



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ  
ΥΠΟΛΟΓΙΣΤΩΝ

ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ  
Γραπτή εξέταση 01/02/22

Επώνυμο: .....

Όνομα: .....

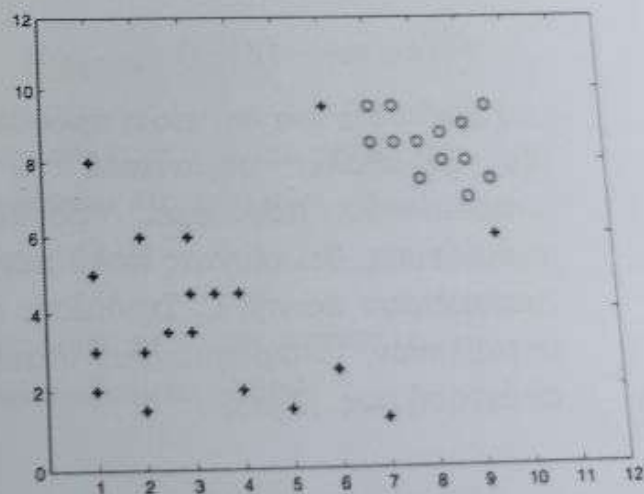
A.M.: .....

Ιδιότητα: .....

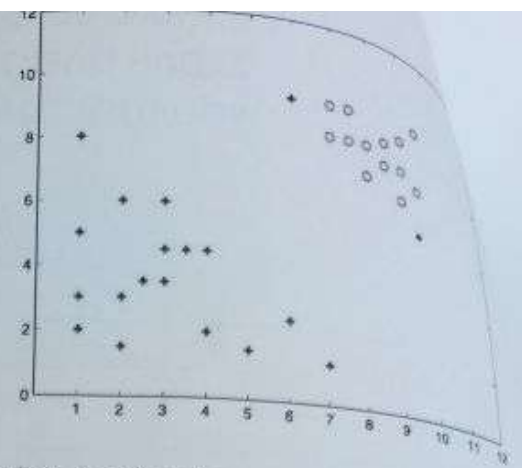
**ΘΕΜΑ 01.** Στο παρακάτω διάγραμμα φαίνονται τα δεδομένα ενός συνόλου εκπαίδευσης που θα χρησιμοποιηθούν για την εκπαίδευση ενός ταξινομητή SVM. Τα δεδομένα προέρχονται από μετρήσεις αισθητήρων που δυνητικά είναι επιρρεπείς σε σφάλματα. Υποθέστε ότι εκπαιδεύουμε ένα SVM με τετραγωνική συνάρτηση πυρήνα (δηλαδή πολυωνυμική συνάρτηση τάξης 2).

Σε καθένα από τα παρακάτω υποερωτήματα απαντήστε σύντομα και περιεκτικά (με 1-2 προτάσεις) στον κενό χώρο πάνω στα θέματα, σχεδιάζοντας ό,τι χρειάζεται στα αντίστοιχα διαγράμματα. [10 pts]

(α) Πού θα περιμένετε να βρίσκεται το διαχωριστικό υπερεπίπεδο για πολύ μεγάλες τιμές της παραμέτρου ποινής  $C$ , δηλαδή όταν  $C \rightarrow \infty$ ; Σχεδιάστε μια πιθανή θέση του διαχωριστικού υπερεπιπέδου στο διπλανό διάγραμμα και αιτιολογήστε την απάντησή σας. [2 pts]

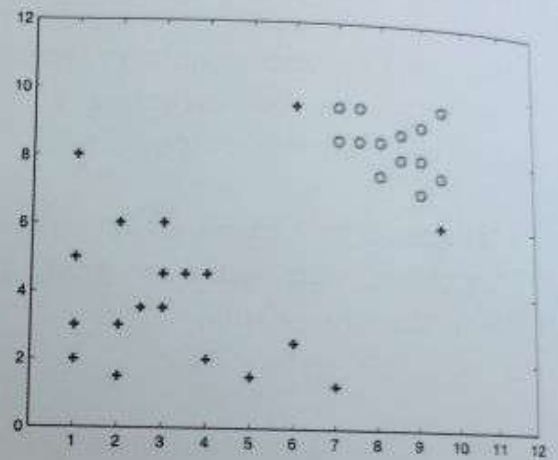


(β) Πού θα περιμένετε να βρίσκεται διαχωριστικό υπερεπίπεδο για πολύ μικρές τιμές της παραμέτρου ποινής  $C$ , δηλαδή όταν  $C \rightarrow 0$ ; Σχεδιάστε μια πιθανή θέση του διαχωριστικού υπερεπίπεδου στο διπλανό διάγραμμα και αιτιολογήστε την απάντησή σας. [2 pts]

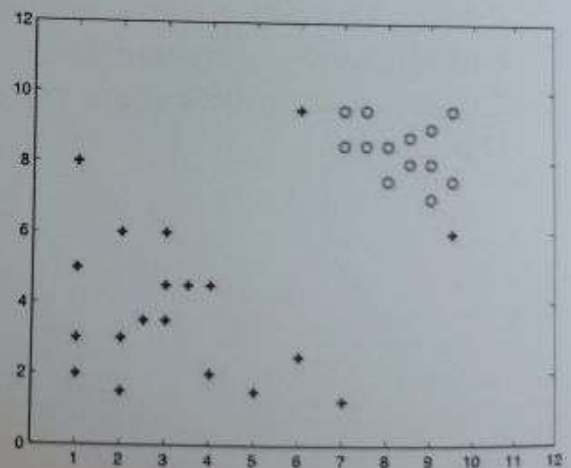


(γ) Ποια από τις παραπάνω δύο περιπτώσεις θεωρείτε ότι θα ήταν καταλληλότερη για το συγκεκριμένο πρόβλημα ταξινόμησης και γιατί; [2 pts]

(δ) Σχεδιάστε ένα σημείο η προσθήκη του οποίου δεν θα μεταβάλει το διαχωριστικό υπερεπίπεδο που έχει προκύψει από την εκπαίδευση, θεωρώντας πολύ μεγάλες τιμές της παραμέτρου ποινής  $C$ . Αιτιολογήστε την απάντησή σας. [2 pts]



(ε) Σχεδιάστε ένα σημείο η προσθήκη του οποίου θα μεταβάλει σημαντικά το διαχωριστικό υπερεπίπεδο που έχει προκύψει από την εκπαίδευση, θεωρώντας πολύ μεγάλες τιμές της παραμέτρου ποινής  $C$ . Σχεδιάστε επίσης το νέο, μεταβληθέν υπερεπίπεδο. Αιτιολογήστε την απάντησή σας. [2 pts]



**ΘΕΜΑ 02.** Επιλέξτε ποια ή ποιες από τις παρακάτω προτάσεις είναι ΣΩΣΤΗ-/ΕΣ κυκλώνοντας τα αντίστοιχα γράμματα πάνω στα θέματα (υπάρχει αρνητική βαθμολογία). [4 pts]



(α) Η αρχικοποίηση των βαρών ενός νευρωνικού δικτύου (ΝΔ) με την ίδια τυχαία αρχική τιμή είναι καλή πρακτική στην εκπαίδευση ΝΔ, καθώς επιτρέπει την κατ' αρχάς «ισότιμη» αντιμετώπιση των διαφορετικών συνάψεων του ΝΔ.

(β) Στην εκπαίδευση ενός MLP καλό είναι να αποφεύγονται πολύ μεγάλες τιμές του ρυθμού μάθησης, καθώς αυτό μπορεί να οδηγήσει σε μεγάλους χρόνους καθόδου προς το τοπικό βέλτιστο.

(γ) Η δυνατότητα της αυτόματης διαφόρισης (automatic differentiation) και η συνάρτηση ενεργοποίησης ReLU είναι δύο παράγοντες που συνετέλεσαν καθοριστικά στην αποτελεσματικότητα και επιτυχία των μεθόδων βαθιάς μάθησης τα τελευταία 15 χρόνια.

(δ) Έστω νευρωνικό δίκτυο με συνολικά 200 συναπτικά βάρη το οποίο εκπαιδεύεται με σύνολο εκπαίδευσης 10.000 δειγμάτων. Ο συνολικός αριθμός ανανεώσεων βαρών στην οπισθοδιάδοση: στη περίπτωση του online training θα είναι 2.000.000, ενώ στην περίπτωση του batch training με μέγεθος batch 25 δείγματα/batch θα είναι 80.000.

**ΘΕΜΑ 03.** Ποια από τις επόμενες συναρτήσεις ενεργοποίησης μπορεί να οδηγήσει σε vanishing gradients; [1pt] (υπάρχει αρνητική βαθμολογία)

- (i) ReLU
- (ii) Tanh
- (iii) Leaky ReLU
- (iv) Καμία από τις άλλες

**ΘΕΜΑ 04.** Έστω η πεπερασμένη κλάση υποθέσεων  $H$  και  $D$  μια αυθαίρετη, καθορισμένη και άγνωστη κατανομή ως προς το  $X$ . Για κάθε  $\epsilon, \delta > 0$  αν λάβουμε δείγμα  $S$  από την  $D$  μεγέθους

$$m \geq \frac{1}{2\epsilon^2} \left( \ln |H| + \ln \frac{2}{\delta} \right)$$

τότε με πιθανότητα το λιγότερο  $1 - \delta$ , όλες οι υποθέσεις  $h \in H$  έχουν  $|\text{err}_D(h) - \text{err}_S(h)| \leq \epsilon$ .

1. Δοκιμάζουμε εκπαίδευση με 100 δείγματα και βλέπουμε απόσταση μεταξύ του σφάλματος εκπαίδευσης και του σφάλματος ελέγχου. Αν θέλουμε να μειώσουμε στο μισό την υπερεκπαίδευση, πόσο δείγματα θα χρειαστούμε; [2 pts]

2. Χρησιμοποιούμε τα δείγματα που υπολογίσαμε στο προηγούμενο ερώτημα ωστόσο υπερεκπαίδευση δεν μειώνεται στο μισό. Γιατί μπορεί να συμβαίνει αυτό; [2 pts]

**ΘΕΜΑ 05.** Έστω ότι έχουμε ένα σύστημα online μάθησης το οποίο καλείται να επιλέξει από 4 πιθανές δράσεις (με κωδικό 0 ως 3). Να γράψετε την κατανομή  $p$ , πάνω στις δράσεις καθώς και τα βάρη  $w_{i,j}$  των δράσεων που προκύπτουν μετά την εφαρμογή του πρώτου γύρου του randomized αλγορίθμου σταθμισμένης πλειοψηφίας, αν το διάνυσμα απώλειας  $I$ , του πρώτου γύρου έχει μη-μηδενική τιμή για τις ενέργειες 1 και 3. Θεωρείστε επίσης ότι η παράμετρος  $\beta$  είναι ίση με 0.  $x$ , όπου  $x = 9 - k$  και  $k$  το τελευταίο ψηφίο του αριθμού μητρώου σας (θέστε  $k = 3$  αν το τελευταίο ψηφίο του αριθμού μητρώου σας είναι ίσο με το 9). [4 pts]

**ΘΕΜΑ 06.** Ποια από τα επόμενα είναι αληθή για ένα layer ενός CNN; (επιλέξτε όσα είναι αληθή, υπάρχει αρνητική βαθμολογία) [2 pts]

- (i) Ο αριθμός των βαρών βασίζεται στο βάθος (depth) του σήματος εισόδου στο δίκτυο.
- (ii) Ο αριθμός των κατωφλίων (biases) είναι ίσος με τον αριθμό των φίλτρων.
- (iii) Ο συνολικός αριθμός των παραμέτρων βασίζεται στο stride.
- (iv) Ο συνολικός αριθμός των παραμέτρων βασίζεται στο padding.

**ΘΕΜΑ 07.** Έστω ότι έχουμε ένα ταξινομητή που για ένα δείγμα επιστρέφει τους  $j$  πλησιέστερους γείτονες και τις κλάσεις τους. Μπορούμε με αυτόν τον ταξινομητή να φτιάξουμε έναν  $k$ -NN, όταν  $j \neq k$  (ναι/όχι) και γιατί; [4 pts]

**ΘΕΜΑ 08.** Έστω ότι έχουμε ένα ομαλοποιημένο μοντέλο γραμμικής παλινδρόμησης  $\arg\min_w \|Y - Xw\|_2^2 + \lambda \|w\|_p^p$ . Ποια είναι η επίδραση της αύξησης του  $p$  ( $p \geq 1$ ) στην πόλωση (bias) και τη διακύμανση (variance) όταν τα βάρη είναι μεγαλύτερα από 1; [2pts] (υπάρχει αρνητική βαθμολογία)

- (α) αύξηση της πόλωσης, αύξηση της διακύμανσης
- (β) αύξηση της πόλωσης, μείωση της διακύμανσης
- (γ) μείωση της πόλωσης, αύξηση της διακύμανσης
- (δ) μείωση της πόλωσης, μείωση της διακύμανσης
- (ε) οι πληροφορίες δεν επαρκούν για να ξέρουμε τί θα γίνει

**ΘΕΜΑ 09.** Έστω ότι έχουμε το παρακάτω μη-γραμμικά διαχωρίσιμο πρόβλημα δυαδικής ταξινόμησης στις δύο διαστάσεις (όπου τα σημεία με 'x' ανήκουν στη μια κλάση και τα σημεία με 'o' στην άλλη).





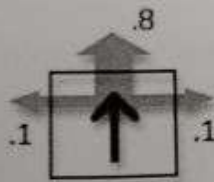
γραμμικά διαχωρίσιμο; [1pt] (υπάρχει αρνητική βαθμολογία)

- A. Ένας πολυωνυμικός πυρήνας
- B. Ένας γκαουσιανός πυρήνας
- Γ. Ένας σιγμοειδής πυρήνας
- Δ. Κανένας από τους παραπάνω

**ΘΕΜΑ 10.** Ένας πράκτορας κινείται σε ένα περιβάλλον αποτελούμενο από ένα πλέγμα τετραγώνων (α). Τα τετράγωνα είναι οι καταστάσεις του πράκτορα, και αναγνωρίζονται από τον αριθμό της γραμμής και της στήλης τους (πρώτη η γραμμή). Ο πράκτορας ξεκινά πάντα στην κατάσταση (1,1). Υπάρχουν δύο τερματικές καταστάσεις η (2,3) με ανταμοιβή +5 και η (1,3) με ανταμοιβή -5. Οι ανταμοιβές είναι 0 στις μη τερματικές καταστάσεις. Η κίνηση του πράκτορα είναι ελαφρά στοχαστική: η συνάρτηση μετάβασης είναι τέτοια ώστε ο πράκτορας κινείται με πιθανότητα 0,8 προς την κατάσταση που θέλει αλλά και με πιθανότητα 0.1 προς κάθε μία από τις κάθετες καταστάσεις ως προς την επιθυμητή (b). Αν συμβεί σύγκρουση με τοίχο, ο πράκτορας παραμένει στην ίδια κατάσταση.

		+5
S		-5

(a)



(b)

1. Ποια είναι η βέλτιστη πολιτική (policy)  $\pi^*$  για κάθε τετράγωνο του πλέγματος; [1 pt]
2. Υπολογίστε δύο γύρους της επανάληψης αξίας (value iteration) αν ο παράγοντας έκπτωσης είναι 0.9. [3 pts]

**ΘΕΜΑ 11.** Σχεδιάζετε ένα σύστημα βαθιάς μάθησης για την ανίχνευση της κόπωσης των οδηγών στα αυτοκίνητα με σκοπό την αποφυγή τυχόν ατυχημάτων. Ποια από τις επόμενες είναι η καταλληλότερη μέτρηση αξιολόγησης:

- (i) Accuracy
- (ii) Precision
- (iii) Recall
- (iv) Loss Value

Να εξηγήσετε την επιλογή σας. [2pts]

**ΘΕΜΑ 12.** Σε μια μηχανή αναζήτησης θέτουμε τα ερωτήματα (queries)  $q_1$  και  $q_2$ . Η μηχανή επιστρέφει έγγραφα (documents) για το κάθε query με την εξής σειρά:

$$q_1 \rightarrow d_1, d_2$$

$$q_2 \rightarrow d_2, d_4, d_5$$

Αν θεωρήσουμε ότι μόνο τα έγγραφα  $d_1, d_4, d_5$  είναι σχετικά ως προς το αντίστοιχο ερώτημα, με ποιον τρόπο μπορούμε να αξιολογήσουμε την ποιότητα των κατατάξεων λαμβάνοντας υπόψη τη σειρά των αποτελεσμάτων και ποια κατάταξη είναι καλύτερη με βάση το κριτήριο αυτό; [2 pts]

**ΘΕΜΑ 13.** Θα χρησιμοποιήσετε ένα ταξινομητή CNN. Για κάθε layer του παρακάτω δικτύου, υπολογίστε τον αριθμό των βαρών, τον αριθμό των biases και το μέγεθος των feature/output maps. [4pts]

- Το CONV-K-N υποδηλώνει ένα συνελκτικό layer με N φίλτρα, το καθένα με μέγεθος  $K \times K$ . Οι παράμετροι padding και stride είναι 0 και 1 αντίστοιχα.

- Το POOL-K υποδηλώνει ένα στρώμα συγκέντρωσης  $K \times K$  με stride K και padding 0.

- Το FC-N είναι ένα πλήρως συνδεδεμένο layer με N νευρώνες.

Είσοδος:  $128 \times 128 \times 3$

CONV-9-32

POOL-2

FC-3

**ΘΕΜΑ 14.** Διαθέτουμε ένα σύνολο δεδομένων χωρίς ετικέτες. Χρησιμοποιούμε έναν αλγόριθμο κεντροειδούς ομαδοποίησης όπως ο k-Means. Τί σημαίνει η ελαχιστοποίηση του κριτηρίου της

αδράνειας (inertia)  $\sum_{i=1}^N (x_i - C_k)^2$ , όπου N το πλήθος των δειγμάτων και  $C_k$  τα κέντρα των k clusters, για την ποιότητα του clustering; [3 pts]

**ΘΕΜΑ 15.** Έστω  $\mathcal{H}$  η κλάση των προσημασμένων διαστημάτων, δηλαδή

$$\mathcal{H} = \{h_{a,b,s} : a \leq b, s \in \{-1, 1\}\}$$

όπου

$$h_{a,b,s}(x) = \begin{cases} s & x \in [a, b] \\ -s & x \notin [a, b] \end{cases}$$

Αποδείξτε ποιο είναι το  $VCdim(\mathcal{H})$ . [5 pts]

**ΘΕΜΑ 16.** Έστω ότι έχουμε πρόβλημα δυαδικής ταξινόμησης (ετικέτες  $\{-1, +1\}$ ), το οποίο επιθυμούμε να μάθουμε τη χρήση ασθενών μοντέλων μάθησης και της τεχνικής AdaBoost. Έστω επίσης ότι διαθέτουμε ένα σύνολο εκπαίδευσης 10 δειγμάτων του εν λόγω προβλήματος και ότι μετά την πρώτη επανάληψη του

αλγορίθμου, το σφάλμα ταξινόμησης  $\epsilon_1$  είναι ίσο με  $\frac{x}{20}$ , όπου x το τελευταίο ψηφίο του αριθμού μητρώου σας (θεωρείστε  $x=3$  στην περίπτωση που το τελευταίο ψηφίο του αριθμού μητρώου σας είναι το 0). Έστω

επίσης ότι  $\text{floor}(\frac{x}{2})$  δείγματα έχουν ταξινομηθεί λανθασμένα και τα υπόλοιπα ορθά. Πως θα

διαμορφωθεί η κατανομή επιλογής των δειγμάτων  $D_2(i)$  στη δεύτερη επανάληψη του AdaBoost; [4 pts]

**ΘΕΜΑ 17.** Έχετε ένα σύνολο δεδομένων D1 με 1 εκατομμύριο labelled training data για ταξινόμηση και ένα άλλο σύνολο δεδομένων D2 με 100 labelled training data. Ενας συνάδελφός σας εκπαιδεύει ένα μοντέλο από την αρχή στο σύνολο δεδομένων D2. Εσείς αποφασίζετε να εκπαιδεύσετε ένα μοντέλο στο D1 και, στη συνέχεια, να εφαρμόστε transfer learning στο D2. Ποιες είναι οι παράμετροι (hyperparameters) που πρέπει να επιλέξετε στην προσέγγισή σας; Αναφέρετε ένα πρόβλημα που είναι πιθανό να βρει ο φίλος σας με την προσέγγισή του. Πώς αντιμετωπίζει η δική σας προσέγγιση αυτό το πρόβλημα; [4 pts]