

Θέμα 1 Regularization

Μια τεχνική για το regularization δικτύων είναι η προσθήκη θορύβου στις εισόδους του δικτύου. Θεωρούμε την απλή περίπτωση γραμμικής παλινδρόμησης και τη συνάρτηση τετραγωνικού σφάλματος, όπου w_i οι συντελεστές του μοντέλου, z η πραγματική επισημείωση (ground truth) και x το διάνυσμα εισόδου:

$$y = \sum_i w_i x_i + b$$
$$J(y, z) = (y - z)^2$$

1. Υποθέστε ότι φτιάχνουμε την είσοδο \tilde{x} για το μοντέλο γραμμικής παλινδρόμησης προσθέτοντας στις συνιστώσες x_i του διανύσματος εισόδου x Γκαουσιανό θόρυβο $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$. $\sigma_1^2 = \sigma_2^2 = \dots \sigma_i^2$. Εκφράστε την έξοδο \tilde{y} ως προς τα y, w_i, ϵ_i .
2. Δεδομένου ότι τώρα η είσοδος του μοντέλου \tilde{x} είναι μια τυχαία μεταβλητή, γράψτε τη νέα συνάρτηση κόστους ως την αναμενόμενη τιμή $\tilde{J} = E[J(\tilde{y}, z)]$ σαν μια έκφραση των J, w_i, ϵ_i .
Υπόδειξη: $E[\epsilon_i] = 0$, ϵ_i ανεξάρτητο με ϵ_j, y, z, w_i
3. Αποδείξτε ότι ο προσθετικός θόρυβος στις εισόδους αυτού του μοντέλου παλινδρόμησης είναι ισοδύναμος με το L2 regularization, εκφράζοντας το κόστος \tilde{J} ως προς τα J, w_i, σ^2 . Ποιος είναι ο συντελεστής λ του L2 regularization σε αυτή την περίπτωση;
Υπόδειξη: $Var[X] = E[X^2] - (E[X])^2$

Θέμα 2 Αναδρομικά Δίκτυα

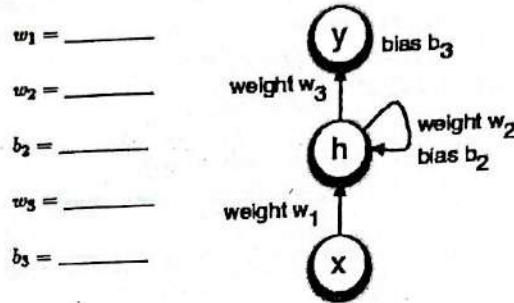
Σας δίδεται ένα αναδρομικό νευρωνικό δίκτυο (recurrent neural network) το οποίο δέχεται ως είσοδο ένα διάνυσμα x_t και ένα διάνυσμα (κρυφής) κατάστασης h_{t-1} και δίνει ως έξοδο ένα διάνυσμα (κρυφής) κατάστασης h_t και ένα διάνυσμα εξόδου y_t . Τα βάρη και τα biases των κρυφών επιπέδων είναι τα w_i και b_i αντίστοιχα.

$$h_t = f(w_1 x_t + w_2 h_{t-1} + b_2),$$

$$y_t = g(w_3 h_t + b_3),$$

όπου f και g είναι συναρτήσεις ενεργοποίησης (activation functions).

1. Το ακόλουθο διάγραμμα περιγράφει ένα ακολουθιακό νευρωνικό δίκτυο στο οποίο οι καταστάσεις x_t, h_{t-1}, h_t καθώς και y_t είναι βαθμωτές ποσότητες (αριθμοί). Σας δίνεται επιπλέον ότι η f



είναι μια δυαδική συνάρτηση κατωφλίου, ενώ η g μια γραμμική ενεργοποίηση. Συγκεκριμένα:

$$f(x) = \begin{cases} 0 & \text{για } x < 0 \\ 1 & \text{για } x \geq 0 \end{cases}$$

$$g(x) = x$$

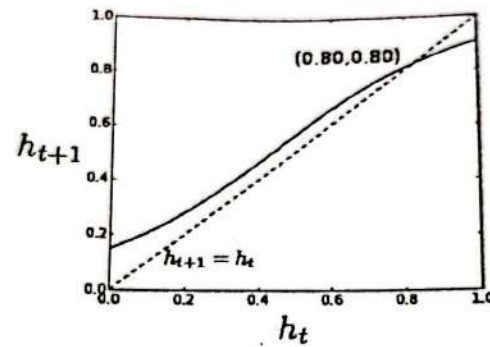
Συμπληρώστε τις τιμές w_1, w_2, b_2, w_3, b_3 ώστε το ακολουθιακό δίκτυο να λειτουργεί ως μια μηχανή κατάστασης η οποία αρχικά δίνει ως έξοδο 0 αλλά μόλις δοθεί ως είσοδος κάποιο 1, η έξοδος αλλάζει σε 1 για όλα τα υπόλοιπα βήματα. Για παράδειγμα εάν η είσοδος είναι 001010 τότε η αντίστοιχη έξοδος θα πρέπει να είναι 001111. Υποθέστε επιπλέον ότι η αρχική κρυφή κατάσταση εκκινεί από την τιμή 0.

Υπόδειξη: Σε μια πιθανή λύση, η κρυφή κατάσταση έχει τιμή ενεργοποίησης $h_t = 0$ έως ότου εμφανιστεί μια είσοδος $x_t = 1$. Από αυτό το σημείο και μετά η κρυφή κατάσταση αλληλίζει και "κλειδώνει" στην τιμή 1. Η έξοδος του δικτύου δίνει πάντοτε την ίδια τιμή με αυτή της κρυφής κατάστασης $y_t = h_t$.

2. Τώρα υποθέστε ότι η συνάρτηση ενεργοποίησης f είναι η σιγμοειδής και επιπλέον υποθέστε πως η είσοδος είναι συνέχεια 0. Το ακόλουθο διάγραμμα περιγράφει πώς μεταβάλλεται η επόμενη κρυφή κατάσταση (h_{t+1}) συναρτήσει της προηγούμενης h_t για τιμές $w_2 = 3$ και $b_2 = -1$, δηλαδή $h_{t+1} = \sigma(3h_t - 1)$.

Υπόδειξη: $\sigma(x) = \frac{1}{1 + e^{-x}}$

- (a) Η τιμή h_{t+1} είναι μεγαλύτερη ή μικρότερη σε σχέση με την τιμή h_t , όταν $h_t = 0.5$?
- (b) Η τιμή h_{t+1} είναι μεγαλύτερη ή μικρότερη σε σχέση με την τιμή h_t , όταν $h_t = 0.9$?
- (c) Συγκλίνει (και αν ναι πού) η τιμή h_{t+1} εάν παίρνουμε σταδιακά όλο και μεγαλύτερες ακολουθίες εισόδου ως $t \rightarrow \infty$?
- (d) Για ποιο πεδίο τιμών αρχικής κρυφής κατάστασης h_0 παρουσιάζει το δίκτυο το φαινόμενο exploding gradients για μεγάλα μήκη ακολουθιών?
- (e) Για ποιο πεδίο τιμών αρχικής κρυφής κατάστασης h_0 παρουσιάζει το δίκτυο το φαινόμενο vanishing gradients για μεγάλα μήκη ακολουθιών?



3. Μιά λύση που προτείνεται στη βιβλιογραφία ώστε να μην παρατηρούμε φαινόμενα exploding/vanishing gradients είναι να αντικαταστήσουμε τη συνάρτηση ενεργοποίησης (τη σιγμοειδή εν προκειμένω), με τη λεγόμενη ReLU. Θα βοηθούσε στην περίπτωση μας να τη χρησιμοποιήσουμε? Εξηγήστε. Υπόδειξη: $ReLU(x) = \max\{0, x\}$

Θέμα 3 HMM

Ο καιρός είναι δυνατό να έχει μία από τις εξής τρεις καταστάσεις: ηλιοφάνεια (S) ή βροχή (R) ή ομίχλη (F). Ο πίνακας μεταβάσεων για τον καιρό μεταξύ δύο διαδοχικών ημερών είναι:

	S	R	F
S	0.70	0.10	0.20
R	0.10	0.70	0.20
F	0.30	0.20	0.50

(Π.χ., η πιθανότητα μετάβασης από την κατάσταση S στην κατάσταση F είναι 0.20.) Υποθέστε πως είστε κλεισμένοι στο εργαστήριο. Η μόνη ένδειξη που μπορείτε να έχετε σχετικά με την κατάσταση του καιρού που επικρατεί βασίζεται στην παρατήρηση αν ένας συγκεκριμένος συνάδελφός σας φέρει ομπρέλα ή όχι, κάθε μέρα που έρχεται στο εργαστήριο. Η πιθανότητα ο συνάδελφός σας να φέρει ομπρέλα δοθείσης της κατάστασης του καιρού την εκάστοτε μέρα είναι:

S	0.10
R	0.80
F	0.30

(Π.χ., η πιθανότητα να φέρει ομπρέλα όταν επικρατεί ηλιοφάνεια είναι 0.10.)

Απαντήστε στα ακόλουθα ερωτήματα χρησιμοποιώντας κρυφό μοντέλο Markov πρώτης τάξεως (HMM-1) όπου χρειάζεται.

1) όπου χρειάζεται.

(α) Υποθέστε πως σήμερα επικρατεί ηλιοφάνεια. Ποιά η πιθανότητα αύριο να έχει ήλιο και μεθαύριο να βρέχει;

(β) Υποθέστε πως σήμερα επικρατεί ομίχλη. Ποιά η πιθανότητα μεθαύριο ο καιρός να είναι βροχερός;

(γ) Υποθέστε πως χτες που κλειστήκατε στο εργαστήριο επικρατούσε ηλιοφάνεια. Έστω ότι σήμερα συνάδελφός σας έφτασε στο εργαστήριο κουβαλώντας ομπρέλα. Υποθέτουμε ότι η prior πιθανότητα ο συνάδελφός σας να κουβαλάει ομπρέλα μια οποιαδήποτε μέρα είναι 0.50, ποιά η πιθανότητα σήμερα να βρέχει;

(δ) Υποθέστε πως την ημέρα που κλειστήκατε στο εργαστήριο (Ημέρα 1) επικρατούσε ηλιοφάνεια. Την Ημέρα 2 ο συνάδελφός σας έφερε ομπρέλα, πράγμα που δεν έκανε την Ημέρα 3. Υποθέτοντας ότι η prior πιθανότητα ο συνάδελφός σας να κουβαλάει ομπρέλα μια οποιαδήποτε μέρα είναι 0.50, υπολογίστε την πιθανότητα να επικρατεί ομίχλη την Ημέρα 3.