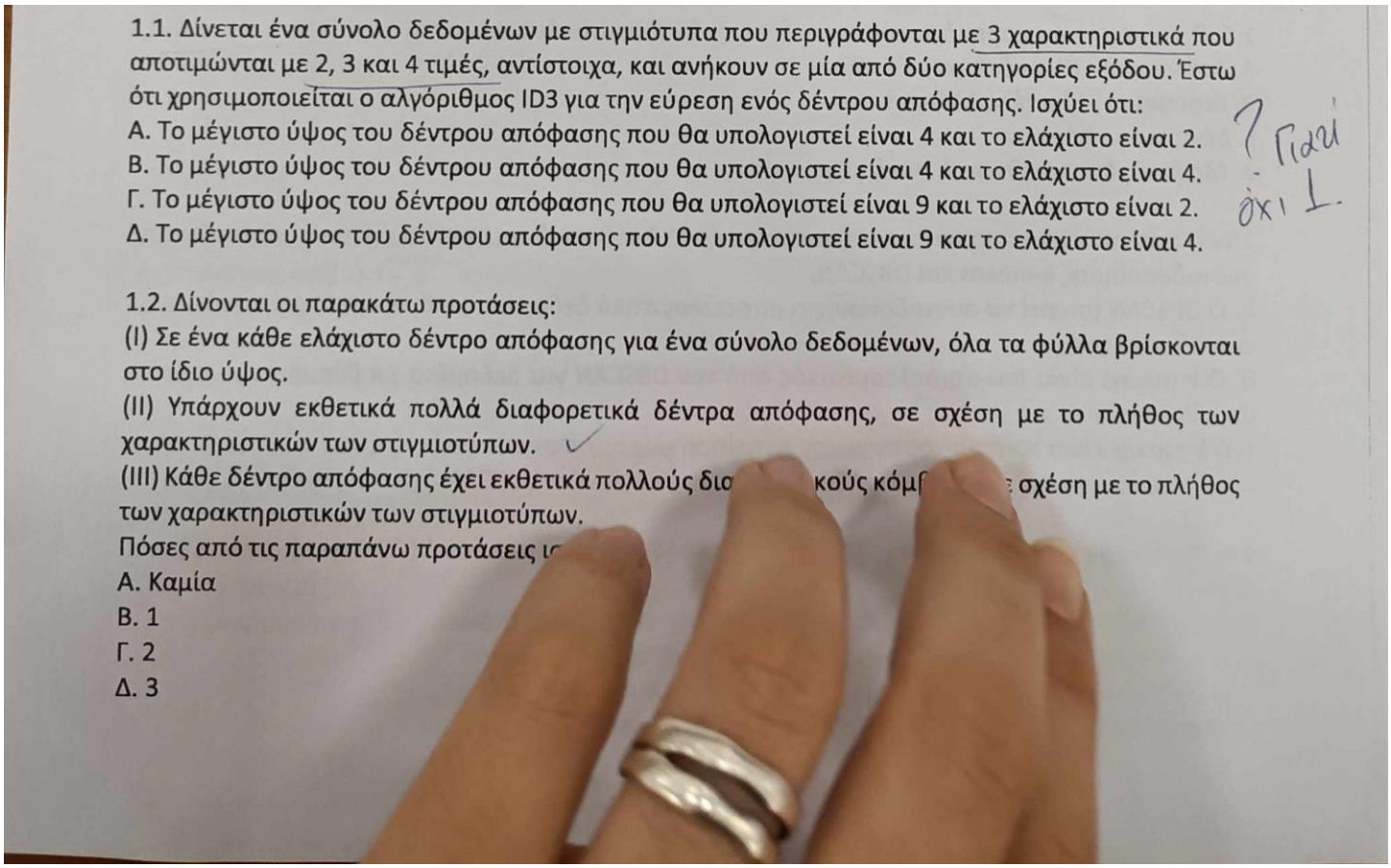
Θέματα ML - ΕΔΕΜΜ 2023-2024 Ιούνιος



1.1. Δίνεται ένα σύνολο δεδομένων με στιγμιότυπα που περιγράφονται με 3 χαρακτηριστικά που αποτιμώνται με 2, 3 και 4 τιμές, αντίστοιχα, και ανήκουν σε μία από δύο κατηγορίες εξόδου. Έστω ότι χρησιμοποιείται ο αλγόριθμος ID3 για την εύρεση ενός δέντρου απόφασης. Ισχύει ότι:

Α. Το μέγιστο ύψος του δέντρου απόφασης που θα υπολογιστεί είναι 4 και το ελάχιστο είναι 2.

Β. Το μέγιστο ύψος του δέντρου απόφασης που θα υπολογιστεί είναι 4 και το ελάχιστο είναι 4.

Γ. Το μέγιστο ύψος του δέντρου απόφασης που θα υπολογιστεί είναι 9 και το ελάχιστο είναι 2.

Δ. Το μέγιστο ύψος του δέντρου απόφασης που θα υπολογιστεί είναι 9 και το ελάχιστο είναι 4.

1.2. Δίνονται οι παρακάτω προτάσεις:

(I) Σε ένα κάθε ελάχιστο δέντρο απόφασης για ένα σύνολο δεδομένων, όλα τα φύλλα βρίσκονται στο ίδιο ύψος.

(II) Υπάρχουν εκθετικά πολλά διαφορετικά δέντρα απόφασης, σε σχέση με το πλήθος των χαρακτηριστικών των στιγμιοτύπων.

(III) Κάθε δέντρο απόφασης έχει εκθετικά πολλούς δια κούς κόμ σχέση με το πλήθος των χαρακτηριστικών των στιγμιοτύπων.

Πόσες από τις παραπάνω προτάσεις ι

Α. Καμία

Β. 1

Γ. 2

Δ. 3

1.3. Δίνεται ένα σύνολο δεδομένων για το οποίο υπολογίζουμε ένα τυχαίο δάσος (random forest) με τη μέθοδο της τμηματοποίησης του συνόλου δεδομένων.
Α. Τα δέντρα απόφασης του τυχαίου δάσους είναι γενικά μικρότερου ύψους από αντίστοιχα δέντρα απόφασης που θα υπολογίζαμε αν δεν τμηματοποιούσαμε το σύνολο δεδομένων.
Β. Τα δέντρα απόφασης του τυχαίου δάσους είναι γενικά μεγαλύτερου ύψους από αντίστοιχα δέντρα απόφασης που θα υπολογίζαμε αν δεν τμηματοποιούσαμε το σύνολο δεδομένων.
Γ. Τα δέντρα απόφασης του τυχαίου δάσους είναι γενικά ίδιου ύψους με αντίστοιχα δέντρα απόφασης που θα υπολογίζαμε αν δεν τμηματοποιούσαμε το σύνολο δεδομένων.
Δ. Καμία από τις άλλες προτάσεις δεν είναι σωστή.

1.4. Δίνεται το σύνολο δεδομένων *D* του Πίνακα 1. Κατά τη διαδικασία εκτέλεσης του αλγορίθμου CART, ποιο χαρακτηριστικό επιλέγεται στη ρίζα του δέντρου απόφασης;

Πίνακας 1: Σύνολο δεδομένων $\mathbb{D}$

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

Α. Το X1.
Β. Το X2.
Γ. Επειδή περισσότερα από ένα χαρακτηριστικά έχουν το ίδιο κέρδος πληροφορίας, ο αλγόριθμος δεν θα τερματίσει.
Δ. Επειδή περισσότερα από ένα χαρακτηριστικά έχουν το ίδιο κέρδος πληροφορίας, θα επιλεγεί τυχαία ένα από αυτά.

1.5. Ποιος από τους παρακάτω ισχυρισμούς για τη μέθοδο ελαχίστων τετραγώνων είναι **εσφαλμένος**;
Α. Το διάνυσμα των βαρών προσδιορίζεται επακριβώς (λύση κλειστής μορφής)
Β. Έχει μεγάλη ευαισθησία σε έκτοπες τιμές
Γ. Βασίζεται στην υπόθεση ότι τα χαρακτηριστικά των δεδομένων υπακούουν στην κανονική κατανομή
Δ. Χρησιμοποιείται για τον προσδιορισμό των χαρακτηριστικών γραμμικών συναρτήσεων απόφασης για προβλήματα δυαδικής ταξινόμησης.

1.6. Ποιοι από τους παρακάτω τύπους ταξινομητών **δεν** είναι διαμεριστικοί (divisive);
Α. Ταξινομητές Πλησιέστερων Γειτόνων
Β. Γκαουσιανά Μοντέλα Μίξης
Γ. Δέντρα Αποφάσεων
Δ. Μηχανές Διανυσμάτων Υποστήριξης

1.7. Ποιος από τους παρακάτω ισχυρισμούς είναι **λανθασμένος** σχετικά με τους αλγόριθμους συσταδοποίησης k-means και DBSCAN;
Α. Ο DBSCAN μπορεί να συσταδοποιήσει αποτελεσματικά δεδομένα με διαφορετικές πυκνότητες, ενώ ο k-means όχι.
Β. Ο k-means είναι πιο αποτελεσματικός από τον DBSCAN για δεδομένα με θόρυβο και ακραία σημεία (outliers).
Γ. Ο k-means είναι ευαίσθητος στην αρχικοποίηση των κέντρων των συστάδων.
Δ. Ο k-means υποθέτει ότι οι συστάδες είναι σφαιρικές, ενώ ο DBSCAN όχι.

1.8. Ποια από τις παρακάτω προτάσεις είναι αληθής για ένα συνελικτικό (CONV) επίπεδο σε ένα CNN;
Α. Ο αριθμός των bias είναι μεγαλύτερος από τον αριθμό των φίλτρων.
Β. Ο συνολικός αριθμός των παραμέτρων εξαρτάται από το stride.
Γ. Ο αριθμός των βαρών εξαρτάται από το βάθος των εικόνων εισόδου.
Δ. Ο συνολικός αριθμός των παραμέτρων εξαρτάται από το padding.

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

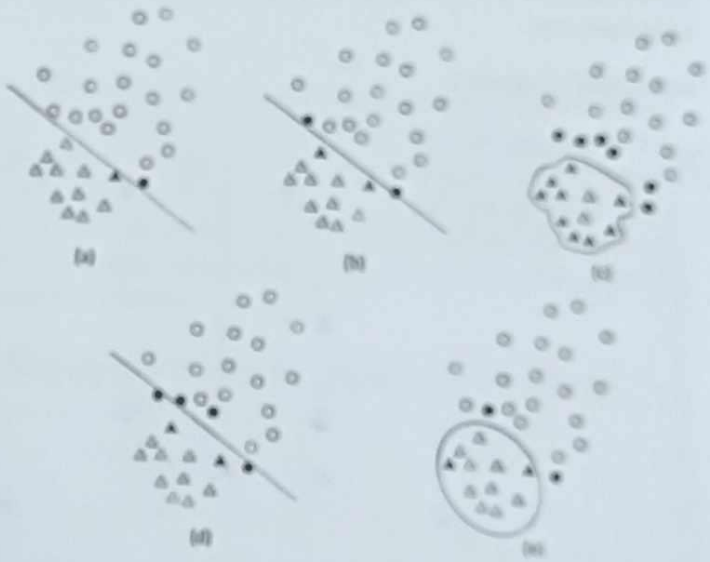
Α. Η συνάρτηση δεν μπορεί να αναπαρασταθεί από απλό perceptron, αλλά μπορεί να αναπαρασταθεί από πολυστρωματικό perceptron (MLP) με ένα κρυμμένο στρώμα δύο νευρώνων και στρώμα εξόδου ενός νευρώνα.

Β. Η συνάρτηση δεν μπορεί να αναπαρασταθεί από απλό perceptron, αλλά μπορεί να αναπαρασταθεί από πολυστρωματικό perceptron (MLP) με δύο κρυμμένα στρώματα δύο νευρώνων και στρώμα εξόδου ενός νευρώνα.

Γ. Το συγκεκριμένο πρόβλημα δεν είναι γραμμικά διαχωρίσιμο, όμως ένα νευρωνικό δίκτυο ADALINE μπορεί να συγκλίνει επιτρέποντας ακριβώς μία λανθασμένη ταξινόμηση.

Δ. Η συνάρτηση μπορεί να αναπαρασταθεί από απλό perceptron.

Στο Σχήμα 1 φαίνονται τα όρια απόφασης πέντε SVMs που προκύπτουν από διαφορετικούς πυρήνες (kernels) ή/και διαφορετικές τιμές ποινής χαλαρότητας C. Οι δύο κλάσεις δεδομένων αναπαρίστανται με τρίγωνα και κύκλους αντίστοιχα, ενώ τα συμπαγή (σκουρόχρωμα) τρίγωνα και κύκλοι αναπαριστούν διανύσματα υποστήριξης.



Σχήμα 1

1.10. Ποιο από τα παρακάτω διαγράμματα του Σχήματος 1 αντιστοιχεί σε γραμμικό SVM με C=0.1;

Α. Το διάγραμμα (a)

Β. Το διάγραμμα (b).

Γ. Το διάγραμμα (c).

Δ. Το διάγραμμα (d).

1.11 Ποιο από τα παρακάτω διαγράμματα του Σχήματος 1 αντιστοιχεί σε γραμμικό SVM με C=5;

Α. Το διάγραμμα (b).

Β. Το διάγραμμα (c).

Γ. Το διάγραμμα (d).

Δ. Το διάγραμμα (e).

1.12 Ποιο από τα παρακάτω διαγράμματα του Σχήματος 1 αντιστοιχεί σε SVM με πυρήνα $K(u, v) = (u^T v)^2$;

Α. Το διάγραμμα (b).

Β. Το διάγραμμα (c).

Γ. Το διάγραμμα (d).

Δ. Το διάγραμμα (e).

Θέμα 2 [8 μονάδες]

Δίνονται τα παρακάτω δεδομένα δύο χαρακτηριστικών για δύο κλάσεις:

Κλάση Α: [2, -4], [4, -2], [0, -1], [4, -4], [-2, -1]
Κλάση Β: [-2, 0], [-2, 2], [2, 5], [-4, -2], [2, 0]

Έστω ότι μας δίνεται ένα καινούργιο στιγμιότυπο [0, 0], του οποίου η κλάση είναι άγνωστη.

2.1. Σχεδιάστε στο επίπεδο τα δεδομένα των δύο κλάσεων καθώς και το νέο στιγμιότυπο.
2.2. Σε ποια κλάση θα τοποθετήσει το νέο στιγμιότυπο ο ταξινομητής πλησιέστερου γείτονα αν ως συνάρτηση απόστασης χρησιμοποιεί την ευκλείδεια απόσταση; Αιτιολογήστε σύντομα την απάντησή σας.
2.3. Σε ποια κλάση θα τοποθετήσει το νέο στιγμιότυπο ο ταξινομητής 3-πλησιέστερων γειτόνων αν ως συνάρτηση απόστασης χρησιμοποιεί την ευκλείδεια απόσταση; Αιτιολογήστε σύντομα την απάντησή σας.

Απάντηση Θέματος 2

Θέμα 3 [10 μονάδες]

Ένας πάροχος κινητής τηλεφωνίας επιθυμεί να φτιάξει ένα σύστημα βασισμένο στον αφελή μπεϋζιανό (Naïve Bayes) ταξινομητή, το οποίο να μπορεί να προβλέπει αν ένας πελάτης πρόκειται να «φύγει» προς ανταγωνιστική εταιρία (churn prediction), δεδομένου του ύψους του μηνιαίου του παγίου και της διάρκειας του τελευταίου του συμβολαίου. Για το σκοπό αυτό έχει συγκεντρώσει στοιχεία για 1000 πελάτες (από τους οποίους οι 800 τελικά έμειναν στον πάροχο και οι 200 τελικά «έφυγαν») αναφορικά με τη διάρκεια του συμβολαίου και το ύψος του μηνιαίου παγίου, τα οποία συνοψίζονται στον παρακάτω πίνακα:

Σύνολο πελατών που:	Έμειναν στον πάροχο	Έφυγαν από τον πάροχο
Είχαν 18μήνο συμβόλαιο πριν	200	100
Είχαν 24μήνο συμβόλαιο πριν	600	100
Είχαν πάγιο άνω των €30 το μήνα	100	150
Είχαν πάγιο κάτω από €30 το μήνα	700	50

Με βάση τα παραπάνω δεδομένα, τι θα προβλέψει ο αφελής μπεϋζιανός ταξινομητής για ένα πελάτη που έχει πάγιο άνω των €30 το μήνα και 18μήνο συμβόλαιο, ότι θα μείνει ή ότι θα «φύγει»; Αιτιολογήστε την απάντησή σας.

Θέμα 4 [10 μονάδες]

Έστω νευρωνικό δίκτυο στο οποίο οι έξοδοι του επιπέδου *U* τροφοδοτούνται στο επίπεδο *V* και έστω *W* τα βάρη των συνδέσεων μεταξύ των επιπέδων *U* και *V*. Στο επίπεδο *U* η συνάρτηση ενεργοποίησης είναι η υπερβολική εφαπτομένη (tanh) ενώ στο επίπεδο *V* η σιγμοειδής. Αν u_i είναι η έξοδος του $i^{στού}$ νευρώνα του επιπέδου *U*, v_k η έξοδος του $k^{στού}$ νευρώνα του επιπέδου *V*, $u_i = 0.8$, $v_k = 0.4$ και το βάρος μεταξύ των δύο νευρώνων $w_{i,k} = 0.3$:

4.1. Υπολογίστε την τιμή της μερικής παραγώγου $\frac{\partial v_k}{\partial u_i}$.

4.2. Υπολογίστε την τιμή της μερικής παραγώγου $\frac{\partial v_k}{\partial w_{i,k}}$.

Θέμα 5 [6 μονάδες]

Θεωρούμε μια παραλλαγή του αλγόριθμου k-means, όπου στον αλγόριθμο, αντί της Ευκλείδειας απόστασης, χρησιμοποιείται η ακόλουθη μετρική στο $\mathbb{R}^l$:

$$d(\mathbf{x}_i, \boldsymbol{\theta}_j) = |x_{im} - \theta_{jm}| + \frac{1}{l/2+1} \sum_{r=1, r\neq m}^{l} |x_{ir} - \theta_{jr}|,$$

όπου $|x_{im} - \theta_{jm}| = \max_{r=1,2,\ldots,l} |x_{ir} - \theta_{jr}|$.

Εφαρμόζουμε τον παραπάνω αλγόριθμο για τον διαχωρισμό των σημείων $x_1 = (0, 1)$, $x_2 = (1, 2)$, $x_3 = (2, 1)$, $x_4 = (0,-1)$, $x_5 = (0,-2)$ σε 2 συστάδες. Αν τα κέντρα των συστάδων αρχικοποιούνται ως $\vartheta_1^{(0)} = (-1, 0)$ και $\vartheta_2^{(0)} = (2, 0)$, να προσδιορίσετε τις θέσεις $\vartheta_1^{(2)}$ και $\vartheta_2^{(2)}$ των δύο κέντρων μετά την ολοκλήρωση της δεύτερης επανάληψης του αλγόριθμου.