

Πρόοδος Μηχανικής Μάθησης Mod 1

Multiple Choice μίας απάντησης: σωστή απάντηση +3, λάθος -1

Multiple Choice πολλαπλών απαντήσεων: κάθε σωστή απάντηση +1, κάθε λάθος -1

* Required

* This form will record your name, please fill your name.

1

Επώνυμο *

2

Όνομα *

3

AM *

The value must be a number

4

Ιδιότητα *

ΔΠΜΣ, ΥΔ κλπ

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

Επιλέξτε ποιες από τις παρακάτω προτάσεις για τα Support Vector Machines (SVMs) είναι **ΛΑΝΘΑΣΜΕΝΕΣ**:

- ☐ 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.
- ☐ Ένα σημαντικό μειονέκτημα των SVMs είναι ότι συχνά παγιδεύονται σε τοπικά ελάχιστα, σε αντίθεση με τα MLPs.
- ☐ Στην εικόνα υπενθυμίζεται το πρόβλημα SVM με χρήση μεταβλητών χαλαρότητας. Όσο αυξάνεται η τιμή του C , τόσο αυξάνεται η ανοχή μας σε λανθασμένες ταξινομήσεις.
- ☐ Αν έχουμε ένα πρόβλημα ταξινόμησης τριών (3) κλάσεων, τότε το πλήθος των SVM που θα πρέπει να εκπαιδεύσουμε αν ακολουθήσουμε τη μέθοδο one-against-one είναι ίσο με το πλήθος των SVM που θα πρέπει να εκπαιδεύσουμε αν ακολουθήσουμε τη μέθοδο one-against-all.
- ☐ Η εισαγωγή μεταβλητών χαλαρότητας (slack variables) καθιστά ένα γραμμικό SVM ικανό να χρησιμοποιηθεί αποτελεσματικά για την επιτυχία (δηλαδή χωρίς εσφαλμένες κατηγοριοποιήσεις) ταξινόμηση δεδομένων κλάσεων που είναι διαχωρίσιμες από μη γραμμικές επιφάνειες.
- ☐ Αν αφαιρέσουμε ένα σημείο που ταξινομείται ορθά και βρίσκεται μακριά από το όριο απόφασης, τότε το όριο απόφασης (βέλτιστο υπερεπίπεδο διαχωρισμού) δεν θα επηρεαστεί.

6

x_1	x_2	$y = x_1 \cup \neg x_2$
0	0	1
0	1	0
1	0	1
1	1	1

Έστω η λογική συνάρτηση του πίνακα.

(α) Μπορεί η συνάρτηση αυτή να αναπαρασταθεί από απλό perceptron; Εξηγήστε.

(β) Αν στο παραπάνω ερώτημα απαντήσατε καταφατικά, σχεδιάστε ένα απλό perceptron που την αναπαριστά. Αλλιώς, σχεδιάστε ένα Multi-Layer Perceptron που την αναπαριστά.

(5 Points)

⤴ Upload file

File number limit: 3 Single file size limit: 10MB Allowed file types: Word,Excel,PPT,PDF,Image,Video,Audio

7

Στη συσταδοποίηση θέλουμε να επιτύχουμε:

- ☐ Συνοχή εντός του κάθε cluster και διαχωρισμό μεταξύ διαφορετικών clusters
- ☐ Τίποτα από τα υπόλοιπα
- ☐ Δεν απαντώ
- ☐ Διαχωρισμός εντός κάθε cluster και συνοχή μεταξύ διαφορετικών clusters

Έστω νευρωνικό δίκτυο που περιέχει συνολικά 150 συναπτικά βάρη (δεν περιέχει πολώσεις – bias terms) και το επίπεδο εξόδου αποτελείται από 5 εξόδους. Ως συνάρτηση κόστους χρησιμοποιείται το μέσο τετραγωνικό σφάλμα (MSE). Κατά τη διάρκεια της εκπαίδευσης, το νευρωνικό δίκτυο επεξεργάζεται συνολικά 400 minibatches, το καθένα μεγέθους 20 δειγμάτων.

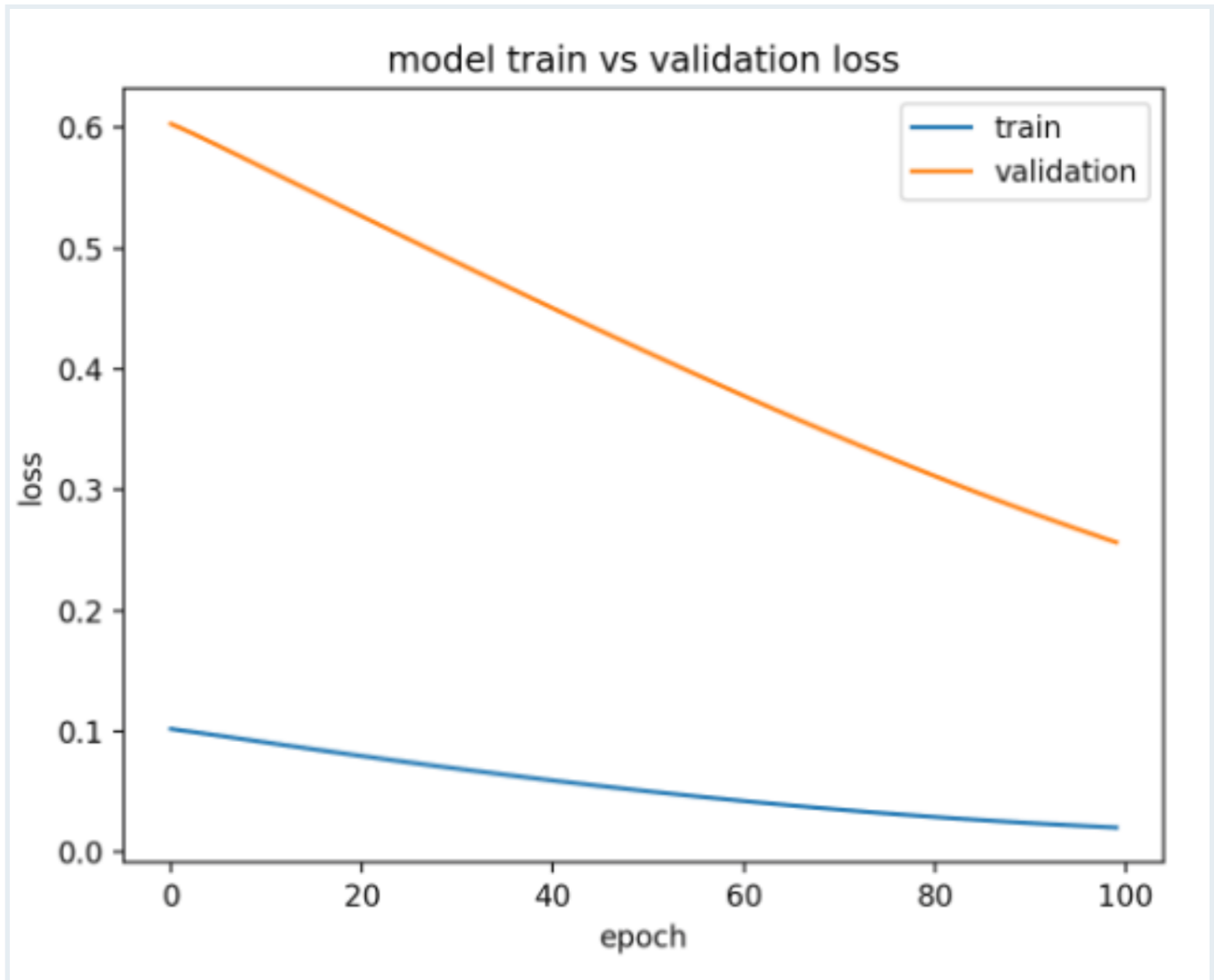
(α) Υπολογίστε τον συνολικό αριθμό των κλίσεων (gradients) του κόστους ως προς ατομικά συναπτικά βάρη που υπολογίζονται στην οπισθοδιάδοση (backpropagation).

(β) Υπολογίστε τον συνολικό αριθμό των ανανεώσεων βαρών που πραγματοποιούνται στην οπισθοδιάδοση (backpropagation).

(4 Points)

Έστω ότι η παρακάτω γραφική παράσταση απεικονίζει τη μεταβολή της συνάρτησης σφάλματος στα δεδομένα εκπαίδευσης (μπλε γραμμή) και ελέγχου (πορτοκαλί γραμμή) για μοντέλο μηχανικής μάθησης, το οποίο εκπαιδεύτηκε για 100 εποχές. Πως θα χαρακτηρίζατε το αποτέλεσμα της εκπαίδευσης;

(3 Points)



STEF

1. Σε ένα πρόβλημα επιβλεπόμενης εκπαίδευσης ενός νευρωνικού δικτύου έχετε ένα σύνολο εκπαίδευσης με μικρό αριθμό από 2-Δ εικόνες. Πώς θα αντιμετωπίζατε αυτό το πρόβλημα;
2. Ποια προβλήματα παρουσιάζει η χρήση της βηματικής συνάρτησης, της συνάρτησης προσήμου, της γραμμικής συνάρτησης, ή της συνάρτησης ReLU ως συνάρτησης ενεργοποίησης των νευρώνων;
(5 Points)

⤴ Upload file

File number limit: 3 Single file size limit: 10MB Allowed file types: Word,Excel,PPT,PDF,Image,Video,Audio

Με ποιούς τρόπους μπορούμε να εξισορροπήσουμε ένα σύνολο δεδομένων;

- ☐ Αφαιρώντας τυχαία δείγματα από όλες τις κατηγορίες που έχουν μεγαλύτερο πλήθος από τη μικρότερη κατηγορία
- ☐ Επιλέγοντας, τυχαία, ορισμένα δείγματα από τις λιγότερο συχνές κατηγορίες και επαναλαμβάνοντας τα.
- ☐ Επιλέγοντας τυχαία δείγματα από όλες τις κατηγορίες που έχουν μεγαλύτερο πλήθος από τη μικρότερη κατηγορία
- ☐ Αφαιρώντας τυχαία δείγματα από τις λιγότερο συχνές κατηγορίες

12

Ποια θα ήταν η σχέση μεταξύ του χρόνου εκπαίδευσης k-NN για $k=1$, $k=2$, $k=3$;

- ☐ 1-NN < 2-NN < 3-NN
- ☐ 1-NN ~ 2-NN ~ 3-NN
- ☐ Τίποτα από τα υπόλοιπα
- ☐ Δεν απαντώ
- ☐ 1-NN > 2-NN > 3-NN

13

Θεωρήστε την κλάση υποθέσεων $H = \{1/10, 2/10, \dots, 9/10\}$, και έστω x πραγματικός αριθμός στο $[0,1]$. Για παράδειγμα ένα από τα μέλη της H είναι η υπόθεση:

$$h(x) = 1 \text{ αν } x \geq 1/10$$

$$h(x) = 0 \text{ αλλιώς}$$

(α) Πόσα δείγματα εκπαίδευσης αρκούν για την αγνωστική εκμάθηση PAC αυτής της κλάσης υποθέσεων για $\epsilon = 0,01$ και $\delta = 0.05$;

(β) Τί ισχύει για τα το πλήθος των δειγμάτων εκπαίδευσης που αρκούν αν $\epsilon = 0,1$;

(γ) Τί ισχύει για τα το πλήθος των δειγμάτων εκπαίδευσης που αρκούν αν το x μπορεί να είναι οποιοσδήποτε πραγματικός αριθμός;(5 Points)

 Upload file

File number limit: 3 Single file size limit: 10MB Allowed file types: Word,Excel,PPT,PDF,Image,Video,Audio

(5 Points)

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 $i^{\text{στου}}$ νευρώνα του επιπέδου U , v_k η έξοδος του $k^{\text{στου}}$ νευρώνα του επιπέδου V , $u_i=0.8$, $v_k=0.4$ και το βάρος μεταξύ των δύο νευρώνων $w_{i,k}=0.3$:

- (α) Υπολογίστε την τιμή της μερικής παραγώγου $\frac{\partial v_k}{\partial w_{i,k}}$
- (β) Υπολογίστε την τιμή της μερικής παραγώγου $\frac{\partial v_k}{\partial u_i}$

⤴ Upload file

File number limit: 3 Single file size limit: 10MB Allowed file types: Word,Excel,PPT,PDF,Image,Video,Audio

Έστω ότι εκπαιδεύουμε μοντέλο μηχανικής μάθησης δέκα παραμέτρων εφαρμόζοντας τεχνικές ομαλοποίησης και οι τιμές τους μετά το πέρας της εκπαίδευσης διαμορφώνονται όπως στο παρακάτω διάνυσμα:

$[-0.59342468, -2.45993074, 0.28591644, -2.17031788, 6.46296243, -0.7626098, -0.28098948, 0.71953103, 0.08476217, -5.32733193]$

Ποια από τις τεχνικές ομαλοποίησης που έχουμε εξετάσει έχει εφαρμοστεί σε αυτή την περίπτωση. Τεκμηριώστε σύντομα την απάντησή σας.

(3 Points)

This content is neither created nor endorsed by Microsoft. The data you submit will be sent to the form owner.

 Microsoft Forms

ΟΜΑΔΑ 2

Με ποιούς τρόπους μπορούμε να εξισορροπήσουμε ένα σύνολο δεδομένων;

- ☐ Επιλέγοντας, τυχαία, ορισμένα δείγματα από τις λιγότερο συχνές κατηγορίες και επαναλαμβάνοντας τα.
- ☐ Επιλέγοντας τυχαία δείγματα από όλες τις κατηγορίες που έχουν μεγαλύτερο πλήθος από τη μικρότερη κατηγορία
- ☐ Αφαιρώντας τυχαία δείγματα από όλες τις κατηγορίες που έχουν μεγαλύτερο πλήθος από τη μικρότερη κατηγορία
- ☐ Αφαιρώντας τυχαία δείγματα από τις λιγότερο συχνές κατηγορίες

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

Επιλέξτε ποιες από τις παρακάτω προτάσεις για τα Support Vector Machines (SVMs) είναι **ΣΩΣΤΕΣ**:



- ☐ Μια διαφορά μεταξύ ενός MLP δύο στρωμάτων και ενός SVM είναι ότι στο πρώτο οι παράμετροι προκύπτουν από εκπαίδευση ενώ στο δεύτερο από την επίλυση προβλήματος τετραγωνικού προγραμματισμού.
- ☐ Αν έχουμε ένα πρόβλημα ταξινόμησης τριών (3) κλάσεων, τότε το πλήθος των SVM που θα πρέπει να εκπαιδεύσουμε αν ακολουθήσουμε τη μέθοδο one-against-one είναι μεγαλύτερο από το πλήθος των SVM που θα πρέπει να εκπαιδεύσουμε αν ακολουθήσουμε τη μέθοδο one-against-all.
- ☐ Αν αφαιρέσουμε ένα σημείο που ταξινομείται ορθά και βρίσκεται μακριά από το όριο απόφασης, τότε το όριο απόφασης (βέλτιστο υπερεπίπεδο διαχωρισμού) δεν θα επηρεαστεί.
- ☐ Η χρήση συναρτήσεων πυρήνα (kernel functions) βοηθά να υπολογίσουμε τις τιμές της συνάρτησης μετασχηματισμού $\Phi()$ όπου αυτή εμφανίζεται μόνη της στις πράξεις.
- ☐ Στην εικόνα υπενθυμίζεται το πρόβλημα SVM με χρήση μεταβλητών χαλαρότητας. Όσο μειώνεται η τιμή του C , τόσο μειώνεται η έμφασή που δίνουμε στην ορθή ταξινόμηση των δειγμάτων.
- ☐ 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.

(Non-anonymous question ⓘ)

(5 Points)

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 $i^{\text{οτού}}$ νευρώνα του επιπέδου U , v_k η έξοδος του $k^{\text{οτού}}$ νευρώνα του επιπέδου V , $u_i=0.7$, $v_k=0.5$ και το βάρος μεταξύ των δύο νευρώνων $w_{i,k}=0.4$:

- (α) Υπολογίστε την τιμή της μερικής παραγώγου $\frac{\partial v_k}{\partial u_i}$
- (β) Υπολογίστε την τιμή της μερικής παραγώγου $\frac{\partial v_k}{\partial w_{i,k}}$

 Upload file

Έστω ότι εκπαιδεύουμε μοντέλο μηχανικής μάθησης δέκα παραμέτρων εφαρμόζοντας τεχνικές ομαλοποίησης και οι τιμές τους μετά το πέρας της εκπαίδευσης διαμορφώνονται όπως στο παρακάτω διάνυσμα:

[4.71338284, 0, 0, -0.97013357, 4.75002806, 0, 0, 0, -3.77642542, 0]

Ποια από τις τεχνικές ομαλοποίησης που έχουμε εξετάσει έχει εφαρμοστεί σε αυτή την περίπτωση. Τεκμηριώστε σύντομα την απάντησή σας.

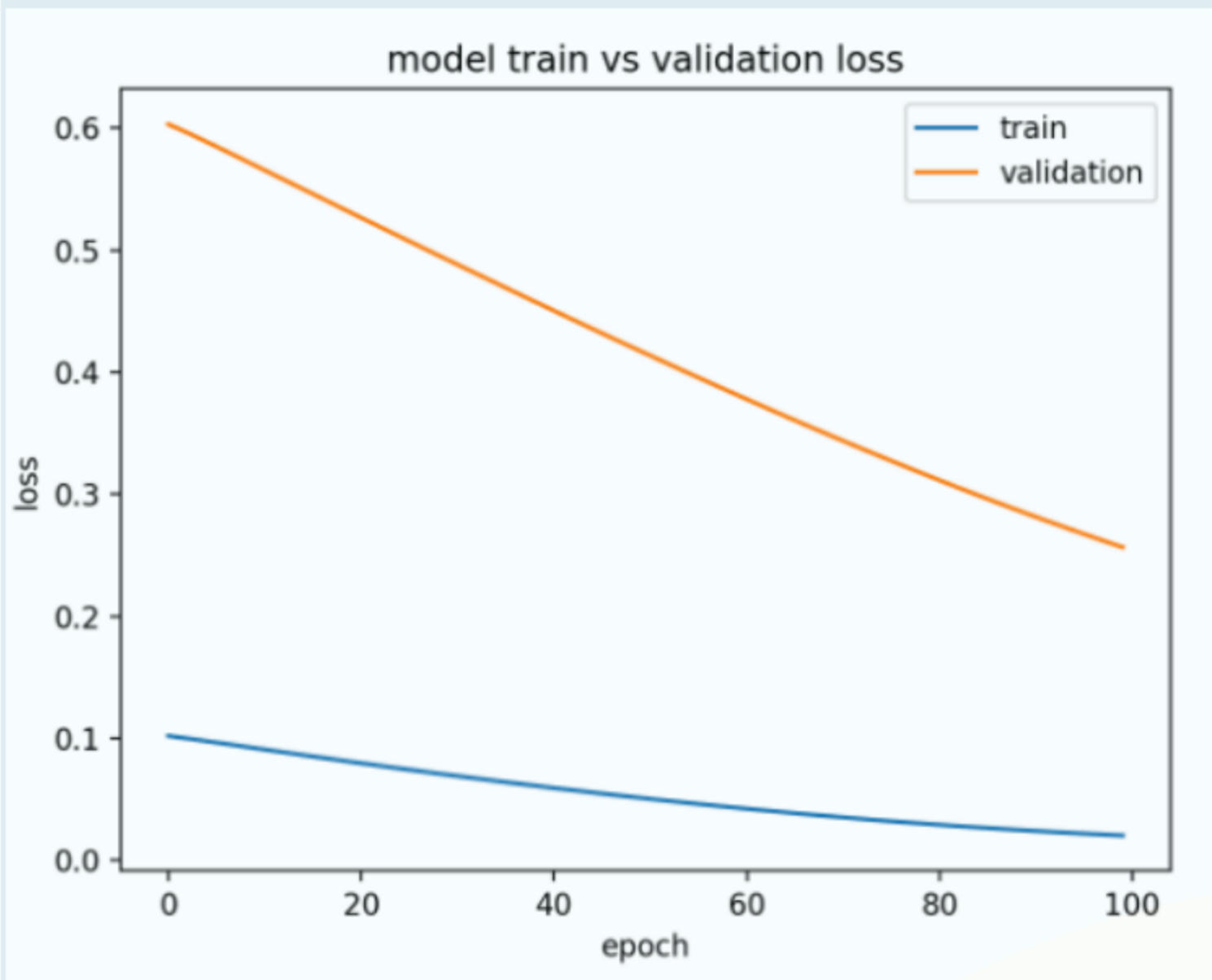


(3 Points)

Enter your answer

Έστω ότι η παρακάτω γραφική παράσταση απεικονίζει τη μεταβολή της συνάρτησης σφάλματος στα δεδομένα εκπαίδευσης (μπλε γραμμή) και ελέγχου (πορτοκαλί γραμμή) για μοντέλο μηχανικής μάθησης, το οποίο εκπαιδεύτηκε για 100 εποχές. Πως θα χαρακτηρίζατε το αποτέλεσμα της εκπαίδευσης;

(3 Points)



Enter your answer

Για οποιουσδήποτε ακέραιους $a \leq b, c \leq d \in [0, n - 1]$, έστω

$h(x_1, x_2) = 1$ αν $a \leq x_1 \leq b$ και $c \leq x_2 \leq d$,

$h(x_1, x_2) = 0$ αλλιώς.

Συμβολίζουμε με H την κλάση όλων των ορθογωνίων που είναι ευθυγραμμισμένα με τους άξονες.

(α) Ως συνάρτηση του n , πόσα διαφορετικά ορθογώνια υπάρχουν στην H ;

(β) Έστω $n = 100$. Πόσα δείγματα εκπαίδευσης αρκούν για να εξασφαλιστεί ότι για κάθε $f \in H$, κάθε συνεπής ταξινομητής που χρησιμοποιεί το H θα εξάγει, με πιθανότητα τουλάχιστον 95%, μια υπόθεση με σφάλμα το πολύ 0,15;

(γ) Περιγράψτε έναν συνεπή ταξινομητή για την κλάση υποθέσεων των ορθογωνίων που είναι ευθυγραμμισμένα στους άξονες.

(Non-anonymous question ①)

(5 Points)

⤴ Upload file

11

Είναι δυνατόν η εκχώρηση παρατηρήσεων σε συστάδες να μην αλλάζει μεταξύ διαδοχικών επαναλήψεων στο K-Means;

- ☐ Όχι
- ☐ Δεν απαντώ
- ☐ Ναι

12

Έστω νευρωνικό δίκτυο που περιέχει συνολικά 150 συναπτικά βάρη (δεν περιέχει πολώσεις – bias terms) και το επίπεδο εξόδου αποτελείται από 5 εξόδους. Ως συνάρτηση κόστους χρησιμοποιείται το μέσο τετραγωνικό σφάλμα (MSE). Κατά τη διάρκεια της εκπαίδευσης, το νευρωνικό δίκτυο επεξεργάζεται συνολικά 400 minibatches, το καθένα μεγέθους 20 δειγμάτων.

(α) Υπολογίστε τον συνολικό αριθμό των κλίσεων (gradients) του κόστους ως προς ατομικά συναπτικά βάρη που υπολογίζονται στην οπισθοδιάδοση (backpropagation).

(β) Υπολογίστε τον συνολικό αριθμό των ανανεώσεων βαρών που πραγματοποιούνται στην οπισθοδιάδοση (backpropagation).


(4 Points)

Enter your answer


13

1. Σε ένα πρόβλημα επιβλεπόμενης εκπαίδευσης ενός νευρωνικού δικτύου έχετε ένα σύνολο εκπαίδευσης με μικρό αριθμό από 2-Δ εικόνες. Πώς θα αντιμετωπίζατε αυτό το πρόβλημα;

2. Ποια προβλήματα παρουσιάζει η χρήση της βηματικής συνάρτησης, της συνάρτησης προσήμου, της γραμμικής συνάρτησης, ή της συνάρτησης ReLU ως συνάρτησης ενεργοποίησης των νευρώνων;

(Non-anonymous question ⓘ) 

(5 Points)

 Upload file

File number limit: 3 Single file size limit: 10MB Allowed file types: Word, Excel, PPT, PDF, Image, Video, Audio

14

Στο k-NN τι θα συμβεί όταν αυξήσετε/μειώσετε την τιμή του k;


- ☐ Η ομαλότητα του ορίου δεν εξαρτάται από την τιμή του k
- ☐ Το όριο γίνεται πιο ομαλό με την αύξηση της τιμής του k
- ☐ Δεν απαντώ
- ☐ Τίποτα από τα υπόλοιπα
- ☐ Το όριο γίνεται πιο ομαλό με την μείωση της τιμής του k

x_1	x_2	$y = \neg x_1 \cup x_2$
0	0	1
0	1	1
1	0	0
1	1	1


Έστω η λογική συνάρτηση του πίνακα

(α) Μπορεί η συνάρτηση αυτή να αναπαρασταθεί από απλό perceptron; Εξηγήστε.

(β) Αν στο παραπάνω ερώτημα απαντήσατε καταφατικά, σχεδιάστε ένα απλό perceptron που την αναπαριστά. Αλλιώς, σχεδιάστε ένα Multi-Layer Perceptron που την αναπαριστά.

(Non-anonymous question ①) 

(5 Points)

 Upload file

ΟΜΑΔΑ 3

4

Ιδιότητα *

ΔΠΜΣ, ΥΔ κλπ

ΔΠΜΣ ΕΔΕΜΜ

5

Είναι δυνατόν η εκχώρηση παρατηρήσεων σε συστάδες να μην αλλάζει μεταξύ διαδοχικών επαναλήψεων στο K-Means;

- ☐ Όχι
- ☐ Δεν απαντώ
- ☐ Ναι

Έστω ότι εκπαιδεύουμε μοντέλο μηχανικής μάθησης δέκα παραμέτρων εφαρμόζοντας τεχνικές ομαλοποίησης και οι τιμές τους μετά το πέρας της εκπαίδευσης διαμορφώνονται όπως στο παρακάτω διάνυσμα:

[4.71338284, 0, 0, -0.97013357, 4.75002806, 0, 0, 0, -3.77642542, 0]

Ποια από τις τεχνικές ομαλοποίησης που έχουμε εξετάσει έχει εφαρμοστεί σε αυτή την περίπτωση. Τεκμηριώστε σύντομα την απάντησή σας.

(3 βαθμοί)

Εισαγάγετε την απάντησή σας

Έστω νευρωνικό δίκτυο που περιέχει συνολικά 100 συναπτικά βάρη (δεν περιέχει πολώσεις – bias terms) και το επίπεδο εξόδου αποτελείται από 4 εξόδους. Ως συνάρτηση κόστους χρησιμοποιείται το μέσο τετραγωνικό σφάλμα (MSE). Κατά τη διάρκεια της εκπαίδευσης, το νευρωνικό δίκτυο επεξεργάζεται συνολικά 500 minibatches, το καθένα μεγέθους 20 δειγμάτων.

(α) Υπολογίστε τον συνολικό αριθμό των κλίσεων (gradients) του κόστους ως προς ατομικά συναπτικά βάρη που υπολογίζονται στην οπισθοδιάδοση (backpropagation).

(β) Υπολογίστε τον συνολικό αριθμό των ανανεώσεων βαρών που πραγματοποιούνται στην οπισθοδιάδοση (backpropagation).



(4 βαθμοί)

Εισαγάγετε την απάντησή σας

Με ποιούς τρόπους μπορούμε να εξισορροπήσουμε ένα σύνολο δεδομένων;

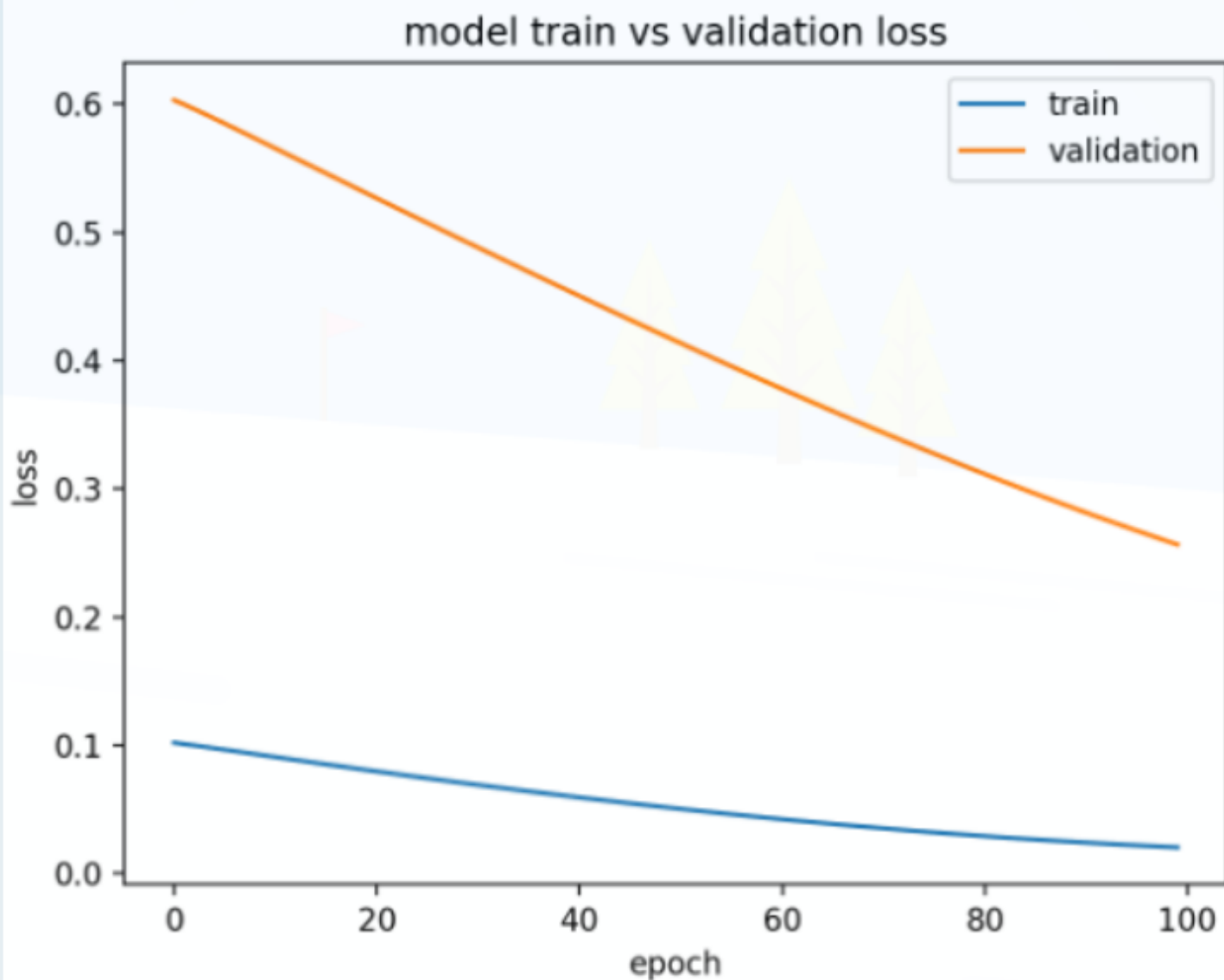


- ☐ Αφαιρώντας τυχαία δείγματα από όλες τις κατηγορίες που έχουν μεγαλύτερο πλήθος από τη μικρότερη κατηγορία
- ☐ Αφαιρώντας τυχαία δείγματα από τις λιγότερο συχνές κατηγορίες
- ☐ Επιλέγοντας τυχαία δείγματα από όλες τις κατηγορίες που έχουν μεγαλύτερο πλήθος από τη μικρότερη κατηγορία
- ☐ Επιλέγοντας, τυχαία, ορισμένα δείγματα από τις λιγότερο συχνές κατηγορίες και επαναλαμβάνοντας τα.

Έστω ότι η παρακάτω γραφική παράσταση απεικονίζει τη μεταβολή της συνάρτησης σφάλματος στα δεδομένα εκπαίδευσης (μπλε γραμμή) και ελέγχου (πορτοκαλί γραμμή) για μοντέλο μηχανικής μάθησης, το οποίο εκπαιδεύτηκε για 100 εποχές. Πως θα χαρακτηρίζατε το αποτέλεσμα της εκπαίδευσης;



(3 βαθμοί)




x_1	x_2	$y = \neg x_1 \cup x_2$
0	0	1
0	1	1
1	0	0
1	1	1

Έστω η λογική συνάρτηση του πίνακα

(α) Μπορεί η συνάρτηση αυτή να αναπαρασταθεί από απλό perceptron; Εξηγήστε.

(β) Αν στο παραπάνω ερώτημα απαντήσατε καταφατικά, σχεδιάστε ένα απλό perceptron που την αναπαριστά. Αλλιώς, σχεδιάστε ένα Multi-Layer Perceptron που την αναπαριστά.


(Μη ανώνυμη ερώτηση ①) 

(5 βαθμοί)

 Αποστολή αρχείου

Ποια θα ήταν η σχέση μεταξύ του χρόνου εκπαίδευσης k-NN για $k=1$, $k=2$, $k=3$;

- ☐ 1-NN > 2-NN > 3-NN
- ☐ Δεν απαντώ
- ☐ Τίποτα από τα υπόλοιπα
- ☐ 1-NN ~ 2-NN ~ 3-NN
- ☐ 1-NN < 2-NN < 3-NN

(Μη ανώνυμη ερώτηση ①) 

(5 βαθμοί)

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 $i^{\text{στου}}$ νευρώνα του επιπέδου U , v_k η έξοδος του $k^{\text{στου}}$ νευρώνα του επιπέδου V , $u_i=0.8$, $v_k=0.4$ και το βάρος μεταξύ των δύο νευρώνων $w_{i,k}=0.3$:

- (α) Υπολογίστε την τιμή της μερικής παραγώγου $\frac{\partial v_k}{\partial w_{i,k}}$
- (β) Υπολογίστε την τιμή της μερικής παραγώγου $\frac{\partial v_k}{\partial u_i}$

 Αποστολή αρχείου

1. Σε ένα πρόβλημα επιβλεπόμενης εκπαίδευσης ενός νευρωνικού δικτύου έχετε ένα σύνολο εκπαίδευσης με μικρό αριθμό από 2-Δ εικόνες. Πώς θα αντιμετωπίζατε αυτό το πρόβλημα;
2. Ποια προβλήματα παρουσιάζει η χρήση της βηματικής συνάρτησης, της συνάρτησης προσήμου, της γραμμικής συνάρτησης, ή της συνάρτησης ReLU ως συνάρτησης ενεργοποίησης των νευρώνων;

(Μη ανώνυμη ερώτηση ⓘ)

(5 βαθμοί)

↑ Αποστολή αρχείου

Όριο πλήθους αρχείων: 3 Όριο μεγέθους μεμονωμένου αρχείου: 10MB Επιτρεπόμενοι τύποι αρχείων: Word, Excel, PPT, PDF, Εικόνα, Βίντεο, Ήχος

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

Επιλέξτε ποιες από τις παρακάτω προτάσεις για τα Support Vector Machines (SVMs) είναι **ΣΩΣΤΕΣ**:

- ☐ 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.
- ☐ Μια διαφορά μεταξύ ενός MLP δύο στρωμάτων και ενός SVM είναι ότι στο πρώτο οι παράμετροι προκύπτουν από εκπαίδευση ενώ στο δεύτερο από την επίλυση προβλήματος τετραγωνικού προγραμματισμού.
- ☐ Η χρήση συναρτήσεων πυρήνα (kernel functions) βοηθά να υπολογίσουμε τις τιμές της συνάρτησης μετασχηματισμού $\Phi()$ όπου αυτή εμφανίζεται μόνη της στις πράξεις.
- ☐ Αν έχουμε ένα πρόβλημα ταξινόμησης τριών (3) κλάσεων, τότε το πλήθος των SVM που θα πρέπει να εκπαιδεύσουμε αν ακολουθήσουμε τη μέθοδο one-against-one είναι μεγαλύτερο από το πλήθος των SVM που θα πρέπει να εκπαιδεύσουμε αν ακολουθήσουμε τη μέθοδο one-against-all.
- ☐ Στην εικόνα υπενθυμίζεται το πρόβλημα SVM με χρήση μεταβλητών χαλαρότητας. Όσο μειώνεται η τιμή του C , τόσο μειώνεται η έμφασή που δίνουμε στην ορθή ταξινόμηση των δειγμάτων.
- ☐ Αν αφαιρέσουμε ένα σημείο που ταξινομείται ορθά και βρίσκεται μακριά από το όριο απόφασης, τότε το όριο απόφασης (βέλτιστο υπερεπίπεδο διαχωρισμού) δεν θα επηρεαστεί.

Για οποιουσδήποτε ακέραιους $a \leq b, c \leq d \in [0, n - 1]$, έστω

$h(x_1, x_2) = 1$ αν $a \leq x_1 \leq b$ και $c \leq x_2 \leq d$,

$h(x_1, x_2) = 0$ αλλιώς.

Συμβολίζουμε με H την κλάση όλων των ορθογωνίων που είναι ευθυγραμμισμένα με τους άξονες.

(α) Ως συνάρτηση του n , πόσα διαφορετικά ορθογώνια υπάρχουν στην H ;

(β) Έστω $n = 100$. Πόσα δείγματα εκπαίδευσης αρκούν για να εξασφαλιστεί ότι για κάθε $f \in H$, κάθε συνεπής ταξινομητής που χρησιμοποιεί το H θα εξαγάγει, με πιθανότητα τουλάχιστον 95%, μια υπόθεση με σφάλμα το πολύ 0,15;

(γ) Περιγράψτε έναν συνεπή ταξινομητή για την κλάση υποθέσεων των ορθογωνίων που είναι ευθυγραμμισμένα στους άξονες.

(Μη ανώνυμη ερώτηση ①)

(5 βαθμοί)

↑ Αποστολή αρχείου

ΟΜΑΔΑ 4

Για οποιουδήποτε ακέραιους $a \leq b, c \leq d \in [0, n - 1]$, έστω

$h(x_1, x_2) = 1$ αν $a \leq x_1 \leq b$ και $c \leq x_2 \leq d$,

$h(x_1, x_2) = 0$ αλλιώς.

Συμβολίζουμε με H την κλάση όλων των ορθογωνίων που είναι ευθυγραμμισμένα με τους άξονες.

(α) Ως συνάρτηση του n , πόσα διαφορετικά ορθογώνια υπάρχουν στην H ;

(β) Έστω $n = 100$. Πόσα δείγματα εκπαίδευσης αρκούν για να εξασφαλιστεί ότι για κάθε $f \in H$, κάθε συνεπής ταξινομητής που χρησιμοποιεί το H θα εξάγει, με πιθανότητα τουλάχιστον 95%, μια υπόθεση με σφάλμα το πολύ 0,15;

(γ) Περιγράψτε έναν συνεπή ταξινομητή για την κλάση υποθέσεων των ορθογωνίων που είναι ευθυγραμμισμένα στους άξονες.

(Μη ανώνυμη ερώτηση ①)

(5 βαθμοί)

↑ Αποστολή αρχείου

x_1	x_2	$y = x_1 \cup \neg x_2$
0	0	1
0	1	0
1	0	1
1	1	1

Έστω η λογική συνάρτηση του πίνακα.

(α) Μπορεί η συνάρτηση αυτή να αναπαρασταθεί από απλό perceptron; Εξηγήστε.

(β) Αν στο παραπάνω ερώτημα απαντήσατε καταφατικά, σχεδιάστε ένα απλό perceptron που την αναπαριστά. Αλλιώς, σχεδιάστε ένα Multi-Layer Perceptron που την αναπαριστά.

(Μη ανώνυμη ερώτηση ①)

(5 βαθμοί)

↑ Αποστολή αρχείου

Όριο πλήθους αρχείων: 3 Όριο μεγέθους μεμονωμένου αρχείου: 10MB Επιτρεπόμενοι τύποι αρχείων: Word, Excel, PPT, PDF, Εικόνα, Βίντεο, Ήχος

Έστω νευρωνικό δίκτυο που περιέχει συνολικά 100 συναπτικά βάρη (δεν περιέχει πολώσεις – bias terms) και το επίπεδο εξόδου αποτελείται από 4 εξόδους. Ως συνάρτηση κόστους χρησιμοποιείται το μέσο τετραγωνικό σφάλμα (MSE). Κατά τη διάρκεια της εκπαίδευσης, το νευρωνικό δίκτυο επεξεργάζεται συνολικά 500 minibatches, το καθένα μεγέθους 20 δειγμάτων.

(α) Υπολογίστε τον συνολικό αριθμό των κλίσεων (gradients) του κόστους ως προς ατομικά συναπτικά βάρη που υπολογίζονται στην οπισθοδιάδοση (backpropagation).

(β) Υπολογίστε τον συνολικό αριθμό των ανανεώσεων βαρών που πραγματοποιούνται στην οπισθοδιάδοση (backpropagation).



(4 βαθμοί)

Εισαγάγετε την απάντησή σας

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

Επιλέξτε ποιες από τις παρακάτω προτάσεις για τα Support Vector Machines (SVMs) είναι **ΛΑΝΘΑΣΜΕΝΕΣ**:

- ☐ Ένα σημαντικό μειονέκτημα των SVMs είναι ότι συχνά παγιδεύονται σε τοπικά ελάχιστα, σε αντίθεση με τα MLPs.
- ☐ Αν αφαιρέσουμε ένα σημείο που ταξινομείται ορθά και βρίσκεται μακριά από το όριο απόφασης, τότε το όριο απόφασης (βέλτιστο υπερεπίπεδο διαχωρισμού) δεν θα επηρεαστεί.
- ☐ 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.
- ☐ 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 Φ() μόνη της.
- ☐ Στην εικόνα υπενθυμίζεται το πρόβλημα SVM με χρήση μεταβλητών χαλαρότητας. Όσο αυξάνεται η τιμή του C, τόσο αυξάνεται η ανοχή μας σε λανθασμένες ταξινομήσεις.
- ☐ Αν έχουμε ένα πρόβλημα ταξινόμησης τριών (3) κλάσεων, τότε το πλήθος των SVM που θα πρέπει να εκπαιδεύσουμε αν ακολουθήσουμε τη μέθοδο one-against-one είναι ίσο με το πλήθος των SVM που θα πρέπει να εκπαιδεύσουμε αν ακολουθήσουμε τη μέθοδο one-against-all.

9

(Μη ανώνυμη ερώτηση ①)
(5 βαθμοί)

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 $i^{\text{στου}}$ νευρώνα του επιπέδου U , v_k η έξοδος του $k^{\text{στου}}$ νευρώνα του επιπέδου V , $u_i=0.7$, $v_k=0.5$ και το βάρος μεταξύ των δύο νευρώνων $w_{i,k}=0.4$:

(α) Υπολογίστε την τιμή της μερικής παραγώγου $\frac{\partial v_k}{\partial u_i}$

(β) Υπολογίστε την τιμή της μερικής παραγώγου $\frac{\partial v_k}{\partial w_{i,k}}$

↑ Αποστολή αρχείου

Όριο πλήθους αρχείων: 3 Όριο μεγέθους μεμονωμένου αρχείου: 10MB Επιτρεπόμενοι τύποι αρχείων: Word, Excel, PPT, PDF, Εικόνα, Βίντεο, Ήχος

10

Στη συσταδοποίηση θέλουμε να επιτύχουμε:

- ☐ Δεν απαντώ
- ☐ Τίποτα από τα υπόλοιπα
- ☐ Συνοχή εντός του κάθε cluster και διαχωρισμό μεταξύ διαφορετικών clusters
- ☐ Διαχωρισμός εντός κάθε cluster και συνοχή μεταξύ διαφορετικών clusters

11

Με ποιούς τρόπους μπορούμε να εξισορροπήσουμε ένα σύνολο δεδομένων;

- ☐ Αφαιρώντας τυχαία δείγματα από τις λιγότερο συχνές κατηγορίες
- ☐ Αφαιρώντας τυχαία δείγματα από όλες τις κατηγορίες που έχουν μεγαλύτερο πλήθος από τη μικρότερη κατηγορία
- ☐ Επιλέγοντας τυχαία δείγματα από όλες τις κατηγορίες που έχουν μεγαλύτερο πλήθος από τη μικρότερη κατηγορία
- ☐ Επιλέγοντας, τυχαία, ορισμένα δείγματα από τις λιγότερο συχνές κατηγορίες και επαναλαμβάνοντας τα.

12

Στο k-NN τι θα συμβεί όταν αυξήσετε/μειώσετε την τιμή του k;



- ☐ Το όριο γίνεται πιο ομαλό με την μείωση της τιμής του k
- ☐ Δεν απαντώ
- ☐ Η ομαλότητα του ορίου δεν εξαρτάται από την τιμή του k
- ☐ Το όριο γίνεται πιο ομαλό με την αύξηση της τιμής του k
- ☐ Τίποτα από τα υπόλοιπα

13

Έστω ότι εκπαιδεύουμε μοντέλο μηχανικής μάθησης δέκα παραμέτρων εφαρμόζοντας τεχνικές ομαλοποίησης και οι τιμές τους μετά το πέρας της εκπαίδευσης διαμορφώνονται όπως στο παρακάτω διάνυσμα:

[4.71338284, 0, 0, -0.97013357, 4.75002806, 0, 0, 0, -3.77642542, 0]

Ποια από τις τεχνικές ομαλοποίησης που έχουμε εξετάσει έχει εφαρμοστεί σε αυτή την περίπτωση. Τεκμηριώστε σύντομα την απάντησή σας.

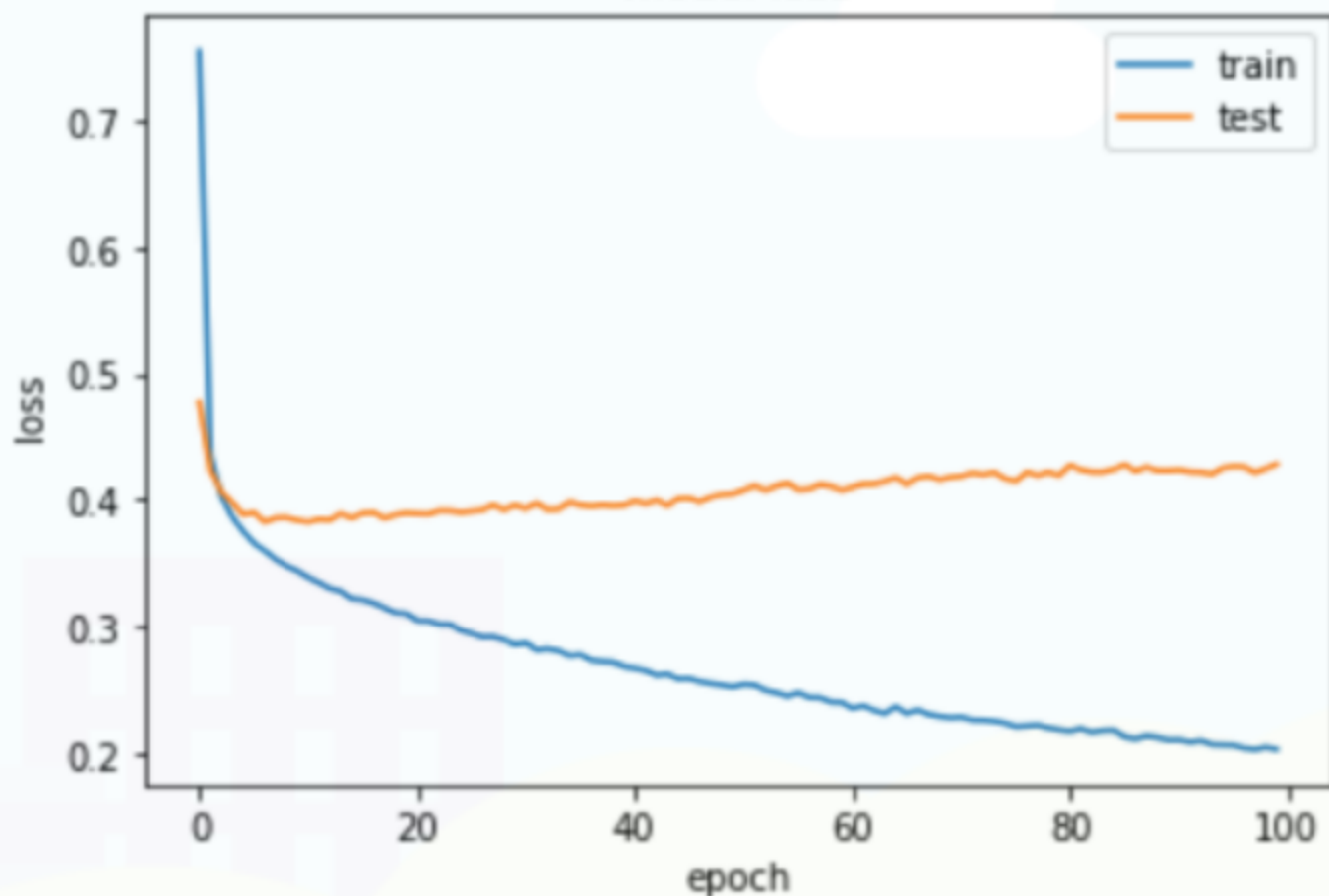
(3 βαθμοί)

Εισαγάγετε την απάντησή σας

Έστω ότι η παρακάτω γραφική παράσταση απεικονίζει τη μεταβολή της συνάρτησης σφάλματος στα δεδομένα εκπαίδευσης (μπλε γραμμή) και ελέγχου (πορτοκαλί γραμμή) για μοντέλο μηχανικής μάθησης, το οποίο εκπαιδεύτηκε για 100 εποχές. Πως θα χαρακτηρίζατε το αποτέλεσμα της εκπαίδευσης;

☐₄₀
(3 βαθμοί)

model loss



Εισαγάγετε την απάντησή σας

1. Σε ένα πρόβλημα επιβλεπόμενης εκπαίδευσης ενός νευρωνικού δικτύου έχετε ένα σύνολο εκπαίδευσης με μικρό αριθμό από 2-Δ εικόνες. Πώς θα αντιμετωπίζατε αυτό το πρόβλημα;

2. Ποια προβλήματα παρουσιάζει η χρήση της βηματικής συνάρτησης, της συνάρτησης προσήμου, της γραμμικής συνάρτησης, ή της συνάρτησης ReLU ως συνάρτησης ενεργοποίησης των νευρώνων;

(Μη ανώνυμη ερώτηση ①)

(5 βαθμοί)

↑ Αποστολή αρχείου

Όριο πλήθους αρχείων: 3 Όριο μεγέθους μεμονωμένου αρχείου: 10MB Επιτρεπόμενοι τύποι αρχείων: Word, Excel, PPT, PDF, Εικόνα, Βίντεο, Ήχος

☐ Να λαμβάνω αποδεικτικό ηλεκτρονικού ταχυδρομείου για τις αποκρίσεις μου

Υποβολή

ΟΜΑΔΑ 5

(Μη ανώνυμη ερώτηση ①)

(5 βαθμοί)

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 $i^{\text{στου}}$ νευρώνα του επιπέδου U , v_k η έξοδος του $k^{\text{στου}}$ νευρώνα του επιπέδου V , $u_i=0.8$, $v_k=0.4$ και το βάρος μεταξύ των δύο νευρώνων $w_{i,k}=0.3$:

(α) Υπολογίστε την τιμή της μερικής παραγώγου $\frac{\partial v_k}{\partial w_{i,k}}$

(β) Υπολογίστε την τιμή της μερικής παραγώγου $\frac{\partial v_k}{\partial u_i}$

Ποια θα ήταν η σχέση μεταξύ του χρόνου εκπαίδευσης k-NN για $k=1$, $k=2$, $k=3$;



- ☐ 1-NN ~ 2-NN ~ 3-NN
- ☐ 1-NN < 2-NN < 3-NN
- ☐ Δεν απαντώ
- ☐ 1-NN > 2-NN > 3-NN
- ☐ Τίποτα από τα υπόλοιπα

Έστω νευρωνικό δίκτυο που περιέχει συνολικά 100 συναπτικά βάρη (δεν περιέχει πολώσεις – bias terms) και το επίπεδο εξόδου αποτελείται από 4 εξόδους. Ως συνάρτηση κόστους χρησιμοποιείται το μέσο τετραγωνικό σφάλμα (MSE). Κατά τη διάρκεια της εκπαίδευσης, το νευρωνικό δίκτυο επεξεργάζεται συνολικά 500 minibatches, το καθένα μεγέθους 20 δειγμάτων.

(α) Υπολογίστε τον συνολικό αριθμό των κλίσεων (gradients) του κόστους ως προς ατομικά συναπτικά βάρη που υπολογίζονται στην οπισθοδιάδοση (backpropagation).

(β) Υπολογίστε τον συνολικό αριθμό των ανανεώσεων βαρών που πραγματοποιούνται στην οπισθοδιάδοση (backpropagation).



(4 βαθμοί)

Εισαγάγετε την απάντησή σας

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


Επιλέξτε ποιες από τις παρακάτω προτάσεις για τα Support Vector Machines (SVMs) είναι **ΛΑΝΘΑΣΜΕΝΕΣ**:




- ☐ 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.
- ☐ Αν αφαιρέσουμε ένα σημείο που ταξινομείται ορθά και βρίσκεται μακριά από το όριο απόφασης, τότε το όριο απόφασης (βέλτιστο υπερεπίπεδο διαχωρισμού) δεν θα επηρεαστεί.
- ☐ Ένα σημαντικό μειονέκτημα των SVMs είναι ότι συχνά παγιδεύονται σε τοπικά ελάχιστα, σε αντίθεση με τα MLPs.
- ☐ Αν έχουμε ένα πρόβλημα ταξινόμησης τριών (3) κλάσεων, τότε το πλήθος των SVM που θα πρέπει να εκπαιδεύσουμε αν ακολουθήσουμε τη μέθοδο one-against-one είναι ίσο με το πλήθος των SVM που θα πρέπει να εκπαιδεύσουμε αν ακολουθήσουμε τη μέθοδο one-against-all.
- ☐ 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.
- ☐ Στην εικόνα υπενθυμίζεται το πρόβλημα SVM με χρήση μεταβλητών χαλαρότητας. Όσο αυξάνεται η τιμή του C , τόσο αυξάνεται η ανοχή μας σε λανθασμένες ταξινομήσεις.

1. Σε ένα πρόβλημα επιβλεπόμενης εκπαίδευσης ενός νευρωνικού δικτύου έχετε ένα σύνολο εκπαίδευσης με μικρό αριθμό από 2-Δ εικόνες. Πώς θα αντιμετωπίζατε αυτό το πρόβλημα;

2. Ποια προβλήματα παρουσιάζει η χρήση της βηματικής συνάρτησης, της συνάρτησης προσήμου, της γραμμικής συνάρτησης, ή της συνάρτησης ReLU ως συνάρτησης ενεργοποίησης των νευρώνων;

(Μη ανώνυμη ερώτηση ⓘ) 

(5 βαθμοί)

 Αποστολή αρχείου

Όριο πλήθους αρχείων: 3 Όριο μεγέθους μεμονωμένου αρχείου: 10MB Επιτρεπόμενοι τύποι αρχείων: Word, Excel, PPT, PDF, Εικόνα, Βίντεο, Ήχος

x_1	x_2	$y = \neg x_1 \cup x_2$
0	0	1
0	1	1
1	0	0
1	1	1

Έστω η λογική συνάρτηση του πίνακα

(α) Μπορεί η συνάρτηση αυτή να αναπαρασταθεί από απλό perceptron; Εξηγήστε.

(β) Αν στο παραπάνω ερώτημα απαντήσατε καταφατικά, σχεδιάστε ένα απλό perceptron που την αναπαριστά. Αλλιώς, σχεδιάστε ένα Multi-Layer Perceptron που την αναπαριστά.

(Μη ανώνυμη ερώτηση ①)

(5 βαθμοί)

↑ Αποστολή αρχείου

Στη συσταδοποίηση θέλουμε να επιτύχουμε:

- ☐ Τίποτα από τα υπόλοιπα
- ☐ Δεν απαντώ
- ☐ Διαχωρισμός εντός κάθε cluster και συνοχή μεταξύ διαφορετικών clusters
- ☐ Συνοχή εντός του κάθε cluster και διαχωρισμό μεταξύ διαφορετικών clusters

Έστω ότι εκπαιδεύουμε μοντέλο μηχανικής μάθησης δέκα παραμέτρων εφαρμόζοντας τεχνικές ομαλοποίησης και οι τιμές τους μετά το πέρας της εκπαίδευσης διαμορφώνονται όπως στο παρακάτω διάνυσμα:

$[-0.59342468, -2.45993074, 0.28591644, -2.17031788, 6.46296243, -0.7626098, -0.28098948, 0.71953103, 0.08476217, -5.32733193]$

Ποια από τις τεχνικές ομαλοποίησης που έχουμε εξετάσει έχει εφαρμοστεί σε αυτή την περίπτωση. Τεκμηριώστε σύντομα την απάντησή σας.

(3 βαθμοί)

Εισαγάγετε την απάντησή σας

Θεωρήστε την κλάση υποθέσεων $H = \{1/10, 2/10, \dots, 9/10\}$, και έστω x πραγματικός αριθμός στο $[0,1]$. Για παράδειγμα ένα από τα μέλη της H είναι η υπόθεση:

$$h(x) = 1 \text{ αν } x \geq 1/10$$

$$h(x) = 0 \text{ αλλιώς}$$

(α) Πόσα δείγματα εκπαίδευσης αρκούν για την αγνωστική εκμάθηση PAC αυτής της κλάσης υποθέσεων για $\epsilon = 0,01$ και $\delta = 0.05$;

(β) Τί ισχύει για τα το πλήθος των δειγμάτων εκπαίδευσης που αρκούν αν $\epsilon = 0,1$;

(γ) Τί ισχύει για τα το πλήθος των δειγμάτων εκπαίδευσης που αρκούν αν το x μπορεί να είναι οποιοσδήποτε πραγματικός αριθμός; (Μη ανώνυμη ερώτηση ①)

(5 βαθμοί)

↑ Αποστολή αρχείου

Με ποιούς τρόπους μπορούμε να εξισορροπήσουμε ένα σύνολο δεδομένων;

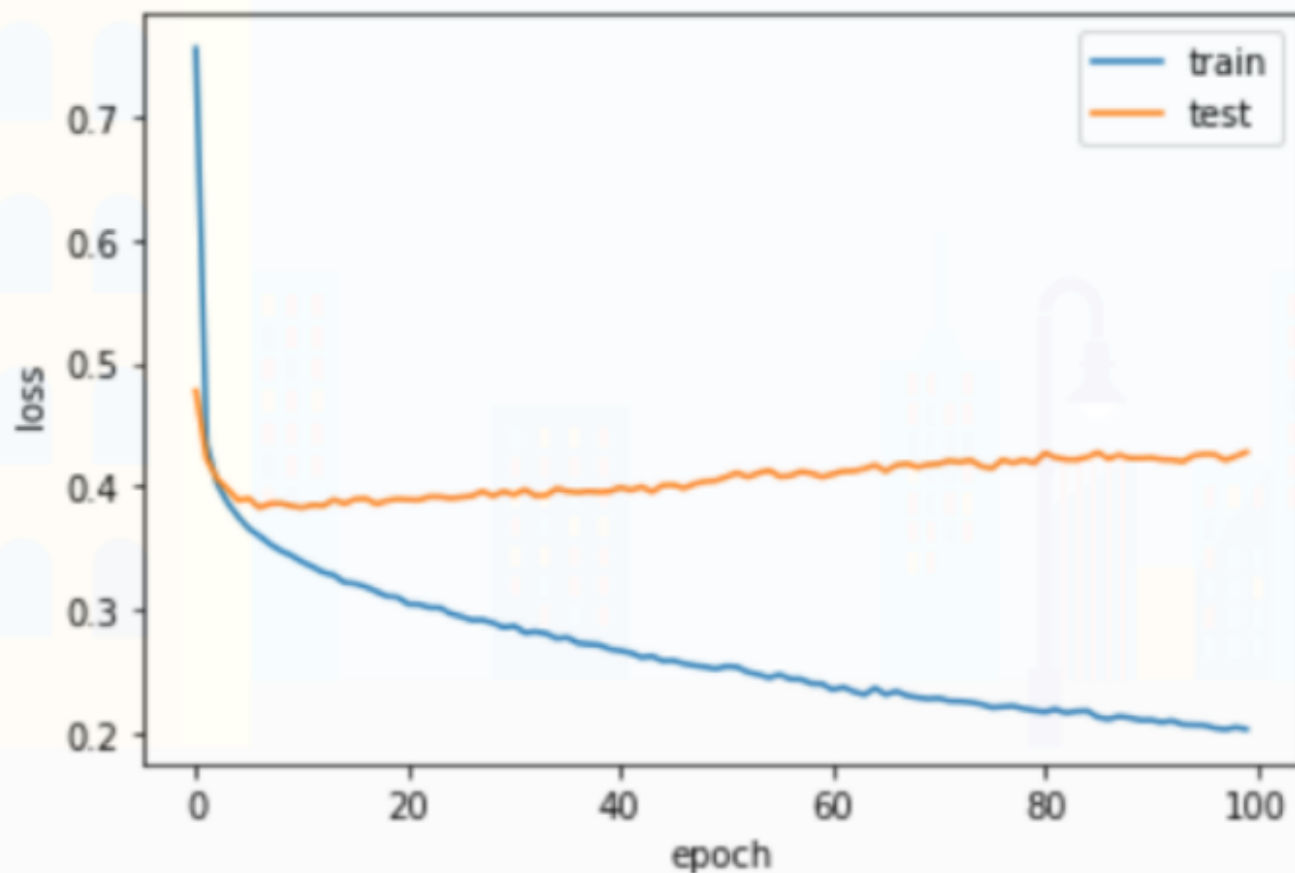
- ☐ Επιλέγοντας τυχαία δείγματα από όλες τις κατηγορίες που έχουν μεγαλύτερο πλήθος από τη μικρότερη κατηγορία
- ☐ Αφαιρώντας τυχαία δείγματα από τις λιγότερο συχνές κατηγορίες
- ☐ Επιλέγοντας, τυχαία, ορισμένα δείγματα από τις λιγότερο συχνές κατηγορίες και επαναλαμβάνοντας τα.
- ☐ Αφαιρώντας τυχαία δείγματα από όλες τις κατηγορίες που έχουν μεγαλύτερο πλήθος από τη μικρότερη κατηγορία

Έστω ότι η παρακάτω γραφική παράσταση απεικονίζει τη μεταβολή της συνάρτησης σφάλματος στα δεδομένα εκπαίδευσης (μπλε γραμμή) και ελέγχου (πορτοκαλί γραμμή) για μοντέλο μηχανικής μάθησης, το οποίο εκπαιδεύτηκε για 100 εποχές. Πως θα χαρακτηρίζατε το αποτέλεσμα της εκπαίδευσης;



(3 βαθμοί)

model loss



Εισαγάγετε την απάντησή σας