

Ποια είναι η μαθηματική σημασία του ρυθμού μάθησης στον αλγόριθμο BK και πώς θα προσαρμόζατε την τιμή του σε πραγματικά προβλήματα αναγνώρισης προτύπων;

Ο ρυθμός μάθησης χρησιμοποιείται για να επιταχύνει ή να επιβραδύνει τη διαδικασία μάθησης. Όσο μικρότερο το η , τόσο μικρότερες θα είναι και οι αλλαγές στα συναπτικά βάρη του δικτύου από τη μία επανάληψη στην επόμενη και τόσο πιο ομαλή η τροχιά στο χώρο των βαρών. Αν το η γίνει υπερβολικά μεγάλο, οι αλλαγές στα συναπτικά βάρη μπορεί να είναι τέτοιες ώστε να κάνουν το δίκτυο ασταθές (να ταλαντώνει). Στην πραγματικότητα, το η δεν είναι σταθερό, αλλά εξαρτάται από τη σύνδεση, δηλαδή $\eta = \eta_{ij}$. Συγκεκριμένα, στην ιδανική περίπτωση, όλοι οι νευρώνες σ' ένα MLP πρέπει να μαθαίνουν με τον ίδιο ρυθμό. Επειδή τα τελευταία επίπεδα έχουν συνήθως μεγαλύτερες τοπικές κλίσεις, το η θα πρέπει να λαμβάνει μικρότερη τιμή στα τελευταία απ' ό,τι στα αρχικά επίπεδα του perceptron. Αντίστοιχα, οι νευρώνες με πολλές εισόδους θα έχουν μικρότερο η απ' ό,τι οι νευρώνες με λίγες εισόδους, ώστε να διατηρείται ο ίδιος χρόνος μάθησης για όλους τους νευρώνες στο δίκτυο. Σε πραγματικά προβλήματα αναγνώρισης προτύπων, το η θα πρέπει να ελαττώνεται σε κάθε επανάληψη της φάσης εκπαίδευσης, ώστε να γίνει πιο ευαίσθητο στις μεταβολές. Ωστόσο, πρέπει να αποφευχθούν οι πολύ μικρές τιμές, ώστε να μην έχουμε υπερπροσαρμογή.

Να δοθούν συνοπτικά οι διαφορές ανάμεσα σε ένα νευρωνικό δίκτυο πρόσθιας τροφοδότησης (όπως το MLP με BK) και ένα αναδρομικό δίκτυο (όπως το δίκτυο Hopfield με τον Χεμπιανό κανόνα). Πώς διαφοροποιούνται στη φάση εκπαίδευσης και πώς στη φάση λειτουργίας;

- MLP προσαρμόζει βάρη, Hopfield χαμηλώνει την ενέργεια των προτύπων (φάση εκπαίδευσης)
- MLP: ρυθμός μάθησης (φάση εκπαίδευσης)
- MLP διαχωρίζει πρότυπα, Hopfield αποθηκεύει πρότυπα (φάση λειτουργίας)
- Hopfield: ταλάντωση αν δεν έχει αποθηκευμένο το πρότυπο (φάση λειτουργίας)
- Hopfield: επιστρέφει ποιο μοιάζει παραπάνω από τα αποθηκευμένα, αν δεν έχει αποθηκευμένο (φάση λειτουργίας)
- MLP: ικανό για ταξινόμηση (φάση λειτουργίας)

Βλ. επίσης σελ.140-141 και 698-699.

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

Συνάρτηση γειτονιάς: για να προσαρμόζεται το SOM σε αυτά που του παρουσιάζεις. Αν είναι πολύ μικρή εξ' αρχής, τότε η σύγκλιση γίνεται σε τοπικό επίπεδο, προσαρμόζεται μόνο κάποιο πρότυπο σε τοπικό επίπεδο και όχι όλα, άρα μειώνεται με την πάροδο του χρόνου για να συγκλίνει πρώτα σε όλο το πλέγμα και κατόπιν σε τοπικά χαρακτηριστικά.

Βλ. επίσης σελ. 429-433

Να εξηγήσετε διαισθητικά το λόγο για τον οποίο η αύξηση της διάστασης του χώρου χαρακτηριστικών σε ένα SVM αυξάνει την πιθανότητα να μετατρέψουμε ένα μη γραμμικά διαχωρίσιμο πρόβλημα σε διαχωρίσιμο

Αύξηση χώρου διαστατικότητας → απομάκρυνση προτύπων (Βλ. και σελ.231-233, θ.Cover)

Τι συμβαίνει στο BK αν τα βάρη αρχικοποιηθούν σε πολύ μεγάλες ή πολύ μικρές τιμές;

- Πολύ μικρές: αργή σύγκλιση
- Πολύ μεγάλες: κορεσμό στις συναρτήσεις εξόδου, άρα δε θα διαδίδεται προς τα πίσω

Αυξάνεται η γενίκευση αν:

- μειωθούν οι εποχές → NAI
- αυξηθούν οι εποχές → OXI
- μειωθούν οι νευρώνες → NAI
- αυξηθούν οι νευρώνες → OXI

Έχουμε ένα νευρωνικό δίκτυο με ένα κρυμμένο επίπεδο και με εξόδους ίσες με τις εισόδους.

Πού νομίζετε ότι μπορεί να χρησιμοποιηθεί και γιατί;

Κωδικοποιητής εισόδου. Βλ. επίσης σελ. 181-183

Τι αντιπροσωπεύει κάθε νευρώνας του κρυμμένου στρώματος ενός SVM;

Κάθε νευρώνας πέρα από το να αναπαριστά το πρότυπο σε μεγαλύτερο χώρο διαστατικότητας, προσπαθεί επίσης να μάθει κι ένα κρυφό χαρακτηριστικό (feature) της εισόδου.

Όταν σχεδιάζουμε έναν αλγόριθμο μάθησης επιλέγουμε το χώρο υποθέσεων. Δώστε ένα πλεονέκτημα και ένα μειονέκτημα του να επιλέξετε ένα μεγάλο ή ένα μικρό χώρο υποθέσεων

Χώρος υποθέσεων είναι ο χώρος στον οποίο ψάχνουμε τις συναρτήσεις που ταιριάζουν με τα δεδομένα μας, ώστε να εξασφαλίσουμε προσαρμογή του δικτύου. Αν είναι μεγάλος, μπορούμε να αναπαραστήσουμε παραπάνω δεδομένα, αλλά είναι δύσκολο να “ψάξουμε” σ' αυτόν. Αντίθετα, ένας μικρός χώρος υποθέσεων επιτρέπει εύκολο “ψάξιμο”, αλλά μικρή αναπαράσταση δεδομένων.

Σε ένα MLP, η ικανότητα γενίκευσης βελτιώνεται από:

- αύξηση αριθμού κόμβων → OXI (overfitting)
- μείωση αριθμού βαρών → NAI (απλότητα δικτύου)
- αύξηση αριθμού εποχών του αλγορίθμου μάθησης → OXI (overfitting)
- μείωση του αριθμού δεδομένων μάθησης → NAI (αποφυγή overfitting)

Σε ποιες περιπτώσεις και γιατί χρησιμοποιούμε μεταβλητές χαλαρότητας στις SVM;

Περιγράψτε χρησιμοποιώντας σχηματικά παραδείγματα.

Όταν θέλουμε να ταξινομήσουμε μη διαχωρίσιμα πρότυπα, καθώς και στην περίπτωση που είναι γραμμικά διαχωρίσιμα, αλλά θέλουμε να βρούμε τη διαχωριστική επιφάνεια.

Βλ. επίσης σελ.277

Περιγράψτε ένα time-delay νευρωνικό δίκτυο και εξηγήστε πώς μαθαίνει και αναγνωρίζει ίδια patterns σε χρονικά μεταβαλλόμενα δεδομένα/σήματα εισόδου

Η είσοδος x_i κατά τη χρονική στιγμή k ισούται με την είσοδο x_{i+1} κατά τη χρονική στιγμή $k+1$. Η συνάρτηση που υλοποιείται από το δίκτυο είναι μη γραμμικό φίλτρο $y =$

$F\{s(k), s(k-1), \dots, s(k-n+1)\}$. Συνεπώς, θέλουμε BK για shifted εκδοχές του σήματος.

Ποιες συνέπειες θα μπορούσε να έχει κατά την εκπαίδευση ενός χάρτη Kohonen η χρήση πολύ μικρής γειτονιάς καθ' όλη τη διάρκεια της εκπαίδευσης; Να αιτιολογηθεί η απάντησή σας.

Πολύ μικρή γειτονιά αποθηκεύει μόνο τα τοπικά χαρακτηριστικά της εισόδου και όχι τη γενική μορφή του πλέγματος.

Να εξηγήσετε με συντομία με ποιο τρόπο οι μεταβλητές χαλάρωσης επηρεάζουν την ευρωστία του περιθωρίου ταξινόμησης σε θορυβώδη δεδομένα

Βλ. σελ. 278

Ποια είναι τα χαρακτηριστικά των βαθιών αρχιτεκτονικών νευρωνικών δικτύων και τι προσπαθούν να επιτύχουν αυτά;

- Χρησιμοποιούνται πολλά επίπεδα νευρώνων
- Κάθε μονάδα μετατρέπει την αναπαράσταση εισόδου της σε μια υψηλότερου επιπέδου

Τα τεχνητά συστήματα που προσομοιώνουν τέτοιες λειτουργίες έχουν ως στόχο να μαθαίνουν να δημιουργούν τις απαραίτητες ενδιάμεσες καταστάσεις και να παράγουν την τελική τους εκτίμηση με επιτυχία. Αυτό γίνεται με επιβλεπόμενη μάθηση, μη επιβλεπόμενη μάθηση με επιβλεπόμενο ταξινομητή στην έξοδο, μη επιβλεπόμενη μάθηση με επιβλεπόμενο εξομαλυντή.

Γενετικοί αλγόριθμοι: Γιατί το θεώρημα των σχημάτων παρέχει ένα κάτω φράγμα για τον αναμενόμενο αριθμό στιγμιοτύπων ενός σχήματος στην επόμενη γενιά;

Κάτι τέτοιο είναι δυνατό επειδή η πιθανότητα αναπαραγωγής είναι ανάλογη του fitness των χρωμοσωμάτων. Ειδικότερα, στον τύπο του θεωρήματος σχημάτων που λαμβάνει υπόψιν και την επίδραση των πιθανοτήτων διασταύρωσης και μετάλλαξης, είναι σαφές ότι λαμβάνουμε ένα ΚΑΤΩ φράγμα για τον αριθμό των χρωμοσωμάτων του σχήματος στην επόμενη γενιά, καθώς λαμβάνονται υπόψιν μόνο οι καταστροφικές ιδιότητες των παραπάνω μηχανισμών. Ως εκ τούτου, μας δίνεται “το χειρότερο σενάριο”. Έτσι, μπορεί να απορρίψει τη (μικρή έστω) μη μηδενική πιθανότητα ότι ένα string που ανήκει στο σχήμα H θα δημιουργηθεί from scratch μέσω μετάλλαξης ενός string (ή recombination) που δεν ανήκε στο H στην προηγούμενη γενιά.

Ποια είναι τα χαρακτηριστικά ενός συνελικτικού βαθιού νευρωνικού δικτύου; Αναφέρετε δύο τεχνικές κανονικοποίησης (regularization) που χρησιμοποιούνται στην υλοποίηση/μάθηση των δικτύων αυτών. Θα χρησιμοποιούσατε τον αλγόριθμο BK για την εκμάθηση αυτών των δικτύων;

- a type of feed-forward artificial neural network
- variations of multilayer perceptrons designed to use minimal amounts of preprocessing
- wide applications in image and video recognition, recommender systems[4] and natural language processing
- **3D volumes of neurons:** The layers of a CNN have neurons arranged in 3 dimensions: width, height and depth. The neurons inside a layer are only connected

to a small region of the layer before it, called a receptive field. Distinct types of layers, both locally and completely connected, are stacked to form a CNN architecture.

- **Local connectivity:** following the concept of receptive fields, CNNs exploit spatially local correlation by enforcing a local connectivity pattern between neurons of adjacent layers. The architecture thus ensures that the learnt "filters" produce the strongest response to a spatially local input pattern. Stacking many such layers leads to non-linear "filters" that become increasingly "global" (i.e. responsive to a larger region of pixel space). This allows the network to first create good representations of small parts of the input, then assemble representations of larger areas from them.
- **Shared weights:** In CNNs, each filter is replicated across the entire visual field. These replicated units share the same parameterization (weight vector and bias) and form a feature map. This means that all the neurons in a given convolutional layer detect exactly the same feature. Replicating units in this way allows for features to be detected regardless of their position in the visual field, thus constituting the property of translation invariance. Weight sharing also helps by dramatically reducing the number of free parameters being learnt, thus lowering the memory requirements for running the network. Decreasing the memory footprint allows the training of larger, more powerful networks.
- + χαρακτηριστικά βαθιών δικτύων

REGULARIZATION

- Empirical → e.g: Artificial data: Since the degree of overfitting of a model is determined by both its power and the amount of training it receives, providing a convolutional network with extra training examples can reduce overfitting. Since these networks are usually already trained with all available data, one approach is to either generate new examples from scratch (if possible) or perturb the existing training samples to create new ones. For example, input images could be asymmetrically cropped by a few percent to create new examples with the same label as the original.
- Explicit → e.g: Early stopping: One of the simplest methods to prevent overfitting of a network is to simply stop the training before overfitting has had a chance to occur. It is very easy to see why this technique works, but it comes with the disadvantage that the learning process is halted.
-

++++++ ΒΛ. ΚΕΦΑΛΑΙΟ 4.17

Θεωρούμε διδιάστατο χάρτη Kohonen σε διδιάστατο χώρο εισόδου. Αν ο χάρτης είναι διατεταγμένος, θα μπορούσε η διάταξη να ανατραπεί μετά από μερικές επαναλήψεις του αλγορίθμου εκπαίδευσης:

Όχι. Ο αλγόριθμος φροντίζει για τη διατήρηση της τοπολογίας, δηλαδή όμοια πρότυπα εισόδου απεικονίζονται σε γειτονικούς νευρώνες. Συνεπώς, εφόσον ο χάρτης είναι ήδη διατεταγμένος, δηλαδή έχει ολοκληρωθεί η φάση τοπολογικής διάταξης, επαναλήψεις του

αλγορίθμου εκπαίδευσης οδηγούν στη φάση σύγκλισης, η οποία οδηγεί σε εξάπλωση του χάρτη ώστε να καλύψει όλο το χώρο εισόδου, χωρίς όμως να αλλάξει την τοπολογία του.

Ομοιότητες και διαφορές γενετικών αλγορίθμων και προσομοιωμένης ανόπτησης

- ΠΑ: μιλάει για λύσεις, κόστη, γείτονες και κινήσεις, ΓΑ: μιλάει για χρωμοσώματα, καταλληλότητα, επιλογή, διασταύρωση και μετάλλαξη
- ΠΑ, ΓΑ: ίδιοι για πληθυσμό ίσο με 1 (στον ΓΑ), αφού δεν υπάρχει διασταύρωση αλλά μόνο μετάλλαξη
- ΠΑ: δημιουργεί νέα λύση αλλάζοντας μία ήδη υπάρχουσα, ΓΑ: συνδυάζει 2 ήδη υπάρχουσες λύσεις
- ΠΑ, ΓΑ: υποθέτουν ότι η λύση βρίσκεται κοντά σε ήδη γνωστές λύσεις
- ΓΑ: θεωρεί ότι τα παιδιά μοιράζονται ιδιότητες των γονιών, άρα 2 καλοί γονείς θα έχουν καλά παιδιά

Μπορείτε να δείξετε με βάση τις αντίστοιχες εξισώσεις ότι η επιλογή των διανυσμάτων υποστήριξης γενικά επηρεάζει τη μορφή της καμπύλης διαχωρισμού;

- Γραμμικά διαχωρίσιμα → Σχέσεις 6.17, 6.18 (σελ.274)
- Μη-γραμμικά διαχωρίσιμα → Σχέση 6.30 (σελ.281)

Ποιο είναι το αποτέλεσμα της υπερπροσαρμογής σε ένα SVM; Να θεωρήσετε μη-γραμμικά διαχωρίσιμο πρόβλημα

Μεγάλη εμπιστοσύνη στην ποιότητα του δείγματος εκπαίδευσης → λάθος ταξινόμηση (δηλαδή ταξινόμηση στη λάθος πλευρά της επιφάνειας απόφασης)

Βλ. επίσης σελ.276++