



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΔΠΜΣ ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Τελική Εξέταση χειμερινού εξαμήνου 2022-2023 στο μάθημα
«Μηχανική Μάθηση»

Ημερομηνία: 31/01/2023 - Διάρκεια: 105 λεπτά

Επώνυμο:
Όνομα:
ΔΠΜΣ/ΥΔ:
Α.Μ.:
e-mail:

Θέμα 1 (8 μον.)

Δίνεται το σύνολο δεδομένων του παρακάτω πίνακα. Να βρείτε το δέντρο απόφασης που θα κατασκευάσει ο αλγόριθμος CART.

Outlook	Temperature	Windy	Play Tennis
Sunny	Hot	False	No
Overcast	Hot	False	Yes
Overcast	Cool	True	Yes
Overcast	Hot	False	Yes

Θέμα 2 (8 μον.)

Να βρεθεί η διάσταση VC (VC dimension) ενός δυαδικού ταξινομητή δύο πραγματικών χαρακτηριστικών που αποτελείται από ένα απλό νευρώνα τύπου perceptron.

Υπόδειξη: Για τον υπολογισμό της πληθικότητας του συνόλου σημείων που κατακερματίζονται από τον ταξινομητή, να θεωρήσετε γενική τοποθέτηση των σημείων στο επίπεδο, το οποίο σημαίνει για $n=5$, για παράδειγμα, να μην θεωρήσετε ότι τα 5 σημεία είναι στην ίδια ευθεία.

Θέμα 3 (14 μον.)

Χαρακτηρίστε καθεμία από τις παρακάτω προτάσεις ως Σωστές (Σ) ή Λανθασμένες (Λ), αιτιολογώντας σε κάθε περίπτωση την απάντησή σας.

(i) Υπενθυμίζεται το πρόβλημα SVM με χρήση μεταβλητών χαλαρότητας:

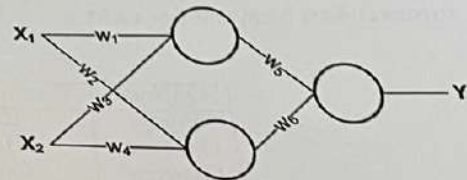
$$\begin{aligned} \min \quad & \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & \xi_i \geq 0, \forall i \in \{1, \dots, n\} \\ & (w^T x_i + b) y_i \geq 1 - \xi_i, \forall i \in \{1, \dots, n\} \end{aligned}$$

Όσο αυξάνεται η τιμή του C, τόσο αυξάνεται η ανοχή μας σε λανθασμένες ταξιν

- (ii) Αν έχουμε ένα πρόβλημα ταξινόμησης τριών (3) κλάσεων, τότε το πλήθος των δυαδικών SVM που θα πρέπει να εκπαιδεύσουμε αν ακολουθήσουμε τη μέθοδο one-against-one είναι μεγαλύτερο από το πλήθος των δυαδικών SVM που θα πρέπει να εκπαιδεύσουμε αν ακολουθήσουμε τη μέθοδο one-against-all.
- (iii) Στο SVM, αν αφαιρέσουμε ένα σημείο που ταξινομείται ορθά και βρίσκεται μακριά από το όριο απόφασης, τότε το όριο απόφασης (βέλτιστο υπερεπίπεδο διαχωρισμού) δεν θα επηρεαστεί.
- (iv) Ένα σημαντικό μειονέκτημα των SVMs είναι ότι συχνά παγιδεύονται σε τοπικά ελάχιστα, σε αντίθεση με τα MLPs.
- (v) Η εισαγωγή μεταβλητών χαλαρότητας (slack variables) καθιστά ένα γραμμικό SVM ικανό να χρησιμοποιηθεί αποτελεσματικά για την επιτυχή (δηλαδή χωρίς εσφαλμένες κατηγοριοποιήσεις) ταξινόμηση δεδομένων κλάσεων που είναι διαχωρίσιμες από μη γραμμικές επιφάνειες.
- (vi) Με τη χρήση συναρτήσεων πυρήνα (kernel functions) γίνεται έμμεσα απεικόνιση των δεδομένων σε μη γραμμικό χώρο χωρίς να εμφανίζεται πουθενά στις πράξεις η συνάρτηση μετασχηματισμού $\Phi()$ μόνη της.
- (vii) Ένα βασικό πλεονέκτημα του ADALINE έναντι του perceptron είναι ότι συγκλίνει πάντα σε περιπτώσεις μη-γραμμικά διαχωρίσιμων προβλημάτων.

Θέμα 4 (7 μον.)

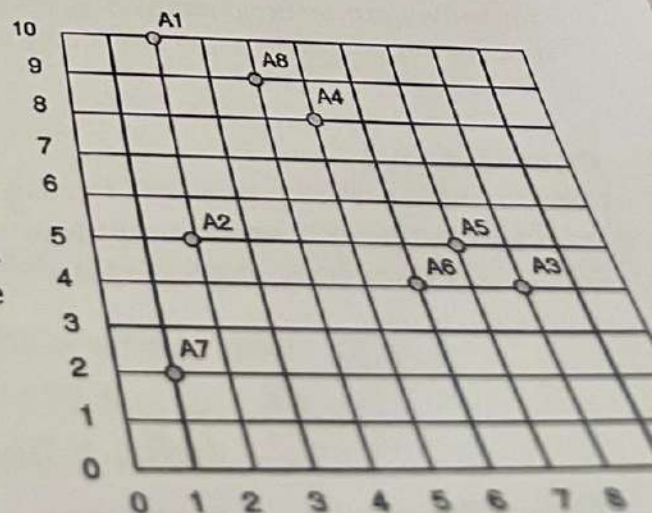
Έστω νευρωνικό δίκτυο με δύο εισόδους, ένα κρυμμένο στρώμα με δύο νευρώνες και έναν νευρώνα στο στρώμα εξόδου, όπως απεικονίζεται στην εικόνα.



- (A) Υποθέστε ότι χρησιμοποιούμε γραμμική συνάρτηση ενεργοποίησης $f(u) = cu$, όπου c σταθερά. Σχεδιάστε ένα νέο νευρωνικό δίκτυο χωρίς κρυμμένο στρώμα, το οποίο να είναι ισοδύναμο με το παραπάνω MLP. Προσδιορίστε τη συνάρτηση ενεργοποίησης και εκφράστε τα βάρη του νέου δικτύου συναρτήσει των w_i και της σταθεράς c .
- (B) Ισχύει ή όχι ότι για οποιοδήποτε MultiLayer Perceptron με γραμμικές συναρτήσεις ενεργοποίησης μπορούμε να βρούμε ένα ισοδύναμο νευρωνικό δίκτυο χωρίς κρυμμένα στρώματα; Εξηγήστε.

Θέμα 5 (8 μον.)

Δίνονται τα σημεία A1 ως A8 στο δισδιάστατο χώρο, τα οποία συσταδοποιούμε με τη χρήση του αλγορίθμου DBSCAN με $MinPts=2$ και $Eps=\sqrt{10}$. Να γράψετε ποιες συστάδες σχηματίζονται, αν υπάρχουν οριακά σημεία (border points), ποια είναι και σε ποια συστάδα είναι οριακά, καθώς και αν υπάρχουν σημεία θορύβου (noise points) και ποια είναι αυτά.



Θέμα 6 (6 μον.)

Θέλουμε να αναπτύξουμε ένα σύστημα που προβλέπει αν ένα μήνυμα ηλεκτρονικού ταχυδρομείου είναι διαφημιστικό ή όχι, ανάλογα με το αν περιέχει τις λέξεις «μοναδική» και «ευκαιρία». Δίνεται ο παρακάτω πίνακας που συνοψίζει τα δεδομένα, δηλαδή τα μέχρι τώρα στοιχεία που έχουμε από τα μηνύματα ηλεκτρονικού ταχυδρομείου που έχουμε λάβει (έχουμε βρει 300 διαφημιστικά και 1500 μη-διαφημιστικά).

Σύνολο μηνυμάτων που:	Είναι διαφημιστικά	Δεν είναι διαφημιστικά
Περιέχουν τη λέξη «ευκαιρία»	285	225
Δεν περιέχουν τη λέξη «ευκαιρία»	15	1275
Περιέχουν τη λέξη «μοναδική»	240	600
Δεν περιέχουν τη λέξη «μοναδική»	60	900

Με βάση τα παραπάνω δεδομένα, τι θα προβλέψει ο αφελής μπεϋζιανός (naïve Bayes) ταξινομητής για ένα μήνυμα που περιέχει τη λέξη «ευκαιρία» αλλά όχι τη λέξη «μοναδική»: ότι είναι διαφημιστικό ή όχι;

Θέμα 7 (5 μον.)

Έστω ότι έχουμε ένα σύστημα online μάθησης το οποίο καλείται να επιλέξει από 5 πιθανές δράσεις. Να γράψετε τις πιθανότητες $p_{2,i}$ επιλογής των δράσεων που προκύπτουν μετά την εφαρμογή του πρώτου γύρου του randomized αλγορίθμου σταθμισμένης πλειοψηφίας αν το διάνυσμα απώλειας είναι $l_2 = (1, 0, 0, 1, 0)$. Θεωρήστε επίσης ότι $\theta = 0,8$.

Θέμα 8 (14 μον.)

Στις παρακάτω ερωτήσεις πολλαπλής επιλογής, επιλέξτε τη σωστή απάντηση (ή τις σωστές απαντήσεις, εάν ζητούνται περισσότερες της μίας). Δεν χρειάζεται να αιτιολογήσετε την απάντησή σας. Κάθε σωστή απάντηση: +2 μονάδες, κάθε λανθασμένη απάντηση: -0.5 μον.

1. Ποιες δύο από τις παρακάτω προτάσεις είναι αληθείς για ένα επίπεδο CONV σε ένα CNN;
- (Α) Ο αριθμός των βαρών εξαρτάται από το βάθος των εικόνων εισόδου.
 - (Β) Ο αριθμός των bias είναι ίσος με τον αριθμό των φίλτρων.
 - (Γ) Ο συνολικός αριθμός των παραμέτρων εξαρτάται από το stride.
 - (Δ) Ο συνολικός αριθμός των παραμέτρων εξαρτάται από το padding.

2. Σχεδιάζετε ένα σύστημα βαθιάς μάθησης για την ανίχνευση της κόπωσης των οδηγών στο οδικό χώρο. Είναι σημαντικό το μοντέλο σας να ανιχνεύει την κόπωση, για την αποφυγή τυχαίων ατυχημάτων. Ποιο από τα ακόλουθα είναι το καταλληλότερο κριτήριο αξιολόγησης:

Accuracy
Precision
Recall
F1 value

- (Α) Στο Batchnorm, ο μέσος όρος εκτέλεσης αργότερα.
 (Β) Είναι ένας άλλος τρόπος εκτέλεσης πιο γρήγορη.
 (Γ) Κάνει την εκπαίδευση του δικτύου πιο γρήγορη.
 (Δ) Είναι ένας μη γραμμικός μετασχηματισμός για να κεντράρει το σύνολο δεδομένων.
 το μηδέν

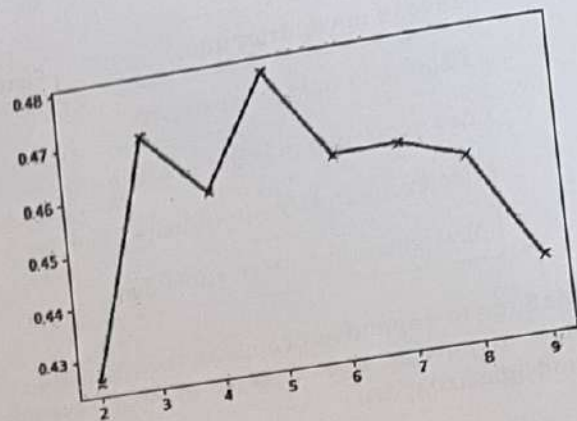
IV. Η διπλανή γραφική παράσταση απεικονίζει την τιμή του silhouette coefficient (κατακόρυφος άξονας) ως προς το πλήθος των συστάδων (οριζόντιος άξονας). Ποιος είναι ο βέλτιστος αριθμός των συστάδων στη συγκεκριμένη περίπτωση;

(Α) 2

(Β) 5

(Γ) 9

(Δ) Δεν μπορούμε να γνωρίζουμε.



V. Έστω ότι εκπαιδεύουμε μοντέλο μηχανικής μάθησης δεκαπέντε παραμέτρων, των οποίων οι τιμές μετά το πέρας της εκπαίδευσης διαμορφώνονται όπως στο παρακάτω διάνυσμα:

$[0, -2.881, 13.951, 0, 6.814, -10.516, 0, 0, -4.165, 0, 2.933, 0, 0, 0, 20.129]$

Με βάση την εμπειρία σας, ποια τεχνική ομαλοποίησης (regularization) είναι το πιθανότερο να έχει χρησιμοποιηθεί;

(Α) Καμία

(Β) L1

(Γ) L2

(Δ) Από τα δεδομένα της εκφώνησης δεν μπορούμε να βγάλουμε ασφαλές συμπέρασμα

VI. Δίνονται οι παρακάτω προτάσεις σχετικά με τον ε-άπληστο (ε-greedy) αλγόριθμο στην ενισχυτική μάθηση. Επιλέξτε την πρόταση που είναι λανθασμένη.

(Α) Στη διάρκεια της εκτέλεσής του, σε κάποιες περιπτώσεις επιλέγεται η ενέργεια με τη μέγιστη εκτιμώμενη ανταμοιβή και σε άλλες κάποια άλλη ενέργεια με τυχαίο τρόπο.

(Β) Για κάποια οριακή τιμή του ϵ εκφυλίζεται στον απλό άπληστο αλγόριθμο.

(Γ) Είναι δυνατό η αύξηση της εξερεύνησης σε κάποιες περιπτώσεις να μην βελτιώνει την επίδοση του αλγορίθμου.

(Δ) Προτείνει έναν τρόπο αντιμετώπισης του μειονεκτήματος της ελλιπούς εκμετάλλευσης που διακρίνει τον απλό άπληστο αλγόριθμο.