

Ασκήσεις μελέτης της 14^{ης} διάλεξης

14.1. Χρησιμοποιούμε το νευρωνικό δίκτυο κυλιόμενου παραθύρου της διαφάνειας 15, για να αναγνωρίζουμε ονόματα προσώπων, οργανισμών και τοποθεσιών. Χρησιμοποιούμε 7 ετικέτες (κατηγορίες) B-I-O, δηλαδή BPerson (beginning of person name, π.χ. «Δημήτριος» στο «Δημήτριος Παπαδόπουλος»), IPerson (inside person name, π.χ. «Παπαδόπουλος» στο «Δημήτριος Παπαδόπουλος»), BOrganization (πρώτη λέξη ονόματος οργανισμού), IOrganization (μη αρχική λέξη ονόματος οργανισμού), BLocation, ILocation και O (other, λέξη που δεν ανήκει σε καμία από τις άλλες κατηγορίες). Το κυλιόμενο παράθυρο καλύπτει 3 λέξεις, την τρέχουσα, την προηγούμενη και την επόμενη. Το μέγεθος του λεξιλογίου είναι $|V| = 100.000$. Κάθε ένθεση λέξης (word embedding) είναι ένα διάνυσμα 300 διαστάσεων. Ο κόμβος “+” συνενώνει (concatenates) τις ενθέσεις των τριών λέξεων του παραθύρου. Το κρυφό επίπεδο αποτελείται από 500 νευρώνες με συνάρτηση ενεργοποίησης \tanh . Ο πίνακας $W^{(1)}$ περιέχει τα βάρη των νευρώνων του κρυφού επιπέδου, ενώ ο πίνακας $W^{(2)}$ τα βάρη των νευρώνων του επιπέδου εξόδου. Ποιες είναι οι διαστάσεις των $E, \vec{e}, W^{(1)}, \vec{\delta}^{(1)}, W^{(2)}, \vec{\delta}^{(2)}$; Αιτιολογήστε πλήρως τις απαντήσεις σας.

Απάντηση: Ο πίνακας E περιέχει (ως στήλες) τις ενθέσεις των 100.000 λέξεων του λεξιλογίου. Κάθε ένθεση είναι διάνυσμα (στήλη) 300 διαστάσεων. Άρα ο E έχει διαστάσεις 300×100.000 .

Το διάνυσμα \vec{e} είναι η συνένωση (concatenation) τριών ενθέσεων λέξεων (τρέχουσα, προηγούμενη, επόμενη), κάθε μία από τις οποίες είναι ένα διάνυσμα-στήλη 300×1 , επομένως το \vec{e} είναι διαστάσεων 900×1 .

Ο πίνακας $W^{(1)}$ έχει διαστάσεις 500×900 , ώστε ο πολλαπλασιασμός $W^{(1)}\vec{e}$ να παράγει διάνυσμα 500×1 , το οποίο περιέχει τις τιμές των 500 νευρώνων του κρυφού επιπέδου πριν τη συνάρτηση ενεργοποίησης \tanh . Μετά την εφαρμογή της \tanh , το διάνυσμα αυτό γίνεται το $\vec{\delta}^{(1)}$ του σχήματος, δηλαδή περιέχει τις εξόδους των νευρώνων του κρυφού επιπέδου (μετά και την εφαρμογή της συνάρτησης ενεργοποίησης σε κάθε νευρώνα), επομένως και το $\vec{\delta}^{(1)}$ είναι διαστάσεων 500×1 .

Ο πίνακας $W^{(2)}$ έχει διαστάσεις 7×500 , ώστε ο πολλαπλασιασμός $W^{(2)}\vec{\delta}^{(1)}$ να παράγει διάνυσμα 7×1 , το οποίο περιέχει τις τιμές των 7 νευρώνων του επιπέδου εξόδου πριν τη συνάρτηση ενεργοποίησης softmax . Μετά την εφαρμογή της softmax , το διάνυσμα αυτό γίνεται το $\vec{\delta}^{(2)}$ του σχήματος, δηλαδή περιέχει τις τιμές των νευρώνων του επιπέδου εξόδου, δηλαδή 7 πιθανότητες, μία για κάθε μία δυνατή ετικέτα (κατηγορία) της τρέχουσας λέξης. Επομένως το $\vec{\delta}^{(2)}$ είναι διαστάσεων 7×1 .

Αν δεν έχετε ολοκληρώσει τη μελέτη των ασκήσεων των προηγούμενων ενοτήτων, είναι ευκαιρία να το κάνετε τώρα...