



Ταξινόμηση προτύπων

2

Προβλήματα στην ταξινόμηση με νευρωνικά δίκτυα

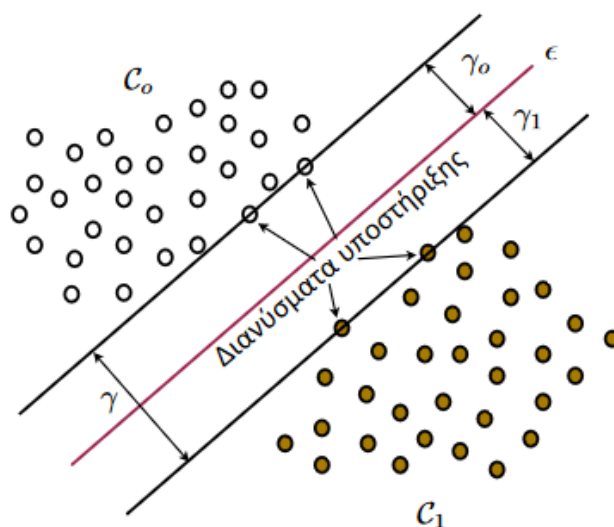
- Η ταξινόμηση με perceptrons δουλεύει μόνο με γραμμικά διαχωρίσιμες κλάσεις
- Η ταξινόμηση με δίκτυα MLP υποφέρει από βραδεία εκπαίδευση (μην ξεχνάμε ότι στην περίπτωση αυτή λύνουμε γενικότερο πρόβλημα από αυτό της ταξινόμησης)

Ιδέα

- Αν επικεντρωθούμε στο πρόβλημα της ταξινόμησης μπορούμε να πετύχουμε καλύτερους χρόνους εκπαίδευσης και καλύτερες ιδιότητες γενίκευσης

9

Βέλτιστο υπερεπίπεδο διαχωρισμού



$$\gamma_0 = \min_{x \in C_0} d(x, \epsilon)$$

$$\gamma_1 = \min_{x \in C_1} d(x, \epsilon)$$

$$\gamma = \gamma_0 + \gamma_1$$

Κανονικό υπερεπίπεδο

$$\gamma_0 = \gamma_1$$

$$w^T x_i + w_0 \geq 1 \text{ αν } x_i \in C_0$$

$$w^T x_i + w_0 \leq -1 \text{ αν } x_i \in C_1$$

Ορίζουμε τη συνάρτηση κόστους:

$$\mathcal{L}(\mathbf{w}, w_o, \lambda_1, \dots, \lambda_P) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^P \lambda_i [d_i(\mathbf{w}^\top \mathbf{x}_i + w_o) - 1]$$

$$\text{με } \lambda_i \geq 0, i = 1, \dots, P$$

Η συνάρτηση αυτή πρέπει να ελαχιστοποιηθεί ως προς τα \mathbf{w} , w_o και να μεγιστοποιηθεί ως προς τα λ_i

Από τις συνθήκες KKT έχουμε:

$$\frac{\partial L}{\partial w_o} = 0 \longrightarrow \sum_{i=1}^P \lambda_i d_i = 0$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \longrightarrow \mathbf{w} = \sum_{i=1}^P \lambda_i d_i \mathbf{x}_i$$

Συνεπώς η βέλτιστη διαχωριστική επιφάνεια δίνεται από τη σχέση:

$$g^*(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + w_o = \sum_{i=1}^P \lambda_i d_i \mathbf{x}_i^\top \mathbf{x} + w_o$$

Για τα διανύσματα υποστήριξης ισχύει ότι:

$$d_i(\mathbf{w}^\top \mathbf{x}_i + w_o) = 1 \longrightarrow w_o = \frac{1}{d_i} - \mathbf{w}^\top \mathbf{x}_i$$

Παρατήρηση

Οι μόνοι πολλαπλασιαστές λ_i που μπορούν να είναι θετικοί είναι αυτοί που αντιστοιχούν σε κάποιο διάνυσμα υποστήριξης \mathbf{x}_i .
Για τους υπόλοιπους ισχύει $\lambda_i = 0$.

Δυϊκό πρόβλημα (2)

6

Ορισμός δυϊκού προβλήματος βελτιστοποίησης

Υπολόγισε το ελάχιστο της συνάρτησης:

$$\mathcal{L}^d(\lambda_1, \dots, \lambda_P) = \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P \lambda_i \lambda_j d_i d_j \mathbf{x}_i^\top \mathbf{x}_j - \sum_{i=1}^P \lambda_i$$

ως προς τα $\lambda_1, \dots, \lambda_P$, υπό τους περιορισμούς

$$\sum_{i=1}^P \lambda_i d_i = 0 \quad \lambda_i \geq 0, i = 1, \dots, P$$

Μεταβλητές χαλαρότητας

Ορίζουμε ένα σύνολο $\{\xi_i\}_{i=1}^N$ από θετικές τιμές και τις εισάγουμε στην εξίσωση της βέλτιστης ευθείας διαχωρισμού ως εξής:

$$d_i(\mathbf{w}^T \mathbf{x}_i + w_o) \geq 1 - \xi_i, i = 1, \dots, P$$

με $\xi_i \geq 0, i = 1, \dots, P$

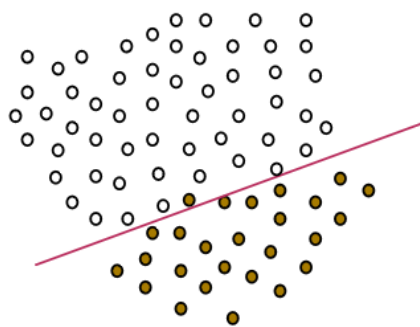
Παρατηρούμε ότι:

Αν $\xi_i \leq 1$ δεν υπάρχει λάθος ταξινόμηση

Αν $\xi_i > 1$ υπάρχει λάθος ταξινόμηση
και το πρότυπο \mathbf{x}_i ταξινομείται σε λάθος κλάση

Μη γραμμικά διαχωρίσιμες κλάσεις

9



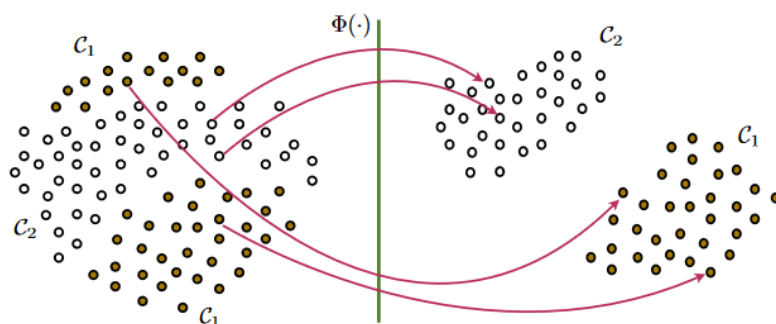
Πέτρα 4-2013

2) μεταβλητές χαλαρότητας → όταν δεν να ταξινομήσεις με διαχωρίσιμη ευθεία
↳ όταν είναι γραμμικά διαχωρίσιμα και δεν να βρεις διαχωριστική ευθεία

Επίσης επηρεάζουν την ευρωστία του περιθωρίου ταξινόμησης σε θορυβώδη δεδομένα, δηλ γίνεται καλύτερη ταξινόμηση λόγω θορύβου επειδή υπάρχει ελαστικότητα.

Θεώρημα Cover

Κάθε πολυδιάστατος χώρος με μη γραμμικά διαχωρίσιμα πρότυπα, μπορεί να μετασχηματιστεί σε ένα νέο χώρο στον οποίο τα πρότυπα είναι γραμμικά διαχωρίσιμα με *υψηλή πιθανότητα*, αρκεί ο μετασχηματισμός να είναι μη γραμμικός και ο νέος αυτός χώρος να έχει την απαραίτητη διάσταση



α) (10%) Να εξηγήσετε διαισθητικά το λόγο για τον οποίο η αύξηση της διάστασης του χώρου χαρακτηριστικών σε ένα δίκτυο διανυσμάτων υποστήριξης αυξάνει την πιθανότητα να μετατρέψουμε ένα μη γραμμικά διαχωρίσιμο πρόβλημα σε γραμμικά διαχωρίσιμο.

Θέμα 4^ο-2020

λύση του χώρου διαστατικότητας → αμοιβαία φασάνων (όχι 233, διαφήμιση Cover)

Λύση δυϊκού προβλήματος

6

Βέλτιστη διαχωριστική επιφάνεια:

$$g^*(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}) + w_o = \sum_{i=1}^P \lambda_i d_i \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}) + w_o$$

Κατώφλι:

$$w_o = \frac{1}{|I_{sv}|} \sum_{i \in I_{sv}} \left(\frac{1}{d_i} - \mathbf{w}^T \Phi(\mathbf{x}_i) \right)$$

Συνάρτηση κόστους του δυϊκού προβλήματος:

$$\mathcal{L}(\lambda_1, \dots, \lambda_P) = \sum_{i=1}^P \lambda_i - \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P \lambda_i \lambda_j d_i d_j \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$$

Ορισμός

Ορίζουμε τη συνάρτηση $k(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x})^T \Phi(\mathbf{y})$, την οποία θα ονομάζουμε συνάρτηση πυρήνα.

Χρησιμοποιώντας τη συνάρτηση πυρήνα κάνουμε οικονομία πράξεων ειδικά όταν η διάσταση του \mathbf{x} είναι μεγαλύτερη από τη διάσταση του $\Phi(\mathbf{x})$ (το οποίο συνήθως συμβαίνει)

Παράδειγμα

28

$$\text{Έστω } \mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$$

$$\Phi(\mathbf{x}) = \begin{bmatrix} x_1^2 & \sqrt{2}x_1x_2 & x_2^2 \end{bmatrix}^T$$

$$\text{Για } \mathbf{x} = \begin{bmatrix} 1 & 2 \end{bmatrix}^T$$

$$\Phi([1 \ 2]^T) = \begin{bmatrix} 1 & 2\sqrt{2} & 4 \end{bmatrix}^T$$

$$k(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x})^T \Phi(\mathbf{y}) = (x_1^2 y_1^2 + 2x_1 y_1 y_2 + x_2^2 y_2^2)$$

$$= (x_1 y_1 + x_2 y_2) = (\mathbf{x}^T \mathbf{y})^2$$

Πρόβλημα SVM

Υπολόγισε το μέγιστο της συνάρτησης:

$$\mathcal{L}_{SVM}(\lambda_1, \dots, \lambda_P) = \sum_{i=1}^P \lambda_i - \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P \lambda_i q_{ij} \lambda_j$$

υπό τους περιορισμούς:

$$0 \leq \lambda_i \leq C \quad \sum_{i=1}^P \lambda_i d_i = 0$$

όπου: $q_{ij} = d_i d_j k(\mathbf{x}_i, \mathbf{x}_j)$

Παρατήρηση

Το πλήθος των στοιχείων του πίνακα $\mathbf{Q} = [q_{ij}]$ είναι P^2 , συνεπώς είναι αρκετά πολύπλοκη η επίλυση του προβλήματος.

Μέθοδοι υλοποίησης SVM



32

Μέθοδος τεμαχισμού

Η συνάρτηση κόστους δεν αλλάζει αν αφαιρέσουμε τις γραμμές και τις στήλες του \mathbf{Q} που αντιστοιχούν σε μηδενικές τιμές του λ_i

Διαλέγουμε σε κάθε βήμα την επίλυση του προβλήματος για το τμήμα του \mathbf{Q} που αντιστοιχεί στα μη μηδενικά λ_i από το προηγούμενο πρόβλημα και επιπλέον στα K χειρότερα λ_i (που παραβιάζουν περισσότερο τις συνθήκες ΚΚΤ

Μέθοδος Osuna

Αν επιλύσουμε ένα μικρότερο πρόβλημα, επιλέγοντας μερικές μόνο γραμμές του \mathbf{Q} έτσι ώστε να περιέχεται τουλάχιστον ένα λ_i που παραβιάζει τις συνθήκες ΚΚΤ τότε η συνάρτηση κόστους μειώνεται και όλοι οι περιορισμοί συνεχίζουν να ικανοποιούνται

Επιλύουμε το πρόβλημα προσθέτοντας μία μεταβλητή λ_i που παραβιάζει τις συνθήκες και αφαιρώντας μία μεταβλητή για την οποία $\lambda_i = 0$ ή $\lambda_j = C$

ΘΕΜΑ 4

Δίνονται οι παρακάτω προτάσεις που αφορούν την επίλυση ενός γραμμικά διαχωρίσιμου προβλήματος με Μηχανές Διανυσμάτων Υποστήριξης. Ποια από όλες πιστεύετε ότι είναι σωστή;

Α. Τα διανύσματα του συνόλου εκμάθησης που δεν είναι διανύσματα υποστήριξης δεν εμπλέκονται στη διαδικασία της εκτέλεσης του αλγόριθμου προσδιορισμού της ευθείας διαχωρισμού.

☒ Β. Αν μας δίνονταν τα διανύσματα υποστήριξης, δεν θα χρειαζόταν να εκτελέσουμε τον αλγόριθμο, θα υπολογίζαμε άμεσα τη βέλτιστη ευθεία διαχωρισμού από αυτά.

Γ. Τα διανύσματα του συνόλου εκμάθησης που δεν είναι διανύσματα υποστήριξης εμπλέκονται στη διαδικασία της εκτέλεσης του αλγόριθμου προσδιορισμού της ευθείας διαχωρισμού και η τελική ευθεία διαχωρισμού εξαρτάται από αυτά.

Competitive – SOM – LVQ

Φάση ανταγωνισμού

Νικητής

⇒ μέγιστη σταθμισμένη είσοδος
(εσωτερικό γινόμενο)

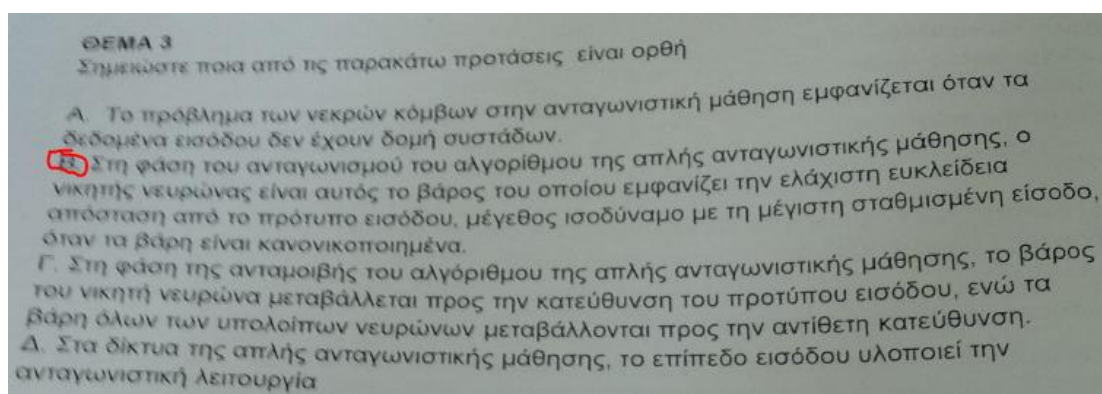
$$\max s_i = \sum_j w_{ij} x_j = \vec{w}_i \vec{x}$$

⇒ ελάχιστη ευκλείδεια απόσταση

$$\min d_i = \left\| \vec{w}_i - \vec{x} \right\|$$

Ισοδύναμα αν τα βάρη

είναι κανονικοποιημένα, π.χ. $\left\| \vec{w}_i \right\| = 1, \quad \forall i$



Για Comp – SOM (Unsupervised)

Φάση ανταμοιβής

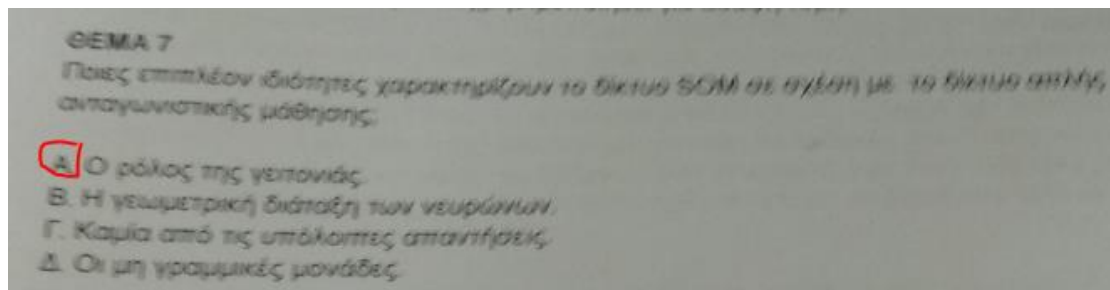
Βασικός κανόνας

$$\vec{w}_i^{(t+1)} = \vec{w}_i^{(t)} + \eta(t) \left[\vec{x} - \vec{w}_i^{(t)} \right]$$

$$0 \leq \eta(t) \leq 1$$

Μπορεί στο SOM να υπάρχει και συντελεστής γειτονιάς γ οπότε η σχέση γίνεται

$$w(t+1) = w(t) + \eta \gamma (x - w(t))$$



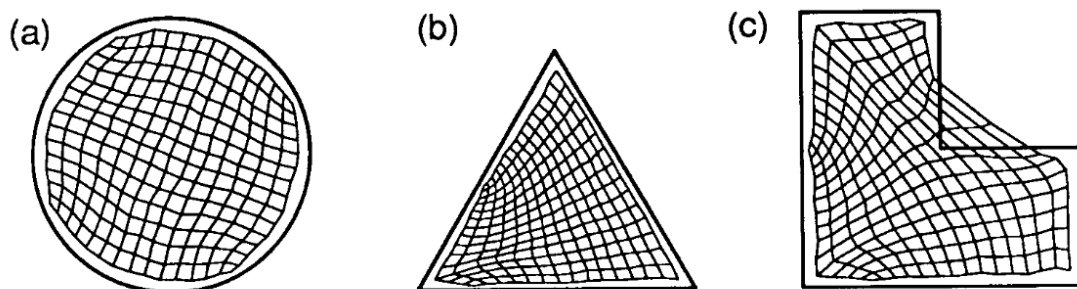
Για LVQ (Supervised)

$$\Delta \mathbf{w}_k(t) = \begin{cases} + \eta(t) (\mathbf{x}(t) - \mathbf{w}_k(t)) & \text{αν η κατηγορία του νικητή είναι σωστή} \\ - \eta(t) (\mathbf{x}(t) - \mathbf{w}_k(t)) & \text{αλλιώς} \end{cases}$$

Δίκτυο SOM

Παράδειγμα

- Πλέγμα εξόδου 15 x 15
- Πρότυπα εισόδου ομοιόμορφα κατανεμημένα σε κύκλο, τρίγωνο και περιοχή σχήματος L



- Μονοδιάστατη έξοδος 50 κόμβων
- Πρότυπα εισόδου ομοιόμορφα κατανεμημένα σε περιοχή σχήματος L

➤ Μείωση διάστασης (dimensionality reduction)

➤ Μερική διατήρηση τοπολογίας

Θέμα 9 - 2015

α) Η Πρώτη γειτονιά → αναζητείται μόνο τα γειτονικά χαρακτηριστικά της εισόδου που έχει η γειτονιά της πρώτης του υπεύθυνου

1.1.3

2) Συνάρτηση γειτονιάς → για να προσαρμόζεται το SOM σε αυτά που του παρέχεται
 αρχική μνήμη → εύρεση σε τοπικό ελάχιστο προσαρμόζεται μόνο πάροδο χρόνου σε
 τοπικό ελάχιστο και όχι όλο, άρα η λειτουργία για να αρχίσει να
 έρθει το υπέρβα και στη συνέχεια τα τοπικά χαρακτηριστικά

β) (4%) Γιατί χρειάζεται η συνάρτηση γειτονιάς στον αλγόριθμο εκπαίδευσης του χάρτη Kohonen; (1%)
 θα μπορούσε να ενημερώνεται μόνο ο νικητής κόμβος; Γιατί το εύρος της γειτονιάς πρέπει να
 μειώνεται με την πάροδο του χρόνου εκπαίδευσης;

Self-organizing maps differ from other artificial neural networks as they apply [competitive learning](#) as opposed to error-correction learning (such as [backpropagation](#) with [gradient descent](#)), and in the sense that they use a neighborhood function to preserve the [topological](#) properties of the input space.

Regardless of the functional form, the neighborhood function shrinks with time.^[9] At the beginning when the neighborhood is broad, the self-organizing takes place on the global scale. When the neighborhood has shrunk to just a couple of neurons, the weights are converging to local estimates.

Δίκτυο SOM: Σύγκλιση

- Σημεία ισορροπίας - τοπικά ελάχιστα

$$\Delta \vec{w}_j = \eta \sum_m \sum_i M(m, i) h_{j, i} \left(\vec{x}^m - \vec{w}_j \right) = 0, \forall j$$

Δύσκολη επίλυση στη γενική περίπτωση
 - Ομοιόμορφη κατανομή προτύπων
 Μονοδιάστατο πλέγμα

- Διαδικασία σύγκλισης:

(i) Εκτύλιξη (untangling)

Συστροφές, στρεβλώσεις

(ii) Λεπτή προσαρμογή (fine tuning)

Μονοδιάστατο πλέγμα: Εκτύλιξη

(a) Κάθε μονότονη περιοχή βαρών παραμένει μονότονη μετά την ενημέρωση.

(b) Το όριο μεταξύ δύο μονότονων περιοχών μπορεί να μετακινηθεί κατά ένα βήμα (προς τη μία ή την άλλη πλευρά) ανά ενημέρωση.

(c) Τα όρια εξαφανίζονται στα άκρα. Δεν δημιουργούνται νέα όρια.

Χρόνος εκτύλιξης

Τυχαίος περίπατος: $O(N^2)$ βήματα για μετακίνηση σε ένα άκρο.

Εκτύλιξη: $O(N^3)$ βήματα συνολικά.

Μπορεί να μειωθεί σε $O(N^2)$ αν η συνάρτηση γειτονιάς δεν είναι συμμετρική.

Ο αλγόριθμος k -μέσων

Σε κάθε βήμα :

- Ταξινόμηση καθενός προτύπου στην ομάδα Ω_k με τη μικρότερη απόσταση

$$d(\vec{x}^i, \vec{w}_k) = \min_j d(\vec{x}^i, \vec{w}_j)$$

- Υπολογισμός των νέων κέντρων των ομάδων

$$\vec{w}_j^{(t+1)} = \frac{1}{N_j^{(t)}} \sum_{\vec{x}^i \in \Omega_j^{(t)}} \vec{x}^i \quad N_j: \text{αριθμός προτύπων στην ομάδα } \Omega_j$$

Συνολική ενημέρωση (batch update) σε κάθε βήμα
 \Rightarrow Κάθοδος κλίσης

Ασαφής ομαδοποίηση (Fuzzy clustering)

- Εύκαμπτη ομαδοποίηση (soft clustering)
 - Κάθε πρότυπο μπορεί να ανήκει σε περισσότερες από μία ομάδες.
 - Βαθμός συμμετοχής (membership degree)
 $u_{ij} \in [0,1]$ του προτύπου i στην ομάδα j

$$\sum_{j=1}^M u_{ij} = 1$$

- Δυνατότητα μετάβασης σε άκαμπτη ομαδοποίηση
Παρέχει περισσότερη πληροφορία (κέντρα ομάδων και βαθμός συμμετοχής).

Ασαφής αλγόριθμος k -μέσων (Fuzzy k -means)

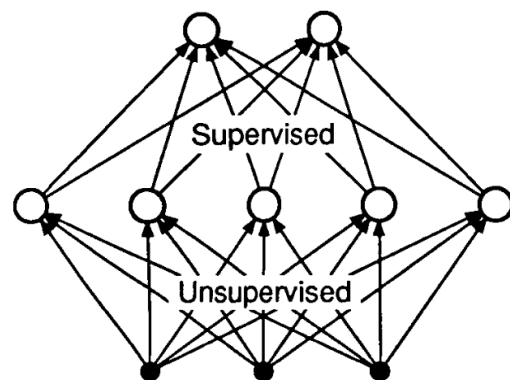
Σε κάθε βήμα :

- Υπολογισμός βαθμών συμμετοχής
- Υπολογισμός των νέων κέντρων των ομάδων

RBF – MLP

Γενική περίπτωση υβριδικού ταξινομητή

- Εκπαίδευση πρώτου (κρυμμένου) στρώματος με ανταγωνιστική μάθηση (είσοδοι x_k , βάρη w_{kj} , τιμές κόμβων z_j)
- Εκπαίδευση στρώματος εξόδου (γραμμική απεικόνιση) με τον κανόνα δέλτα (βάρη v_{ji} , τιμές κόμβων εξόδου y_i)



Εκπαίδευση στρώματος εξόδου με τον κανόνα δέλτα:

$$\Delta v_{ji} = \eta (d_i - y_i) z_j = \eta (d_i - v_{ji}) z_j, \quad 0 < \eta < 1$$

προκύπτει αν πάρεις ενέργεια $E = \frac{1}{2} (d - y)^2$ και χρησιμοποιήσεις τον τύπο του back propagation για ανανέωση βαρών $\Delta w = -\eta \frac{dE}{dw}$

Δίκτυα RBF

- Συνήθως χρησιμοποιούνται (κανονικοποιημένες) γκαουσιανές συναρτήσεις βάσης

$$\varphi_j(\|\vec{x} - \vec{\mu}_j\|) = R_j(\vec{x}) \approx \exp\left(-\frac{\|\vec{x} - \vec{\mu}_j\|^2}{2\sigma_j^2}\right)$$

Δίκτυα RBF

- Εκπαίδευση δύο σταδίων
 - Κρυμμένο επίπεδο
 - ⇒ Καθορισμός κέντρων: ανταγωνιστική μάθηση (π.χ. αλγόριθμος k -μέσων)
 - ⇒ Καθορισμός ακτίνων: ευρετικά (π.χ. μέση τετραγωνική απόσταση από τους p -πλησιέστερους γείτονες)
 - Επίπεδο εξόδου
 - Εκπαίδευση γραμμικού ταξινομητή (κανόνας δέλτα)
- Ενιαία εκπαίδευση
 - Ταυτόχρονη ενημέρωση όλων των παραμέτρων του δικτύου (κάθοδος κλίσης)
 - Χρήση ανταγωνιστικής μάθησης για αρχικοποίηση κρυμμένου επιπέδου

Σύγκριση MLP και RBF

- MLP
 - Κρυμμένες μονάδες: υπερεπίπεδα
 - Κατανεμημένη αναπαράσταση
 - Ανοχή σε σφάλματα
- RBF
 - Κρυμμένες μονάδες: σφαιρικά δεκτικά πεδία
 - Τοπική αναπαράσταση (διαμερισμός χώρου)
 - Υψηλή ταχύτητα εκπαίδευσης

Σύγκριση MLP και RBF

Ομοιότητες

- Μη γραμμικά, στρωματικά δίκτυα πρόσθιας τροφοδότησης
- Καθολικοί προσεγγιστές (universal approximators): Οποιαδήποτε συνάρτηση μπορεί να προσεγγιστεί με αυθαίρετα μικρό σφάλμα εφόσον υπάρχουν αρκετοί κρυμμένοι νευρώνες.
- Παρόμοιες περιοχές εφαρμογών

ΘΕΜΑ 9
Σημειώστε ποια από τις παρακάτω προτάσεις είναι λανθασμένη.

Σε σύγκριση με τα Πολυεπίπεδα Perceptrons (MLP), τα δίκτυα ακτινικών συναρτήσεων βάσης (RBF):

Α. σχηματίζουν δεκτικά πεδία γύρω από τις κρυμμένες μονάδες.
Β. έχουν υψηλότερη ταχύτητα εκπαίδευσης.
Γ. διαμερίζουν τον χώρο της εισόδου, επιτυγχάνοντας τοπικές αναπαραστάσεις των δεδομένων.
Δ. εμφανίζουν μεγαλύτερη ανοχή σε σφάλματα.

Επίλυση XOR με RBF

Συναρτήσεις βάσης: $\varphi_j(\mathbf{x}) = \exp(-\|\mathbf{x} - \mu_j\|^2)$, $j = 1, 2, 3, 4$

Κέντρα: $\mu_1 = [1, 1]$, $\mu_2 = [0, 0]$, $\mu_3 = [0, 1]$, $\mu_4 = [1, 0]$

Έξοδος δικτύου: $y = v_1 \varphi_1(\mathbf{x}) + v_2 \varphi_2(\mathbf{x}) + v_3 \varphi_3(\mathbf{x}) + v_4 \varphi_4(\mathbf{x})$

Πρότυπο εισόδου	$\mathbf{x} = [x_1, x_2]$	$\varphi_1(\mathbf{x})$	$\varphi_2(\mathbf{x})$	$\varphi_3(\mathbf{x})$	$\varphi_4(\mathbf{x})$	Επιθυμητή έξοδος
1	[1, 1]	1	0,1353	0,3678	0,3678	0
2	[0,0]	0,1353	1	0,3678	0,3678	0
3	[0, 1]	0,3678	0,3678	1	0,1353	1
4	[1, 0]	0,3678	0,3678	0,1353	1	1

$$v_1 \cdot 1 + v_2 \cdot 0,1353 + v_3 \cdot 0,3678 + v_4 \cdot 0,3678 = 0$$

$$v_1 \cdot 0,1353 + v_2 \cdot 1 + v_3 \cdot 0,3678 + v_4 \cdot 0,3678 = 0$$

$$v_1 \cdot 0,3678 + v_2 \cdot 0,3678 + v_3 \cdot 1 + v_4 \cdot 0,1353 = 1$$

$$v_1 \cdot 0,3678 + v_2 \cdot 0,3678 + v_3 \cdot 0,1353 + v_4 \cdot 1 = 1$$

$$V = \Phi^{-1} d$$

Τελικά: $v_1 = v_2 = -0,9843$

$$v_3 = v_4 = 1,5188$$

Genetics

Σχήμα (Schema) = Ακολουθία που περιλαμβάνει 0, 1 και "*" ("don't care")

Π.χ. το σχήμα

1 * * 0

παριστάνει το σύνολο
των δυαδικών ακολουθιών
(χρωμοσωμάτων)

στην γλώσσα

1	1	1	0
1	1	0	0
1	0	1	0
1	0	0	0

Άσκηση 1

Θεωρούμε την κωδικοποίηση χρωμοσωμάτων με δυαδικές ακολουθίες μήκους L .

1. Πόσα είναι τα δυνατά σχήματα;
2. Με πόσα σχήματα ταιριάζει ένα δεδομένο χρωμόσωμα;
3. Με πόσα χρωμοσώματα ταιριάζει ένα σχήμα που περιέχει r αστερίσκους;
4. Μπορεί οποιοδήποτε σύνολο χρωμοσωμάτων να περιγραφεί ως σχήμα;
5. Η τάξη ενός σχήματος είναι το πλήθος των ορισμένων (0 ή 1) ψηφίων του. Πόσα είναι τα δυνατά σχήματα τάξης k ;

Να αιτιολογηθούν οι απαντήσεις σας.

Απάντηση

Δίνονται τα αποτελέσματα. (Η αιτιολόγηση αφήνεται σε σας.)

1. 3



vasil 8/26/2017, 11:46:16 PM

2. 2^L

3^L επειδή παίζει και ο αστερίσκος.
Οποτε έχω 3 συμβολα 1,0,* ????

3. 2^r

4. Όχι

5. 2^k

2^{L-k}

Άσκηση 4

Για την επίλυση ενός προβλήματος βελτιστοποίησης θεωρούμε κωδικοποίηση χρωμοσωμάτων με δυαδικές ακολουθίες μήκους L και υποθέτουμε ότι η βέλτιστη λύση αντιστοιχεί στην ακολουθία με '1' σε όλες τις θέσεις. Τα άτομα του αρχικού πληθυσμού περιέχουν σε κάθε θέση ένα από τα σύμβολα '0', '1' και '*' (don't care), με αντίστοιχες πιθανότητες p_0 , p_1 και p^* ($p_0+p_1+p^*=1$). Μια ακολουθία που περιέχει μόνο '1' και '*' χαρακτηρίζεται ως «δυναμικό βέλτιστο».

1. Ποιά είναι η πιθανότητα ένα οποιοδήποτε αρχικό χρωμόσωμα να αποτελεί δυναμικό βέλτιστο;
2. Σε πόσες λύσεις αντιστοιχεί ένα δυναμικό βέλτιστο που περιέχει '1' σε m θέσεις;

Απάντηση

1. $(1-p_0)^L$

2. 2^{L-m}

Κάτω φράγμα

$$E[\alpha(s, t+1)] \geq \frac{f(s, t)}{F(t)} \alpha(s, t) \left(1 - p_c \frac{l(s)}{L-1} \right) (1 - p_m)^{\tau(s)}$$

όπου

$\alpha(s,t)$: Αριθμός στιγμιοτύπων του σχήματος s
στον πληθυσμό τη στιγμή (γενιά) t

$f(x)$: Τιμή προσαρμογής της ακολουθίας x

$f(s,t)$: Μέση τιμή προσαρμογής των στιγμιοτύπων
του σχήματος s στον πληθυσμό τη στιγμή t

$F(t)$: Μέση τιμή προσαρμογής όλων των μελών
του πληθυσμού τη στιγμή t

p_c : Πιθανότητα εφαρμογής του τελεστή διασταύρωσης
σε ένα μέλος του πληθυσμού

$l(s)$: Ορίζον μήκος (defining length) του σχήματος s
(απόσταση μεταξύ των ακραίων ορισμένων ψηφίων)

p_m : Πιθανότητα μετατροπής ενός ψηφίου ενός μέλους
του πληθυσμού με εφαρμογή του τελεστή μετάλλαξης

$\tau(s)$: Αριθμός των ορισμένων ψηφίων του σχήματος s

Άσκηση 6

Θεωρούμε κωδικοποίηση χρωμοσωμάτων με ακολουθίες 6 δυαδικών ψηφίων. Για κάθε χρωμόσωμα, η τιμή της συνάρτησης προσαρμογής είναι το πλήθος των '1'. Ο τρέχων πληθυσμός περιλαμβάνει τα εξής χρωμοσώματα:

101101, 011100, 010110, 011011, 010010

Αν η πιθανότητα διασταύρωσης είναι 0,7 και η πιθανότητα μετάλλαξης ανά δυαδικό ψηφίο είναι 0,001, ζητείται κάτω φράγμα για τον μέσο αριθμό στιγμιοτύπων του σχήματος $s=[0^{****}0]$ στην επόμενη γενιά.

Απάντηση

$$E[\alpha(s,t+1)] \geq \frac{f(s,t)}{F(t)} \alpha(s,t) \left(1 - p_c \frac{l(s)}{L-1}\right) (1 - p_m)^{\tau(s)}$$

Έχουμε: $L=6$ $p_c=0,7$ $p_m=0,001$

$f(s,t)=(3+3+2)/3=2,67$ $F(t)=16/5=3,2$

$\alpha(s,t)=3$

$l(s)=5$ $\tau(s)=2$

Τελικά: $E[\alpha(s,t+1)] \geq 0,75$

ΘΕΜΑ 1

Θεωρήστε ένα σχήμα s και έστω $a(s,t)$ ο αριθμός στιγμιοτύπων του s στον πληθυσμό τη στιγμή (γενιά) t . Ποια από τις ακόλουθες απαντήσεις περιλαμβάνει μόνο προτάσεις που επιδρούν θετικά ως προς την αύξηση του κάτω φράγματος του $a(s, t+1)$;

- A. Μικρή μέση τιμή προσαρμογής των στιγμιοτύπων του σχήματος s στον πληθυσμό τη στιγμή t . Μικρή πιθανότητα εφαρμογής του τελεστή διασταύρωσης. Μεγάλος αριθμός ορισμένων ψηφίων του s .
- B. Μεγάλη μέση τιμή προσαρμογής των στιγμιοτύπων του σχήματος s στον πληθυσμό τη στιγμή t . Μεγάλο ορίζον μήκος. Μικρός αριθμός ορισμένων ψηφίων του s .
- ☒ Γ. Μεγάλη μέση τιμή προσαρμογής των στιγμιοτύπων του σχήματος s στον πληθυσμό τη στιγμή t . Μικρό ορίζον μήκος του s . Μικρή πιθανότητα εφαρμογής του τελεστή μετάλλαξης.
- ☒ Δ. Μεγάλη μέση τιμή προσαρμογής των στιγμιοτύπων του σχήματος s στον πληθυσμό τη στιγμή t . Μεγάλη πιθανότητα εφαρμογής του τελεστή διασταύρωσης. Μεγάλη πιθανότητα εφαρμογής του τελεστή μετάλλαξης.

Στρατηγικές Εξέλιξης

ΕΠΕΚΤΑΣΗ ΣΕ ΠΛΗΘΥΣΜΙΑΚΟ ΣΧΗΜΑ

μ γονείς – λ παιδιά ($\mu < \lambda$)

- Αναπαραγωγή λύσεων με εφαρμογή μετάλλαξης και επανασυνδυασμού (recombination) σε κάθε βήμα (Πολλοί τρόποι υλοποίησης του τελεστή επανασυνδυασμού)
- Επιλογή των καλύτερων λύσεων

$(\mu+\lambda)$ -ES (plus strategy): Τα μ άτομα της επόμενης γενιάς επιλέγονται από το σύνολο των $\mu+\lambda$ γονέων και παιδιών

(μ,λ) -ES (comma strategy): Τα μ άτομα της επόμενης γενιάς επιλέγονται από το σύνολο των λ παιδιών

ΘΕΜΑ 2

Ποια από τις ακόλουθες απαντήσεις σε σχέση με τους γενετικούς αλγόριθμους και τις στρατηγικές εξέλιξης περιλαμβάνει μόνο ορθές προτάσεις;

- A. Η εφαρμογή ελιτισμού ενισχύει την ανακάλυψη. Η υψηλή πιθανότητα εφαρμογής των γενετικών τελεστών ενισχύει την εκμετάλλευση. Η στρατηγική $\mu+\lambda$ ευνοεί περισσότερο την εκμετάλλευση από τη μ,λ .
- ☒ B. Η εφαρμογή ελιτισμού ενισχύει την εκμετάλλευση. Η υψηλή πιθανότητα εφαρμογής των γενετικών τελεστών ενισχύει την ανακάλυψη. Η στρατηγική μ,λ ευνοεί περισσότερο την ανακάλυψη από τη $\mu+\lambda$.
- Γ. Η εφαρμογή ελιτισμού ενισχύει την ανακάλυψη. Η υψηλή πιθανότητα εφαρμογής των γενετικών τελεστών ενισχύει την ανακάλυψη. Η στρατηγική $\mu+\lambda$ ευνοεί περισσότερο την ανακάλυψη από τη μ,λ .
- Δ. Η εφαρμογή ελιτισμού ενισχύει την εκμετάλλευση. Η υψηλή πιθανότητα εφαρμογής των γενετικών τελεστών ενισχύει την εκμετάλλευση. Η στρατηγική μ,λ ευνοεί περισσότερο την εκμετάλλευση από τη $\mu+\lambda$.

Genetic algorithm (GA) vs simulated annealing (SA)

- SA → τρέχει για άμεσες, μέσες, γρήγορες και μικρές ενώ σε GA → τρέχει για μεγάλους, μακροχρόνια, αργή, λίγη και μερικές
- SA ίδιο με GA με συνδυασμό από 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100
- SA συνδυασμός άμεσων και μέσων, ενώ GA συνδυασμός 2 μεθόδων
- SA και GA υποθέτουν ότι η λύση βρίσκεται κοντά σε γρήγορα
- GA θεωρεί ότι τα καλύτερα αποτελέσματα είναι τα πρώτα, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100

12/10 5-9010

γενετικός αλγόριθμος για clustering

→ αναζητείται κάθε μέντρο ξ σε 1 διαμόρφωση (χρωμάτωση)

λεπτομέρεια: χωρίζεις σε clusters με βάση τα χρωμάτιστα
και προσπαθείς κάθε χρωμάτιστα να φτάσει στο μέντρο
από ερμηνεύσεις των κληρονομιάς

Hopfield

Μάθηση

Αποθήκευση p N -διάστατων προτύπων ξ^k

Ο Χεμπιανός κανόνας (D. Hebb)

$$w_{ij} = \frac{1}{N} \sum_{k=1}^P (2\xi_i^k - 1)(2\xi_j^k - 1), \quad i, j = 1, \dots, N$$

(δυαδικά πρότυπα)

$$w_{ij} = \frac{1}{N} \sum_{k=1}^P \xi_i^k \xi_j^k, \quad i, j = 1, \dots, N$$

(διπολικά πρότυπα)

$$w_{ii} = 0, \quad i = 1, \dots, N$$

Λειτουργία (Δυναμική συμπεριφορά)

$$u_i(t+1) = \sum_{j=1}^N w_{ij} y_j(t) + \theta_i$$

✓ ΔΕΝ ΤΟ
ΒΑΖΟΥΜΕ
ΑΝ ΔΕΝ
ΔΕΙ/ΚΑΤΙ

$$y_i(t+1) = f(u_i(t+1)) \quad f: \text{step / signum}$$

Ενημέρωση: ασύγχρονη
σύγχρονη

Ευστάθεια – Διακριτό μοντέλο

Συνάρτηση ενέργειας (Lyapunov)

$$E(\vec{y}) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N w_{ij} y_i y_j - \sum_{i=1}^N \theta_i y_i$$

Τοπικά ελάχιστα \Rightarrow Ευσταθείς καταστάσεις

Η ευστάθεια είναι εγγυημένη για συμμετρικά δίκτυα ($w_{ij}=w_{ji}$ $i,j=1,\dots,N$) με ασύγχρονη λειτουργία.

Ευστάθεια – Αναλογικό μοντέλο

Συνάρτηση ενέργειας (Lyapunov)

$$E(\vec{y}) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N w_{ij} y_i y_j + \sum_{i=1}^N \int_0^{y_i} f^{-1}(y) dy - \sum_{i=1}^N \theta_i y_i$$

Τοπικά ελάχιστα \Rightarrow Ευσταθείς καταστάσεις

Η ευστάθεια είναι εγγυημένη για συμμετρικά δίκτυα ($w_{ij}=w_{ji}$ $i,j=1,\dots,N$).

Μόνο ο πρώτος όρος \wedge μας ενδιαφέρει στις ασκήσεις εκτός αν λέει κάτι συγκεκριμένο (θα δίνεται)

ΘΕΜΑ 5

Αναφορικά με την ευστάθεια διακριτού δικτύου Hopfield, ποια από τις ακόλουθες απαντήσεις περιλαμβάνει μόνο προτάσεις που μπορούν να ισχύουν ταυτόχρονα;

- A. Τα βάρη είναι ασύμμετρα. Η ενημέρωση των κόμβων γίνεται σύγχρονα. Το δίκτυο οδηγείται σε ταλάντωση.
- B. Τα βάρη είναι συμμετρικά. Η ενημέρωση των κόμβων γίνεται σύγχρονα. Το δίκτυο οδηγείται σε χαοτική συμπεριφορά.
- ☒ C. Τα βάρη είναι συμμετρικά. Η ενημέρωση των κόμβων γίνεται ασύγχρονα. Το δίκτυο οδηγείται σε ευσταθή κατάσταση.
- D. Τα βάρη είναι ασύμμετρα. Η ενημέρωση των κόμβων γίνεται ασύγχρονα. Το δίκτυο οδηγείται σε ευσταθή κατάσταση.

Συνθήκη ευστάθειας για ένα πρότυπο ξ^m

$$\text{sgn}\left(\sum_j w_{ij} \xi_j^m\right) = \xi_i^m, \quad i = 1, \dots, N$$

$$\sum_j w_{ij} \xi_j^m = \frac{1}{N} \sum_j \sum_k \xi_i^k \xi_j^k \xi_j^m$$

Όρος διαφωνίας (crosstalk)

$$= \xi_i^m + \frac{1}{N} \sum_j \sum_{k \neq m} \xi_i^k \xi_j^k \xi_j^m =$$

$$= \xi_i^m (1 - C_i^m)$$

$$C_i^m = -\xi_i^m \frac{1}{N} \sum_j \sum_{k \neq m} \xi_i^k \xi_j^k \xi_j^m$$

Αν $C_i^m > 1$ το πρότυπο ξ^m μπορεί να είναι ασταθές.

Χωρητικότητα (Χεμπιανός κανόνας)

$$P \leq 0,15N \quad (\text{Hopfield})$$

$$P = O(N/\log N) \quad (\text{McEliece et al.})$$

ΘΕΜΑ 8

Ποια από τις ακόλουθες απαντήσεις είναι αληθής ως προς την περιγραφόμενη ιδιότητα του δικτύου Hopfield;

- ☒ A. Το δίκτυο Hopfield, ως αυτοσυσχετιστική μνήμη, χαρακτηρίζεται από σχετικά χαμηλή χωρητικότητα.
- B. Το δίκτυο Hopfield, ως αυτοσυσχετιστική μνήμη, αποτελείται από δύο στρώματα με βάρη συμμετρικά μεταξύ κόμβων των δύο στρωμάτων.
- Γ. Το δίκτυο Hopfield μπορεί να περιλαμβάνει κρυφούς κόμβους, των οποίων τα βάρη εκπαιδεύονται με τον Χεμπιανό κανόνα.
- Δ. Το δίκτυο Hopfield μπορεί να περιλαμβάνει κρυφούς κόμβους, των οποίων τα βάρη εκπαιδεύονται με μη επιβλεπόμενη μάθηση.

Προσομοιωμένη ανόπτηση

Στοχαστική κατάβαση κλίσης

$$P_x/P_y = \exp[-(E_x - E_y)/T] = \exp[-(\Delta E)/T]$$

Ικανή συνθήκη θερμικής ισορροπίας

$$P_x P(x \rightarrow y) = P_y P(y \rightarrow x)$$

Στοχαστικό δίκτυο Hopfield Διακριτό Hopfield + Simulated Annealing

- Αν $\Delta E_i < 0$ η ενημέρωση γίνεται δεκτή
- Αν $\Delta E_i \geq 0$ η ενημέρωση γίνεται δεκτή με πιθανότητα $p(\Delta E, T) = \exp(-\Delta E_i / T)$
(Ικανοποίηση συνθήκης θερμικής ισορροπίας)

Διακριτό Hopfield για $T \rightarrow 0$

Θεώρημα (Ασυμπτωτική σύγκλιση)

Για $T \rightarrow 0$ η στατική κατανομή πιθανότητας συγκλίνει στο σύνολο των βέλτιστων λύσεων.

ΘΕΜΑ 10

Σε ένα στοχαστικό δίκτυο Hopfield στο οποίο γίνεται ενημέρωση των τιμών των κόμβων ποια από τις ακόλουθες απαντήσεις περιλαμβάνει μόνο ορθές προτάσεις;

- ☒ Α. Αν η ενέργεια του συστήματος μειώνεται η ενημέρωση γίνεται δεκτή. Για σταθερή θερμοκρασία ανόπτησης, μεγαλύτερη αύξηση της ενέργειας μετά την ενημέρωση είναι πιο εύκολο να γίνει αποδεκτή από μικρότερη. Μια ορισμένη μεταβολή ενέργειας είναι πιο δύσκολο να γίνει αποδεκτή σε υψηλότερη θερμοκρασία από ότι σε μικρότερη.
- ☐ Β. Αν η ενέργεια του συστήματος μειώνεται η ενημέρωση γίνεται δεκτή. Για σταθερή θερμοκρασία ανόπτησης, μεγαλύτερη αύξηση της ενέργειας μετά την ενημέρωση είναι πιο δύσκολο να γίνει αποδεκτή από μικρότερη. Μια ορισμένη μεταβολή ενέργειας είναι πιο δύσκολο να γίνει αποδεκτή σε υψηλότερη θερμοκρασία από ότι σε μικρότερη.
- ☐ Γ. Αν η ενέργεια του συστήματος μειώνεται η ενημέρωση απορρίπτεται. Για σταθερή θερμοκρασία ανόπτησης, μεγαλύτερη αύξηση της ενέργειας μετά την ενημέρωση είναι πιο εύκολο να γίνει αποδεκτή από μικρότερη. Μια ορισμένη μεταβολή ενέργειας είναι πιο εύκολο να γίνει αποδεκτή σε υψηλότερη θερμοκρασία από ότι σε μικρότερη.
- ☐ Δ. Αν η ενέργεια του συστήματος μειώνεται η ενημέρωση απορρίπτεται. Για σταθερή θερμοκρασία ανόπτησης, μεγαλύτερη αύξηση της ενέργειας μετά την ενημέρωση είναι πιο εύκολο να γίνει αποδεκτή από μικρότερη. Μια ορισμένη μεταβολή ενέργειας είναι πιο δύσκολο να γίνει αποδεκτή σε υψηλότερη θερμοκρασία από ότι σε μικρότερη.

Θέμα 2

perceptron \neq Hopfield

- perceptron διαχωρίζει ενώ το Hopfield να τα αναθνήκει
- perceptron υποσχεστέος βάση, Hopfield σταθμίζεις ενέργεια υποσχεστέων } από εισόδους
- perceptron πολύ λιανής
- Hopfield → ταξινόμηση αν δες το έχει αναθνήκει
↳ da σου επιστρέφει ποιο πόση να μην αφήνεις } από γενεαλογίας
- perceptron → ταξινόμηση

