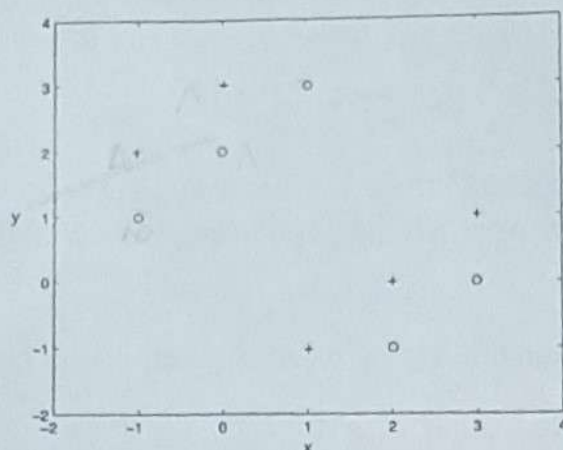


Τα θέματα τα έχω ανεβάσει μαζί με τις δικές μου λύσεις από την εξέταση. Δεν είναι απαραίτητα όλα σωστά και ακριβή

ΘΕΜΑ 1. [8 μονάδες] Έστω ότι εφαρμόζουμε τον ταξινομητή k -πλησιέστερων γειτόνων (k -NN) χρησιμοποιώντας Ευκλείδεια απόσταση για το σύνολο δεδομένων (10 δείγματα) του παρακάτω διαγράμματος.



Σχήμα 1: Δεδομένα για εφαρμογή k -NN

(α) Ποιο θα είναι το σφάλμα στο leave one out cross validation (10-fold cross validation) εάν εφαρμόσουμε 1-NN;

σε κάθε βήμα του cross validation κρατάμε ένα από τα δεδομένα για test. Αυτό θα ταξινομηθεί λάθος καθώς τα δεδομένα μας στο χώρο είναι εναλλάξ με ίση μεταξύ τους απόσταση. Άρα το test θα ταξινομηθεί στην κλάση του διπλανού του (κάποιες φορές έχει δύο διπλανά δεδομένα ίδια μεταξύ τους), η οποία είναι διαφορετική από την δική του άρα θα είναι λάθος. Άρα σε κάθε ένα από τα 10 folds έχουμε 1 λάθος.

(β) Ποια από τις παρακάτω τιμές του k οδηγεί σε μικρότερη τιμή σφάλματος: 3, 5 ή 9; Ποια είναι η τιμή του σφάλματος για αυτό το k ;

$k=3$ σε κάθε fold 1 λάθος

$k=5$ σε κάθε fold 1 λάθος

$k=9$ σε κάθε fold 1 λάθος

Για κάθε k οδηγούμαστε σε ίδιο σφάλμα

10 λόγω της κατανομής των δεδομένων στο χώρο

ΘΕΜΑ II. [12 μονάδες] Σας ζητείται να σχεδιάσετε νευρωνικό δίκτυο MLP (το κατά το δυνατόν πιο απλό) για ταξινόμηση διδιάστατων δεδομένων $X = (x_1, x_2)$ σε δύο κλάσεις. Δίνονται τα δεδομένα εκπαίδευσης που φαίνονται στο Σχήμα. Περιγράψτε τη δομή του δικτύου, αναφερθείτε στις εισόδους, εξόδους και στο πλήθος νευρώνων κάθε στρώματος, εξηγώντας (ποιοτικά) το ρόλο (λειτουργία) κάθε νευρώνα.



Σχήμα 2: Δεδομένα εκπαίδευσης για το νευρωνικό δίκτυο

Βλέπουμε ότι η κλάση των σταυρών περιβάλλεται από την άλλη, οπότε να την έχει περικεκλωθεί. Συνεπώς πρέπει το MLP να δημιουργήσει περίπου ένα κύκλο γύρω από τους σταυρούς.

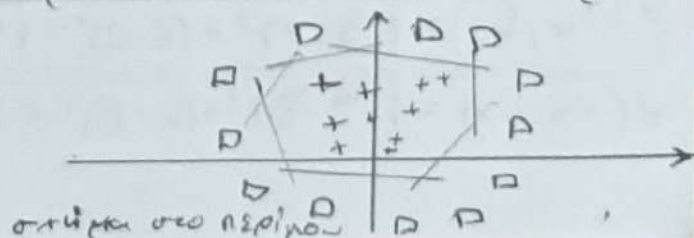
Κάθε νευρώνας του MLP θα τραβάει 1 γραμμική διαχωριστική επιφάνεια στο χώρο.

Είσοδος θα είναι οι συντεταγμένες x_1 και x_2 , δηλαδή θα έχουμε 2 εισόδους.

Η έξοδος θα είναι 1, το γινόμενο ή 0 ή 1 ανάλογα την κλάση στην οποία θα ταξινομήσουμε το στιγμιότυπο.

Θα έχουμε μία ~~κρυφή~~ κρυφή στρώση που θα περιέχει 19 νευρώνες, όσα και τα δεδομένα της κλάσης των ρόμβων. Κάθε ένας από αυτούς τους νευρώνες τραβάει μια διαχωριστική επιφάνεια ανάμεσα στον αντίστοιχο ρόμβο και τους σταυρούς. Αν ενώσουμε τις τομές αυτών των διαχωριστικών ευθειών θα έχει δημιουργηθεί κάτι σαν κύκλο γύρω από τους σταυρούς.

Ίσως είναι επαρκείς και λιγότεροι νευρώνες αφού στην τελική τομή των ευθειών, κάποιες είναι περίσies.



ΘΕΜΑ III. [10 μονάδες] (α) Δίνεται ο παρακάτω πίνακας εγγύτητας 5 σημείων $x_i \in \mathbb{R}^2, i = 1, 2, \dots, 5$:

$$P(X) = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0 & 3.42 & 2.24 & 3.61 & 5.83 \\ 3.42 & 0 & 2.83 & 3.16 & 3.61 \\ 2.24 & 2.83 & 0 & 1.41 & 4.12 \\ 3.61 & 3.16 & 1.41 & 0 & 2.15 \\ 5.83 & 3.61 & 4.12 & 2.15 & 0 \end{bmatrix} \end{matrix}$$

Αν με βάση τον $P(X)$ εφαρμόσουμε τον ιεραρχικό συσσωρευτικό αλγόριθμο πλήρους δεσμού (complete link - MAX), να δείξετε ποια θα είναι η συσταδοποίηση που προκύπτει μετά το δεύτερο βήμα του αλγόριθμου.

Βήμα 1
 $\min d = d(x_2, x_4) \rightarrow \text{cluster}(x_2, x_4)$

	x_1	x_2	x_3+x_4	x_5
x_1	0	3.42	3.61	5.83
x_2	3.42	0	3.16	3.61
x_3+x_4	3.61	3.16	0	4.12
x_5	5.83	3.61	4.12	0

Βήμα 2
 $\min d = d(x_2, x_3+x_4) = 3.16 \rightarrow \text{cluster}(x_2, x_3, x_4)$

άρα παίρνω τα clusters $\{x_1, (x_2, x_3, x_4), x_5\}$

(β) Εφαρμόζουμε τον αλγόριθμο k-means για τον διαχωρισμό των σημείων $x_1 = (2, 3), x_2 = (3, 3), x_3 = (6, 7), x_4 = (8, 8), x_5 = (7, 5), x_6 = (9, 6)$ σε δύο ομάδες. Αν τα κέντρα των ομάδων αρχικοποιούνται ως $\theta_1^{(0)} = (2, 3)$ και $\theta_2^{(0)} = (8, 8)$, να υπολογίσετε και να δείξετε ποιες θα είναι θέσεις των δύο κέντρων μετά την ολοκλήρωση της πρώτης επανάληψης του αλγόριθμου.

$$\left. \begin{aligned} d(x_1, \theta_1) &= \sqrt{(2-2)^2 + (3-3)^2} = 0 \\ d(x_2, \theta_1) &= \sqrt{(3-2)^2 + (3-3)^2} = 1 \end{aligned} \right\} x_1 \text{ στο } \underline{\theta_1}$$

$$\left. \begin{aligned} d(x_2, \theta_1) &= \sqrt{(3-2)^2 + (3-3)^2} = 1 \\ d(x_2, \theta_2) &= \sqrt{(3-8)^2 + (3-8)^2} = \sqrt{50} \end{aligned} \right\} x_2 \text{ στο } \underline{\theta_1}$$

$$\left. \begin{aligned} d(x_3, \theta_1) &= \sqrt{(6-2)^2 + (7-3)^2} = \sqrt{16+16} \\ d(x_3, \theta_2) &= \sqrt{(6-8)^2 + (7-8)^2} = \sqrt{4+1} \end{aligned} \right\} x_3 \text{ στο } \underline{\theta_2}$$

$$\left. \begin{aligned} d(x_4, \theta_1) &= \sqrt{(8-2)^2 + (8-3)^2} \\ d(x_4, \theta_2) &= 0 \end{aligned} \right\} x_4 \text{ στο } \underline{\theta_2}$$

$$\left. \begin{aligned} d(x_5, \theta_1) &= \sqrt{(7-2)^2 + (5-3)^2} = \sqrt{25+4} \\ d(x_5, \theta_2) &= \sqrt{(7-8)^2 + (5-8)^2} = \sqrt{1+9} \end{aligned} \right\} x_5 \text{ στο } \underline{\theta_1}$$

$$\left. \begin{aligned} d(x_6, \theta_1) &= \sqrt{(9-2)^2 + (6-3)^2} = \sqrt{49+9} \\ d(x_6, \theta_2) &= \sqrt{(9-8)^2 + (6-8)^2} = \sqrt{1+4} \end{aligned} \right\} x_6 \text{ στο } \underline{\theta_2}$$

$$\begin{aligned} \theta_1^{(1)} &= \left(\frac{2+3}{2}, \frac{3+3}{2} \right) \\ &= (2.5, 3) \\ \theta_2^{(1)} &= \left(\frac{6+8+7+9}{4}, \frac{7+8+5+6}{4} \right) \\ &= (7.5, 6.5) \end{aligned}$$

ΘΕΜΑ IV. [8 μονάδες] Έστω ότι έχουμε τα παρακάτω δεδομένα 2 διαστάσεων:

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

Να εφαρμόσετε συνέλιξη με το φίλτρο

4	1
3	5

και να υπολογίσετε τον παραγόμενο χάρτη χαρακτηριστικών για padding ίσο με το μηδέν και για stride ίσο με το 1 και στις δύο διαστάσεις.

$$\begin{array}{cccc}
 4+0+0+5 & 0+1+3+5 & 4+0+3+5 & 0+0+3+5 \\
 0+3+3+0 & 4+1+0+5 & 4+1+3+5 & 4+1+3+0 \\
 4+0+0+0 & 0+1+0+5 & 4+3+3+0 & 4+0+0+0 \\
 0+0+3+5 & 0+1+3+0 & 4+0+0+5 & 0+0+3+0
 \end{array}$$

$$\begin{array}{cccc}
 9 & 9 & 12 & 8 \\
 4 & 10 & 13 & 8 \\
 4 & 6 & 8 & 4 \\
 8 & 4 & 9 & 3
 \end{array}$$

↙ χάρτης
χαρακτηριστικών

→

ΘΕΜΑ V. [30 μονάδες] Στις ακόλουθες ερωτήσεις επιλέξτε ευκρινώς ποια (μοναδική) απάντηση (Α, Β, Γ ή Δ) θεωρείτε ότι είναι η ορθή. Ορθή απάντηση: 3 μονάδες. Λάθος απάντηση: -0.75 μονάδα. Κενή απάντηση: 0 μονάδες.

1. Δίνεται το σύνολο δεδομένων του Πίνακα 1. Το πλήθος των διαφορετικών δέντρων απόφασης που ταξινομούν σωστά τα δεδομένα είναι:

- (A) 2. ~
(B) 4.
(Γ) 8.
(Δ) Καμία από τις άλλες απαντήσεις δεν είναι σωστή.

Πίνακας 1: Σύνολο δεδομένων D

X1	X2	Y
TRUE	TRUE	No
TRUE	FALSE	Yes
FALSE	TRUE	Yes
FALSE	FALSE	No

2. Δίνονται οι παρακάτω προτάσεις:

- (A) Κάθε δέντρο απόφασης μπορεί να μετασχηματιστεί άμεσα σε μία λογική πρόταση σε κανονική διαζευκτική μορφή. Σ
(B) Το πρόβλημα εύρεσης του ελάχιστου δέντρου απόφασης δεν μπορεί πάντα να επιλυθεί. Λ
(Γ) Ο αλγόριθμος CART χρησιμοποιεί το μέτρο της εντροπίας για την επιλογή του χαρακτηριστικού απόφασης. Λ

Πόσες από τις παραπάνω προτάσεις ισχύουν;

- (A) Καμία

- (B) 1

- (Γ) 2 ~

- (Δ) 3

Αν έχουμε άπειρο χρόνο να κάνουμε όλες τις δοκιμές, μπορούμε πάντα να βρούμε το ελάχιστο δέντρο.

3. Έστω ένα σύνολο δεδομένων και δύο διαφορετικά δέντρα απόφασης που ταξινομούν σωστά όλα τα δεδομένα, χωρίς σφάλμα. Ισχύει ότι:

- (A) Τα δύο δέντρα έχουν την ίδια ικανότητα γενίκευσης, ανεξάρτητα από το ύψος τους.
(B) Το υψηλότερο δέντρο έχει καλύτερη γενίκευση.
(Γ) Το υψηλότερο δέντρο έχει χειρότερη γενίκευση. ~
(Δ) Καμία από τις άλλες προτάσεις δεν είναι σωστή.

4. Ένα τυχαίο δάσος κατασκευάζεται με τμηματοποίηση χαρακτηριστικών (feature bagging). Ισχύει ότι:

- (A) Τα δέντρα που το απαρτίζουν έχουν γενικά κοινούς κόμβους χαρακτηριστικών (παρατηρούμε ίδιους κόμβους σε διαφορετικά δέντρα).
(B) Τα δέντρα που το απαρτίζουν έχουν διαφορετικούς κόμβους χαρακτηριστικών (δεν παρατηρούμε ίδιους κόμβους σε διαφορετικά δέντρα).
(Γ) Τα δέντρα που το απαρτίζουν έχουν τους ίδιους κόμβους χαρακτηριστικών (παρατηρούμε ακριβώς τους ίδιους κόμβους σε όλα τα δέντρα, πιθανά σε διαφορετική διάταξη).
(Δ) Καμία από τις άλλες προτάσεις δεν είναι σωστή.

5. Θεωρούμε το πρόβλημα δυαδικής ταξινόμησης κατά Bayes σε δύο κλάσεις ω_1 και ω_2 . Υποθέτουμε ότι οι εκ των προτέρων πιθανότητες των δύο κλάσεων είναι $P(\omega_1)$ και $P(\omega_2)$ και οι πιθανοφάνειες παρατήρησης ενός χαρακτηριστικού $x \in \mathbb{R}$ στις δύο κλάσεις είναι $P(x|\omega_1)$ και $P(x|\omega_2)$, αντίστοιχα. Σύμφωνα με το θεώρημα Bayes, ποιες από τις παρακάτω εκφράσεις δίνουν την εκ των υστέρων πιθανότητα $P(\omega_1|x)$;

I. $P(\omega_1|x) = \frac{P(x|\omega_1)P(\omega_1)}{P(x)}$ Σ

II. $P(\omega_1|x) = \frac{P(x|\omega_1)P(\omega_1)}{P(x|\omega_1)+P(x|\omega_2)}$ \wedge

III. $P(\omega_1|x) = \frac{P(x|\omega_1)P(\omega_1)}{P(\omega_1)+P(\omega_2)}$ \wedge

IV. $P(\omega_1|x) = \frac{P(x|\omega_1)P(\omega_1)}{P(x|\omega_1)P(\omega_1)+P(x|\omega_2)P(\omega_2)}$ Σ

(A') I και II

(B') I και III

☒ (Γ') I και IV

(Δ') III και IV

6. Ποιοι από τους παρακάτω ισχυρισμούς είναι λανθασμένοι σχετικά με τον αλγόριθμο ομαδοποίησης k -means;

(I): Ο k -means βρίσκει εγγυημένα την ολικά βέλτιστη λύση για τα κέντρα των κλάσεων. \wedge

(II): Ο k -means αναθέτει κάθε σημείο στην εγγύτερή του κλάση με βάση την Ευκλείδεια απόσταση. $\times \Sigma$

(III): Ο k -means μπορεί να χειριστεί αποτελεσματικά μη-σφαιρικές κλάσεις, ανεξάρτητα από το σχήμα τους. \wedge

(IV): Η υπολογιστική πολυπλοκότητα του k -means είναι $O(nklt)$, όπου n είναι ο αριθμός των σημείων, k ο αριθμός των κλάσεων, l ο αριθμός των χαρακτηριστικών και t ο αριθμός των επαναλήψεων του αλγόριθμου. Σ

(A') I και II

☒ (B') I και III

(Γ') III και IV

(Δ') I, III και IV

7. Έστω Multi Layer Perceptron $MLP1$ με γραμμικές συναρτήσεις ενεργοποίησης. Τι από τα παρακάτω ισχύει;

(A') Μπορούμε να βρούμε ένα ισοδύναμο νευρωνικό δίκτυο χωρίς κρυμμένα στρώματα μόνο εάν πληρούνται συγκεκριμένες προϋποθέσεις για τη δομή του $MLP1$.

(B') Μπορούμε να βρούμε ένα ισοδύναμο νευρωνικό δίκτυο χωρίς κρυμμένα στρώματα μόνο εάν πληρούνται συγκεκριμένες προϋποθέσεις για τις εισόδους του $MLP1$.

☒ (Γ') Μπορούμε να βρούμε ένα ισοδύναμο νευρωνικό δίκτυο χωρίς κρυμμένα στρώματα ανεξαρτήτως της δομής και των εισόδων του $MLP1$.

(Δ') Δεν μπορούμε να βρούμε ισοδύναμο νευρωνικό δίκτυο χωρίς κρυμμένα στρώματα.

8. Ποιοι από τους παρακάτω ισχυρισμούς για τα SVMs είναι ορθοί;

(I): Αν αφαιρέσουμε ένα σημείο που ταξινομείται ορθά και βρίσκεται μακριά από το όριο απόφασης, τότε το όριο απόφασης (βέλτιστο υπερεπίπεδο διαχωρισμού) δεν θα επηρεαστεί. Σ

(II): 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. Σ

(III): 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 μικρότερο από το πλήθος των δυαδικών SVM που θα πρέπει να εκπαιδεύσουμε αν ακολουθήσουμε τη μέθοδο one-against-all. \wedge

☒ (A') I και II

(B') I και III

(Γ') II και III

(Δ') I, II και III

9. Έστω η boolean συνάρτηση $y = x_1 \cup (\neg x_2)$, με $x_1, x_2 \in \{0, 1\}$. Ποια από τις παρακάτω προτάσεις είναι σωστή;

- (A') Το πρόβλημα είναι γραμμικά διαχωρίσιμο.
- (B') Το πρόβλημα, αν και όχι αυστηρά γραμμικά διαχωρίσιμο, είναι σχεδόν γραμμικά διαχωρίσιμο, γι' αυτό και ένα δίκτυο ADALINE μπορεί να συγκλίνει επιτρέποντας μικρό αριθμό λανθασμένων ταξινομήσεων.
- (Γ') Η συνάρτηση δεν μπορεί να αναπαρασταθεί από απλό perceptron, αλλά μπορεί να αναπαρασταθεί από πολυστρωματικό perceptron (MLP) με ένα κρυμμένο στρώμα δύο νευρώνων και στρώμα εξόδου ενός νευρώνα.
- (Δ') Καμία από τις παραπάνω.

10. Τι από τα παρακάτω θα συμβεί εάν αυξήσουμε την τιμή της υπερπαραμέτρου C σε ένα Support Vector Machine (SVM);

- (A') Το σφάλμα εκπαίδευσης τείνει να μειωθεί.
- (B') Το περιθώριο διαχωρισμού margin θα παραμείνει σίγουρα αμετάβλητο.
- (Γ') Και οι δύο παραπάνω προτάσεις είναι σωστές.
- (Δ') Καμία από τις παραπάνω προτάσεις δεν είναι σωστή.
-