***) δείγμα $\hat{x}$ με κατανομή $p_\theta(\hat{x})$
- **Discriminator** (**D**) που προσπαθεί να ταξινομήσει με ***επιβλεπόμενη μάθηση*** τη διαφορά μεταξύ ***αληθινών δεδομένων μάθησης*** $x\sim p(x)$ και ***εικονικών δεδομένων*** $\hat{x}\sim p_\theta(\hat{x})$

Όσο ο **D** καταλαβαίνει τη διαφορά (έξοδος ***Generated***), ο παίκτης **G** τροποποιεί τις παραμέτρους του και επαναλαμβάνει μέχρι να τον εξαπατήσει (έξοδος ***Real***)



Cost Functions (Loss) for D - G Game:

D: $max\ \left\{\log D(x) + \log\left(1 - D\big(G(z)\big)\right)\right\}$

maximize probability $\hat{x}$ classified as fake

G: $min\left\{\log\left(1 - D\big(G(z)\big)\right)\right\}$

minimize probability $\hat{x}$ classified as fake

Εφαρμογές: Computer vision, virtual reality, computer graphics, interactive games, scientific simulations, ….

http://sibgrapi.sid.inpe.br/col/sid.inpe.br/sibgrapi/2017/09.05.22.09/doc/_2017_sibgrapi__Tutorial_Deep_Learning_for_CV___Survey_Paper_CRP.pdf