

5

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

☐ Δεν απαντώ

☐ Όχι

☒ Ναι

6

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 ⓘ)

(5 Points)

 6.pdf

 Upload file

(Non-anonymous question ⓘ)

(5 Points)

Έστω νευρωνικό δίκτυο στο οποίο οι έξοδοι του επιπέδου U τροφοδοτο στο επίπεδο V και έστω W τα βάρη των συνδέσεων μεταξύ των επιπέδ και V . Στο επίπεδο U η συνάρτηση ενεργοποίησης είναι η υπερβ εφαπτομένη (\tanh) ενώ στο επίπεδο V η σιγμοειδής. Αν u_i είναι η έξοδος $i^{\text{στού}}$ νευρώνα του επιπέδου U , v_k η έξοδος του $k^{\text{στού}}$ νευρώνα του επιπέδ $u_i=0.8$, $v_k=0.4$ 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 $\frac{\partial v_k}{\partial u_i}$



7.pdf

Upload file

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

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

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)

Βλέπουμε ότι οι περισσότερες από τις παραμέτρους είναι 0 (6 στις 10 για την ακρίβεια) που σημαίνει ότι χρησιμοποιήθηκε L1 κανονικοποίηση.

πχ. Αν έχουμε w θετικό, η παραμετρος κανονικοποίησης $\lambda > 0$ πιέζει το βάρος w να είναι λιγότερο θετικό αφαιρώντας λ από το w . Αυτό οδηγεί σε πιθανό μηδενισμό του w . Αντιστοίχως όταν έχουμε w αρνητικό πιέζεται πάλι προς το 0 το w προσθέτοντας το λ .

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).

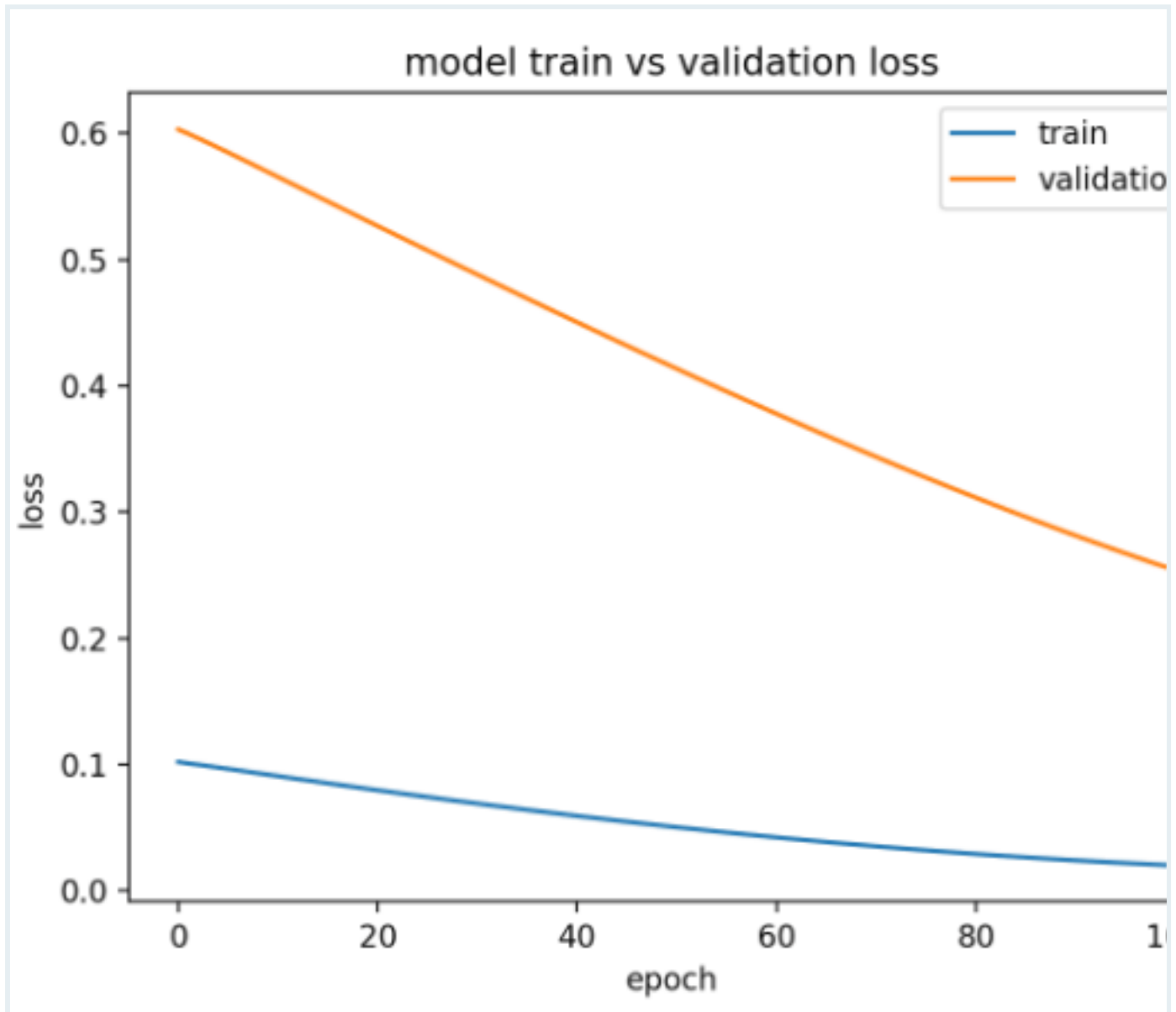
(β) Υπολογίστε τον συνολικό αριθμό των ανανεώσεων βαρών που πραγματοποιούνται στην οπισθοδιάδοση (backpropagation).

(4 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)



Το μοντέλο είναι underfitted μιας και δε γενικεύει καλά στο validation set. Στην περίπτωση αυτή, όπως φαίνεται και από το διάγραμμα, το performance μπορεί να βελτιωθεί με την αύξηση των εποχών εκπαίδευσης.

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

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



- ☐ 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.
- ☒ 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 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.
- ☒ 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.

13

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 ⓘ)

(5 Points)

14

Με ποιούς τρόπους μπορούμε να εξισορροπήσουμε ένα σύνολο δεδομένων;

- ☒ Επιλέγοντας, τυχαία, ορισμένα δείγματα από τις λιγότερο συχνές κατηγορίες και επαναλαμβάνοντας τα.
- ☐ 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

15

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)



15.pdf

↑ Upload file

☒ Send me an email receipt of my responses

Submit

This content is created by the owner of the form. The data you submit will be sent to the form owner. Microsoft is not responsible for the privacy or security practices of its customers, including those of this form owner. Never give out your password.

Powered by Microsoft Forms | [Privacy and cookies](#) | [Terms of use](#)

6

(α) Το a επιλέγεται με n τρόπους, ~~από το σύνολο $\{0, 1, \dots, n-1\}$~~ .
 Ομοίως για το b .
 Άρα τελικά
 επιλέγεται ~~από το σύνολο $\{0, 1, \dots, n-1\}$~~ $a \leq b \leq n-1$.
 άρα επιλέγεται

Όταν $a=0$: επιλέγεται με n τρόπους
 όταν $a=1$: b : $n-1$ τρόπους
 όταν $a=2$: b : $n-2$ τρόπους.

...
 όταν $a=n-2$: b : 2 τρόπους (μπορεί να έχει μόνο $n-1$)
 όταν $a=n-1$: b : 1 τρόπο ($b=a=n-1$) αφού $b \geq a$.

Άρα επιλέγουμε τα a, b με:
 $(n-1) + (n-2) + (n-3) + \dots + 2 + 1 = \frac{(n-1)(n-2)}{2} + n + (n-1) + (n-2) + 2 + 1 = \frac{n(n-1)}{2}$

Ομοίως για τα c, d , επιλέγεται ανεξάρτητα από a, b με τον ίδιο αριθμό τρόπων

Τελικά τα διαφορετικά ορθογώνια θα είναι:

$$\left[\frac{(n-1)(n-2)}{2} \right]^2 = \left(\frac{n(n-1)}{2} \right)^2 = \binom{n}{2}^2 = |H|$$

(β) $n=100$, $1-\delta=0.95 \Rightarrow \delta=0.05$, $\varepsilon=0.15$.

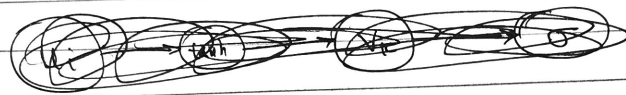
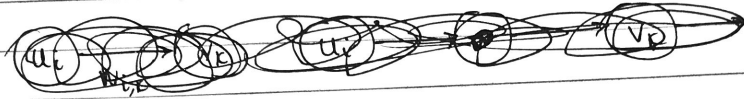
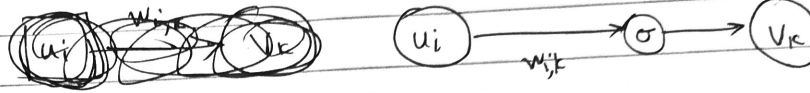
$$m_H(\varepsilon, \delta) \leq \frac{2}{\varepsilon^2} \ln \left(2 \frac{|H|}{\delta} \right) = \frac{2}{0.15^2} \ln \left(2 \frac{\binom{n}{2}^2}{0.05} \right) =$$

$$= \frac{2}{0.15^2} \cdot \ln \left(2 \frac{\left(\frac{100 \cdot 99}{2} \right)^2}{0.05} \right) = 88.88 \ln \left(2 \frac{\left(\frac{100 \cdot 99}{2} \right)^2}{0.05} \right) \approx 88.88 \cdot 20.7 =$$

$$\approx 1838.16$$

(7) $u_i = 0.8$, $v_k = 0.4$, $w_{i,k} = 0.3$.

(a) $\frac{\partial v_k}{\partial w_{i,k}}$



$$v_k = \frac{1}{1 + e^{-w_{i,k} \cdot u_i}}$$

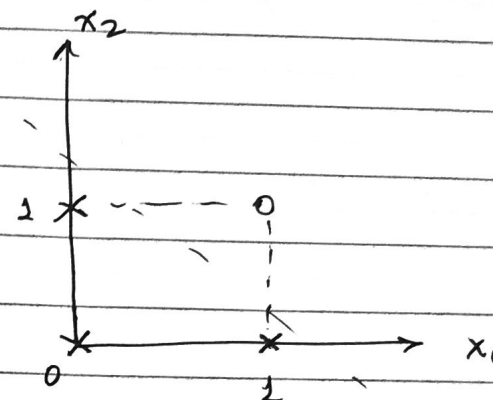
(a) $\frac{\partial v_k}{\partial u_i} = \frac{1}{1 + e^{-w_{i,k} \cdot u_i}} \cdot w_{i,k}$

$$\begin{aligned} \frac{\partial v_k}{\partial u_i} &= - \left(\frac{1}{1 + e^{-w_{i,k} \cdot u_i}} \right)^2 \frac{\partial}{\partial u_i} (e^{-w_{i,k} \cdot u_i}) = \\ &= \left(\frac{1}{1 + e^{-w_{i,k} \cdot u_i}} \right)^2 \cdot w_{i,k} = 0.4 \cdot \left(\frac{1}{1 + e^{-0.3 \cdot 0.8}} \right)^2 = \\ &= 0.4 \cdot \left(\frac{1}{1 + e^{-0.24}} \right)^2 = 0.4 \cdot 0.5695^2 = 0.4 \cdot 0.324 \approx 0.1296 \end{aligned}$$

$$\begin{aligned} \frac{\partial v_k}{\partial w_{i,k}} &= \left(\frac{1}{1 + e^{-w_{i,k} \cdot u_i}} \right)^2 \cdot u_i = 0.7 \cdot v_k^2 = 0.7 \cdot (0.5)^2 = 0.175 \\ &= 0.7 \cdot 0.25 = 0.175. \end{aligned}$$

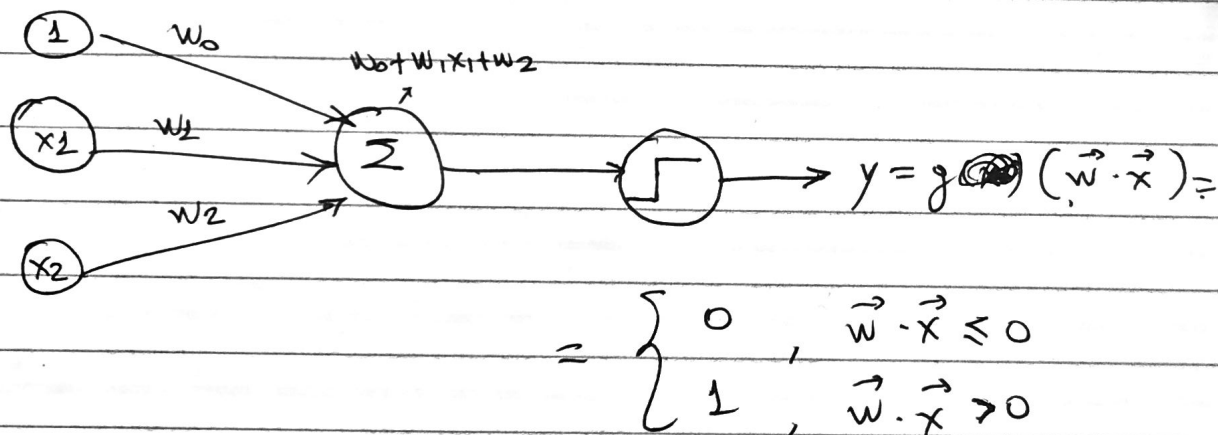
13

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



(a) Βλέπουμε ότι οι δύο κλάσεις είναι γραμμικά διαχωρίσιμες άρα μπορεί να αναπαρασταθεί η συνάρτηση από απλό perceptron.

(b)



- /4.33

15) 1. 2-D εικόνες.

Εξαρτάται το πρόβλημα.

Γενικά, είναι σημαντικό να διαλέγουμε ποια features είναι σημαντικά αν έχουμε το traditional approach.

Με το nn θα πρέπει να δούμε θέματα όπως :

- αν κλάσεις είναι unbalanced \rightarrow oversampling.
- αν οι εποχές είναι αρκετές για εκπαίδευση \rightarrow αυτήνη εποχή
- δειν ή (πχ. κρυφά layers) \rightarrow πειραματισμός με διάφορες δομές.

2. Πρόβλημα :

Backprop :

- zero gradients.
- SGD δεν μπορεί να κάνει ανανέωση βαρών γιατί $\frac{\partial l}{\partial w} = 0$.

ReLU :

- Δεν είναι διαφορίσιμη στο 0
- Νεκροί νευρώνες όταν $x \leq 0$

σφαμική :

- Έξυφος δεν είναι περιορισμένη σε κάποιο range

πρόβλημα :

- Δίνει σταδιακά αποτελέσματα οπότε τελικοί παραληψες να μην έχεις συνάρτηση βαρών.