



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Επεξεργασία Φωνής & Φυσικής Γλώσσας

Ροή Σ: Σήματα, Έλεγχος, Ρομποτική

7^ο Εξάμηνο

Προπαρασκευή 3^{ης} Εργαστηριακής Άσκησης

ΘΕΜΑ:

RNNs/LSTMs for Sentiment Analysis

Χρήστος Δημόπουλος

03117037

Ζητούμενο 1: Συμπληρώστε τα κενά στη θέση `main.py:EX1` και τυπώστε τα πρώτα 10 labels από τα δεδομένα εκπαίδευσης και τις αντιστοιχίες τους σε αριθμούς.

Για το dataset “MR” λαμβάνουμε τα εξής labels από τα δεδομένα εκπαίδευσης:

First 10 Labels of Train Set:

1
1
1
1
1
1
1
1
1
1
1

Ενώ το matching των αριθμών αυτών είναι:

Labels - Numbers Matching:

0 - negative
1 - positive

Αντιστοίχως για το dataset “Semeval2017A”:

First 10 Labels of Train Set:

1
1
0
1
1
2
2
1
2
2

Labels - Numbers Matching:

0 - negative
1 - neutral
2 - positive

Ζητούμενο 2: Συμπληρώστε τα κενά στη θέση `dataloading.py:EX2` και τυπώστε τα πρώτα 10 παραδείγματα από τα δεδομένα εκπαίδευσης. (Tokenization)

Για το dataset “MR” λαμβάνουμε:

['the', 'rock', 'is', 'destined', 'to', 'be', 'the', 'st', 'century', 's', 'new', 'conan', 'and', 'that', 'he', 'is', 'going', 'to', 'make', 'a', 'splash', 'even', 'greater', 'than', 'arnold', 'schwarzenegger', 'jean', 'claud', 'van', 'damme', 'or', 'steven', 'segal']
 ['the', 'gorgeously', 'elaborate', 'continuation', 'of', 'the', 'lord', 'of', 'the', 'rings', 'trilogy', 'is', 'so', 'huge', 'that', 'a', 'column', 'of', 'words', 'cannot', 'adequately', 'describe', 'co', 'writer', 'director', 'peter', 'jackson', 's', 'expanded', 'vision', 'of', 'j', 'are', 'are', 'tolkien', 's', 'middle', 'earth']
 ['effective', 'but', 'too', 'tepid', 'biopic']
 ['if', 'you', 'sometimes', 'like', 'to', 'go', 'to', 'the', 'movies', 'to', 'have', 'fun', 'wasabi', 'is', 'a', 'good', 'place', 'to', 'start']
 ['emerges', 'as', 'something', 'rare', 'an', 'issue', 'movie', 'that', 'is', 'so', 'honest', 'and', 'keenly', 'observed', 'that', 'it', 'does', 'not', 'feel', 'like', 'one']
 ['the', 'film', 'provides', 'some', 'great', 'insight', 'into', 'the', 'neurotic', 'mindset', 'of', 'all', 'comics', 'even', 'those', 'who', 'have', 'reached', 'the', 'absolute', 'top', 'of', 'the', 'game']
 ['offers', 'that', 'rare', 'combination', 'of', 'entertainment', 'and', 'education']
 ['perhaps', 'no', 'picture', 'ever', 'made', 'has', 'more', 'literally', 'showed', 'that', 'the', 'road', 'to', 'hell', 'is', 'paved', 'with', 'good', 'intentions']
 ['steers', 'turns', 'in', 'a', 'snappy', 'screenplay', 'that', 'curls', 'at', 'the', 'edges', 'it', 'is', 'so', 'clever', 'you', 'want', 'to', 'hate', 'it', 'but', 'he', 'somehow', 'pulls', 'it', 'off']
 ['take', 'care', 'of', 'my', 'cat', 'offers', 'a', 'refreshingly', 'different', 'slice', 'of', 'asian', 'cinema']

Αντιστοίχως για το dataset “Semeval2017A”:

['won', 'the', 'match', 'getin', 'plus', 'u', 'c', 'tomorrow', 'is', 'a', 'very', 'busy', 'day', 'u', 'c', 'with', 'awareness', 'day', 'u', 's', 'and', 'debates', 'gulp', 'debates']
 ['some', 'areas', 'of', 'new', 'england', 'could', 'see', 'the', 'first', 'flakes', 'of', 'the', 'season', 'tuesday']
 ['francesco', 'con', 'nd', 'worst', 'qb', 'definitely', 'tony', 'romo', 'the', 'man', 'who', 'likes', 'to', 'share', 'the', 'ball', 'with', 'everyone', 'including', 'the', 'other', 'team']
 ['thailand', 'washington', 'us', 'president', 'barack', 'obama', 'vowed', 'wednesday', 'as', 'he', 'visited', 'storm', 'ravaged', 'new', 'jersey', 'shore', 'to', 'http', 't', 'co', 'xzl', 'lfhs']
 ['did', 'y', 'u', 'all', 'hear', 'what', 'tony', 'romo', 'dressed', 'up', 'as', 'for', 'halloween', 'a', 'giants', 'quarterback', 'because', 'that', 'u', 's', 'all', 'he', 'could', 'throw', 'to', 'sunday', 'night']
 ['tim', 'tebow', 'may', 'be', 'available', 'wow', 'jerry', 'u', 'c', 'what', 'the', 'heck', 'you', 'waiting', 'for', 'http', 't', 'co', 'a', 'z', 'fbl']
 ['mariakaykay', 'aga', 'tayo', 'tomorrow', 'ah', 'good', 'night', 'u', 'c', 'ces', 'love', 'you', 'd']
 ['tina', 'fey', 'amy', 'poehler', 'are', 'hosting', 'the', 'golden', 'globe', 'awards', 'on', 'january', 'what', 'do', 'you', 'think']
 ['lunch', 'from', 'my', 'new', 'lil', 'spot', 'the', 'cotton', 'bowl', 'pretty', 'good', 'st', 'time', 'will', 'be', 'going', 'back', 'http', 't', 'co', 'dbbj', 'xlz']
 ['snc', 'halloween', 'pr', 'pumped', 'let', 'u', 's', 'work', 'it', 'for', 'sunday', 'packers', 'vs', 'who', 'knows', 'or', 'caresn', 'snc', 'cheerpracticeonhalloween']

Ζητούμενο 3: Υλοποιήστε τη μέθοδο `getitem` της κλάσης `SentenceDataset` (θέση `dataloading.py:EX3`) και τυπώστε 5 παραδείγματα στην αρχική τους μορφή και όπως τα επιστρέφει η κλάση `SentenceDataset`.

Για το dataset “**MR**” λαμβάνουμε:

['the', 'film', 'provides', 'some', 'great', 'insight', 'into', 'the', 'neurotic', 'mindset', 'of', 'all', 'comics', '--', 'even', 'those', 'who', 'have', 'reached', 'the', 'absolute', 'top', 'of', 'the', 'game', '.']

Code Form: [14, 1252, 20611, 192, 342, 6068, 490, 14, 196607, 24682, 40, 76, 17971, 443, 240, 559, 128, 65, 6395, 14, 6612, 608]

Label: 1

Sentence Original Length: 22

['while', 'the', 'isle', 'is', 'both', 'preposterous', 'and', 'thoroughly', 'misogynistic', ',', 'its', 'vistas', 'are', 'incredibly', 'beautiful', 'to', 'look', 'at', '.']

Code Form: [655, 14, 37613, 33, 1013, 209015, 27, 34866, 225649, 5, 222, 39878, 71, 12811, 533, 17, 274, 67, 2, 0, 0, 0]

Label: 1

Sentence Original Length: 19

['in', 'a', 'normal', 'screen', 'process', ',', 'these', 'bromides', 'would', 'be', 'barely', 'enough', 'to', 'sustain', 'an', 'interstitial', 'program', 'on', 'the', 'discovery', 'channel', '.', 'but', 'in', 'imax', '3-d', ',', 'the', 'clichés', 'disappear', 'into', 'the', 'vertiginous', 'perspectives', 'opened', 'up', 'by', 'the', 'photography', '.']

Code Form: [36, 12, 1326, 3417, 7015, 5, 376, 1193515, 197, 57, 5405, 799, 17, 70213, 173, 768840, 4542, 47, 14, 16731, 2813, 2]

Label: 1

Sentence Original Length: 22

['if', 'it', 'is', 'possible', 'for', 'a', 'sequel', 'to', 'outshine', 'the', 'original', ',', 'then', 'sl2', 'does', 'just', 'that', '.']

Code Form: [75, 34, 33, 2437, 38, 12, 25645, 17, 155980, 14, 2898, 5, 241, 1193515, 291, 60, 46, 2, 0, 0, 0, 0]

Label: 1

Sentence Original Length: 18

['an', 'exciting', 'and', 'involving', 'rock', 'music', 'doc', ',', 'a', 'smart', 'and', 'satisfying', 'look', 'inside', 'that', 'tumultuous', 'world', '.']

Code Form: [173, 5026, 27, 34451, 1287, 435, 10520, 5, 12, 2154, 27, 41593, 274, 1734, 46, 351988, 368, 2, 0, 0, 0, 0]

Label: 1

Sentence Original Length: 18

Αντιστοίχως για το dataset “**Semeval2017A**”:

['tim', 'tebow', 'may', 'be', 'available', '!', 'wow', 'jerry', '\\u002c', 'what', 'the', 'heck', 'you', 'waiting', 'for', '!', 'http://t.co/a7z9fbl4']

Code Form: [2919, 17871, 531, 57, 652239, 10, 611, 6390, 1193515, 87, 14, 7060, 16, 943, 38, 10, 1193515, 0, 0, 0, 0, 0]

Label: 2

Sentence Original Length: 17

['@dannyb618', 'sure', 'absolutely--', 'i', 'meant', 'out', 'of', 'the', 'bachmann\\u002c', 'perry\\u002c', 'santorum\\u002c', 'herman', 'cain', 'bunch', 'this', 'election.', 'and', 'romney', 'was', 'not', 'my', '1st', 'choice']

Code Form: [1193515, 539, 1193515, 11, 2012, 100, 40, 14, 1193515, 1193515, 1193515, 35359, 39078, 4372, 54, 1193515, 27, 3134, 94, 79, 30, 1193515]

Label: 0

Sentence Original Length: 22

['@h0tlikepayne:', 'it\\u2019s', '#confirmed', 'that', 'you', 'can', 'listen', 'to', 'the', 'deluxe', 'version', 'of', 'tmh', 'on', 'itunes', '9pm', 'gmt', 'on', 'nov', '5th\\u002c', 'monday.']

Code Form: [1193515, 1193515, 1193515, 46, 16, 103, 1058, 17, 14, 14298, 2791, 40, 21272, 47, 3762, 1193515, 18778, 47, 6997, 1193515, 1193515, 0]

Label: 2

Sentence Original Length: 21

['http://t.co/hzorig6w', 'its', 'going', 'down', 'in', '#deathvalley', 'this', 'saturday!', 'geaux', 'tigers', '@lsufball', '@jacobhester22', '@lsucoachmiles']

Code Form: [1193515, 222, 212, 314, 36, 1193515, 54, 1193515, 97160, 10726, 1193515, 1193515, 1193515, 0, 0, 0, 0, 0, 0, 0, 0, 0]

Label: 2

Sentence Original Length: 13

['3', 'days', 'of', 'halloween', 'starts', 'tomorrow:', 'day', '1', '-', 'david', 'bowie\\u002c', 'day', '2', '-', 'poker', 'face\\u002c', 'day', '3', '-', 'skeleton', '#holla']

Code Form: [1193515, 456, 40, 4087, 2041, 1193515, 126, 1193515, 29, 2007, 1193515, 126, 1193515, 29, 10438, 1193515, 126, 1193515, 29, 44718, 1193515, 0]

Label: 1

Sentence Original Length: 21

Ζητούμενο 4: Συμπληρώστε τα κενά στις θέσεις models.py:EX4 και απαντήστε στα παρακάτω ερωτήματα.

- Γιατί αρχικοποιούμε το embedding layer με τα προ-εκπαιδευμένα word embeddings?

Αρχικοποιώντας το embedding layer με τα προ-εκπαιδευμένα word embeddings, χρησιμοποιούμε προϋπάρχουσα πληροφορία σχετικά με τις σχέσεις των λέξεων, με αποτέλεσμα **να επιταχύνουμε τη διαδικασία της μάθησης** του νευρωνικού δικτύου στο στάδιο του training και εν τέλει να λάβουμε καλύτερα αποτελέσματα ακρίβειας. Τα προ-εκπαιδευμένα word embeddings, άλλωστε έχουν εκπαιδευτεί πάνω σε **τεράστια text corpora** και έχουν μάθει εύχρηστα **word associations** ανάμεσα στις λέξεις.

- Γιατί κρατάμε παγωμένα τα βάρη του embedding layer κατά την εκπαίδευση?

Επιλέγοντας να εκπαιδεύσουμε το προεκπαιδευμένο embedding layer, εμπεριέχεται ο κίνδυνος να αλλοιώσουμε τα word associations που έχουν ήδη διαμορφωθεί. Αυτό συμβαίνει διότι το text corpus που χρησιμοποιείται για την εκπαίδευση του Νευρωνικού στη συνέχεια είναι πολύ μικρότερο από αυτό που χρησιμοποιήθηκε για τα προ-εκπαιδευμένα word embeddings, με αποτέλεσμα να ενημερώνονται λέξεις που συναντάμε και στα 2 corpora, αλλά όχι οι υπόλοιπες. Ως εκ τούτου, **παραμορφώνονται οι λεκτικές σχέσεις** που είχαν διαμορφωθεί και γι'αυτό επιλέγουμε να κρατάμε frozen την εκ νέου εκπαίδευση του embedding layer.

Ζητούμενο 5: Συμπληρώστε τα κενά στις θέσεις models.py:EX5 και απαντήστε στα παρακάτω ερωτήματα.

- Γιατί βάζουμε μία μη γραμμική συνάρτηση ενεργοποίησης στο προτελευταίο layer? Τι διαφορά θα είχε αν είχαμε 2 ή περισσότερους γραμμικούς μετασχηματισμούς στη σειρά?

Σύμφωνα με το **Universal Approximation Theorem**, κάτω από συγκεκριμένες συνθήκες, κάθε συνεχής συνάρτηση f μπορεί να μοντελοποιηθεί από ένα νευρωνικό δίκτυο, το οποίο αποτελείται από ένα hidden Layer και έναν ικανοποιητικό αριθμό από units. Βασική προϋπόθεση, ωστόσο, είναι **η συνάρτηση ενεργοποίησης να εμπεριέχει κάποια μορφή μη γραμμικότητας**. Με άλλα όγια, προκειμένου να προσεγγίσουμε ένα μη γραμμικό πρόβλημα, πρέπει να υπάρχει ένα είδος μη γραμμικότητας μέσα στο δίκτυο.

Μολονότι η συνάρτηση ReLU είναι τμηματικά γραμμική (στα διαστήματα $(-\infty, 0)$ και $(0, \infty)$), συνολικά είναι μια μη γραμμική συνάρτηση και έτσι ικανοποιεί την προϋπόθεση του παραπάνω θεωρήματος. Ωστόσο, δεν θα μπορούσε να αντικατασταθεί από 2 γραμμικούς μετασχηματισμούς στη σειρά, καθώς κάτι τέτοιο δεν θα εισήγαγε την επιθυμητή μη γραμμικότητα στο δίκτυο:

Έστω ότι έχουμε το διάνυσμα χαρακτηριστικών x_0 και το διάνυσμα βαρών w_1 . Περνώντας από ένα πρώτο γραμμικό layer του Νευρωνικού Δικτύου λαμβάνουμε:

$$y_1 = w_1^T x_0$$

Στη συνέχεια, περνώντας από ένα δεύτερο διαδοχικό γραμμικό layer λαμβάνουμε:

$$y_2 = w_2^T y_1$$

Συνολικά, συνδυάζοντας τις δύο παραπάνω σχέσεις, έχουμε:

$$y_2 = w_2 w_1 x_0 = W x_0$$

Ως εκ τούτου, παρατηρούμε ότι **παραμένει μια γραμμική συσχέτιση** μεταξύ εισόδου και εξόδου. Δεδομένου ότι η συνάρτηση που θέλουμε να μοντελοποιήσουμε είναι μη γραμμική, κάτι τέτοιο δεν είναι επιθυμητό. Γι'αυτό τον λόγο, καταφεύγουμε στη μη γραμμική συνάρτηση ενεργοποίησης ReLU.

Ζητούμενο 6: Συμπληρώστε τα κενά στις θέσεις `models.py:EX6` στη μέθοδο `forward` και απαντήστε στα παρακάτω ερωτήματα

- Αν θεωρήσουμε ότι κάθε διάσταση του embedding χώρου αντιστοιχεί σε μία αφηρημένη έννοια, μπορείτε να δώσετε μία διαισθητική ερμηνεία για το τι περιγράφει η αναπαράσταση που φτιάξατε (κέντρο-βάρους).

Αν θεωρήσουμε ότι κάθε διάσταση του embedding χώρου αντιστοιχεί σε μια αφηρημένη έννοια, τότε κάθε λέξη ως σημείο του χώρου ποσοτικοποιεί το πόσο "εκφράζει" καθε μια από αυτές τις έννοιες. Με την αναπαράσταση που χρησιμοποιούμε, η κάθε πρόταση θα συνιστά ένα centroid (κέντρο βάρους) των λέξεων που την απαρτίζουν, με αποτέλεσμα να εκφράζει τον μέσο όρο των ποσοτήτων αφηρημένων έννοιων του χώρου που εκφράζουν οι λέξεις της.

- Αναφέρετε πιθανές αδυναμίες της συγκεκριμένης προσέγγισης για να αναπαραστήσουμε κείμενα.

Βασικό μειονέκτημα της εν λόγω αναπαράστασης είναι το γεγονός ότι δεν λαμβάνονται υπόψη οι συσχετίσεις των λέξεων που ανήκουν στην ίδια πρόταση, ούτε η σειρά με την οποία αυτές εμφανίζονται. Για παράδειγμα, οι προτάσεις "You came here to teach not to shout" και "You came here to shout not to teach", μολονότι έχουν τελείως διαφορετικό νοήμα, απαρτίζονται από τις ίδιες ακριβώς λέξεις, με αποτέλεσμα στην αναπαράσταση που διαλέξαμε να αντιμετωπίζονται ως ίδιες. Επομένως, η αναπαράσταση αυτή δεν καλύπτει πλήρως το νοηματικό επίπεδο των προτάσεων.

Ζητούμενο 7: Συμπληρώστε τα κενά στις θέσεις `models.py:EX6` στη μέθοδο `forward` και απαντήστε στα παρακάτω ερωτήματα.

- Τι συνέπειες έχουν τα μικρά και μεγάλα `mini-batches` στην εκπαίδευση των μοντέλων;

Κατά κύριο λόγο, τα **mini-batches μικρού μεγέθους** χρησιμοποιούνται διότι:

1. Τα μικρότερα `batches` είναι θορυβώδη (noisy), παρέχοντας ένα regularization effect και κατ'επέκταση μικρότερο σφάλμα γενίκευσης.
2. Τα μικρότερα `mini-batches` χωράνε πιο εύκολα στη μνήμη (όταν χρησιμοποιούμε GPU) επιταχύνοντας τη διαδικασία εκπαίδευσης.
3. Στα αρνητικά της υπόθεσης, όταν το batch size είναι υπερβολικά μικρό ελλοχεύει ο κίνδυνος το μοντέλο να κάνει update με «ακανόνιστο» τρόπο τα βάρη του (όπως στην περίπτωση που βλέπει μόνο 2 τυχαία instances πριν κάθε ενημέρωση).

Όσον αφορά τα **batches μεγάλου μεγέθους**, ισχύουν τα εξής:

1. Τα πολύ μεγάλα `batches` οδηγούν το μοντέλο σε χαμηλότερο ασυμπτωτικό test accuracy, κάτι το οποίο εν μέρει μπορεί να αποτραπεί αυξάνοντας το learning rate.
2. Στα θετικά της υπόθεσης, ξεκινώντας με μεγάλα `batches` αποτρέπουμε το μοντέλο από το να εγκλωβιστεί σε τοπικά ελάχιστα της συνάρτησης σφάλματος κατά την εκπαίδευση του. Αυτό συμβαίνει διότι, μεγαλύτερα μεγέθη batch συνεπάγονται μεγαλύτερα βήματα του gradient από ότι μικρότερα μεγέθη batch, για τον ίδιο αριθμό δειγμάτων που βλέπει το μοντέλο.
3. Με μεγάλα `batches` το μοντέλο κάνει είτε πολύ μεγάλα είτε πολύ μικρά βήματα για το gradient update. Αντιθέτως, μικρά `batches` κάνουν updates περίπου ιδίου μεγέθους.
4. Η κατανομή του gradient για μεγάλα batch sizes έχει heavier tail.

- Συνήθως ανακατεύουμε την σειρά των `mini-batches` στα δεδομένα εκπαίδευσης σε κάθε εποχή. Μπορείτε να εξηγήσετε γιατί;

Ανακατεύοντας τη σειρά με την οποία εμφανίζονται τα `mini-batches` στα δεδομένα εκπαίδευσης σε κάθε εποχή, επιτυγχάνουμε τα εξής:

1. Η διαδικασία της εκπαίδευσης συγκλίνει πιο γρήγορα.
2. Αποτρέπουμε το bias κατά την εκπαίδευση του μοντέλου.
3. Αποτρέπουμε το μοντέλο από το να μάθει μια συγκεκριμένη σειρά δεδομένων και να εκπαιδευτεί αποκλειστικά πάνω σε αυτήν.

Επιπλέον, γνωρίζοντας ότι ο στόχος της διαδικασίας εκπαίδευσης είναι να βρεθούν οι τιμές των βαρών W για τις οποίες ελαχιστοποιείται η συνάρτηση σφάλματος (η οποία στις περισσότερες περιπτώσεις δεν είναι κυρτή), ανακατεύοντας τη σειρά των `mini-batches` αποτρέπουμε το μοντέλο από το να εγκλωβιστεί σε local minimum της loss function. Αν αντί αυτού τα `batches` έρχονταν διαρκώς με την ίδια σειρά το gradient update δεν θα κατάφερνε να απεγκλωβιστεί από κάποιο τοπικό ελάχιστο που ίσως έχει εγκλωβιστεί. Αντιθέτως, ανακατεύοντας τη σειρά ωθούμε το μοντέλο στο να κάνει συχνά μεγαλύτερα βήματα στο gradient update, αυξάνοντας τις πιθανότητες να καταλήξει σε κάποιο καλύτερο ελάχιστο της συνάρτησης σφάλματος.

Ζητούμενο 10: Για κάθε ένα από τα 2 datasets που σας παρέχονται αναφέρετε τις επιδόσεις του μοντέλου στις μετρικές: accuracy, F1 score (macro average), recall (macro average). Επίσης, δημιουργήστε γραφικές παραστάσεις, στις οποίες θα φαίνονται οι καμπύλες εκπαίδευσης του μοντέλου (training και test loss) ανά εποχή.

- *Για το Dataset 'MR':*

Έχοντας επιλέξει **EPOCHS = 100** και **BATCH_SIZE = 36**, και χρησιμοποιώντας τα glove.6B.100d.txt embeddings και Adam optimizer με learning rate = 0.01, στο πέρας της εκπαίδευσης λαμβάνουμε τα ακόλουθα αποτελέσματα:

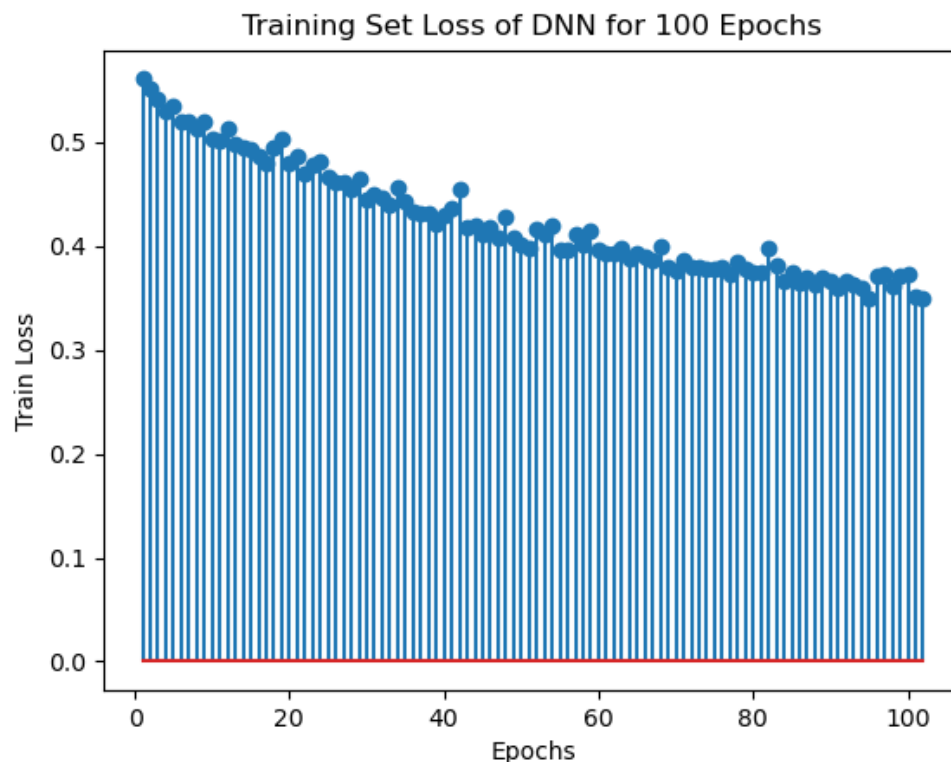
TRAIN DATASET EVALUATION:

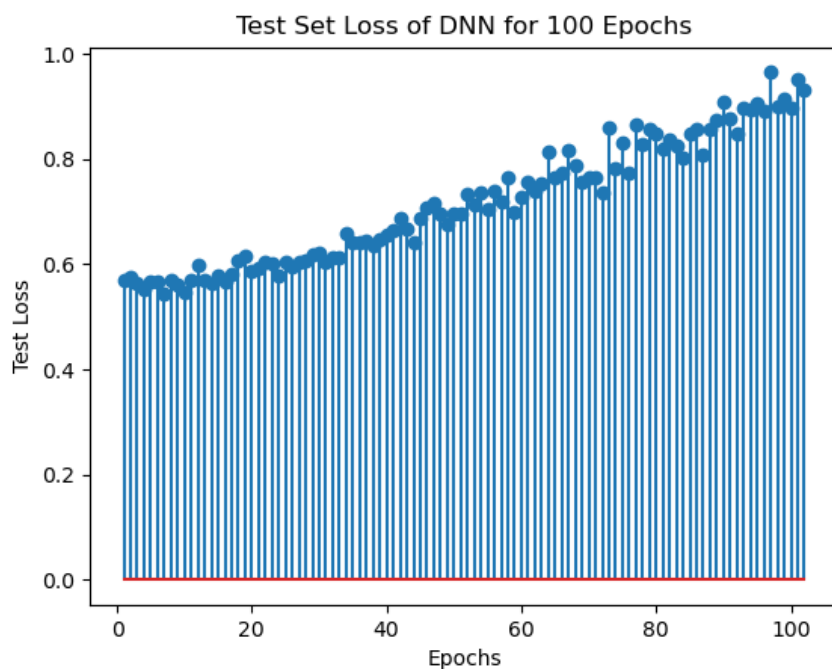
Epoch Loss = 0.3497799572219952
Accuracy = 0.8334189791024316
Recall = 0.833607668898285
F1 Score = 0.829220576452432

TEST DATASET EVALUATION:

Epoch Loss = 0.9324116157857996
Accuracy = 0.7009189640768588
Recall = 0.7009055424822492
F1 Score = 0.6953624779012958

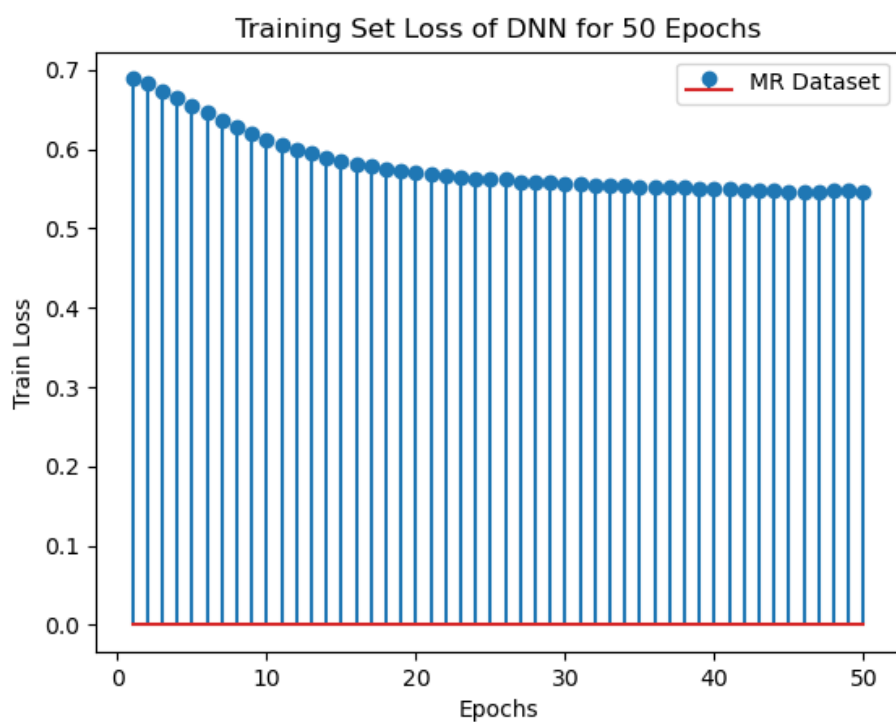
Παρακάτω φαίνονται οι γραφικές παραστάσεις των training Loss και test Loss, ανά τις 100 εποχές:

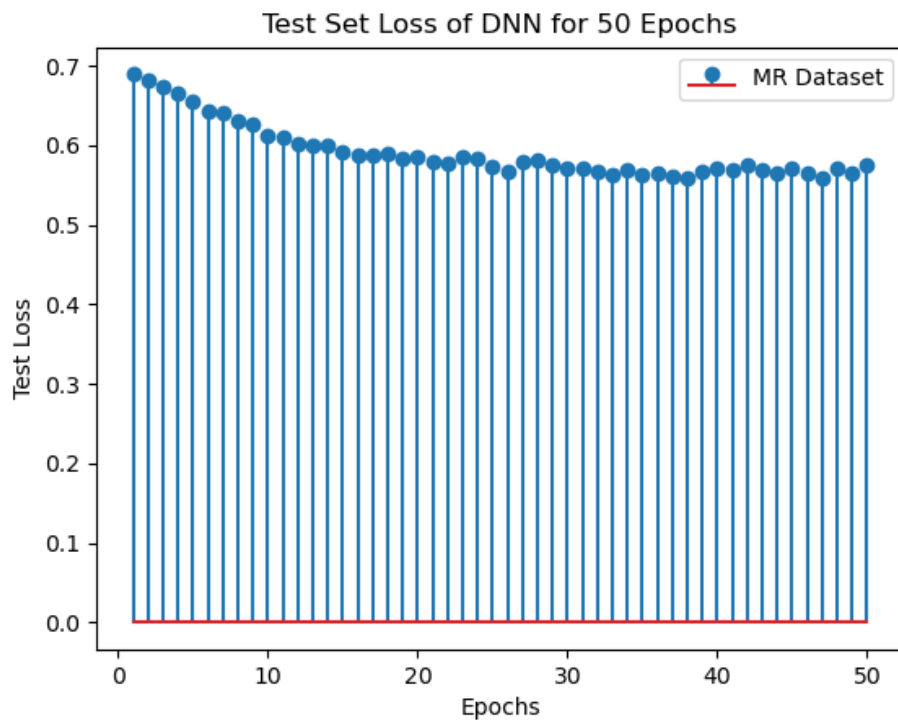




Παρατηρούμε ότι ενώ το training loss διαρκώς μειώνεται καθώς περνάνε οι εποχές, το test loss μετά από ένα σημείο αρχίζει να αυξάνεται. Αυτό συμβαίνει διότι το μοντέλο μας παθαίνει **overfitting** και χάνει τη δυνατότητα του να γενικεύει την γνώση που συσώρευσε και σε καινούρια δεδομένα.

Προκειμένου να επιλύσουμε το πρόβλημα αυτό, μειώνουμε το learning rate σε 0.0001 και αυξάνουμε το batch size στην τιμή 64. Σε διάστημα 50 εποχών, λοιπόν, εκπαιδεύουμε εκ νέου το μοντέλο μας και λαμβάνουμε τα κάτωθι αποτελέσματα:





TRAIN DATASET EVALUATION:

Epoch Loss = 0.5467184979444856

Accuracy = 0.7209394904458599

Recall = 0.7213421381911869

F1 Score = 0.7179208306505276

TEST DATASET EVALUATION:

Epoch Loss = 0.559230013327165

Accuracy = 0.7336002066115703

Recall = 0.7330914043836677

F1 Score = 0.7318165751506425

Τα αποτελέσματα αυτά είναι σαφώς πολύ καλύτερα, επιτυγχάνοντας accuracy μέχρι και 73% για τα δύο datasets. Σημειώνεται ότι ως criterion χρησιμοποιήσαμε τη συνάρτηση *BCEwithlogits()* της βιβλιοθήκης torch, καθώς πρόκειται για ένα πρόβλημα κατηγοριοποίησης 2 κλάσεων.

- Για το Dataset 'Semeval2017':

Έχοντας επιλέξει **EPOCHS** = 50 και **BATCH_SIZE** = 64, και χρησιμοποιώντας τα glove.twitter.6B.100d.txt embeddings και SGD optimizer με learning rate = 0.001, στο πέρας της εκπαίδευσης λαμβάνουμε τα ακόλουθα αποτελέσματα, σταματώντας την εκπαίδευση προτού αρχίσει να αυξάνεται το σφάλμα του test dataset, ώστε να αποφύγουμε το **overfitting**:

TRAIN DATASET EVALUATION:

Epoch Loss = 0.7204548695010524

Accuracy = 0.6700260910815938

Recall = 0.612707683756316

F1 Score = 0.6218353541600296

TEST DATASET EVALUATION:

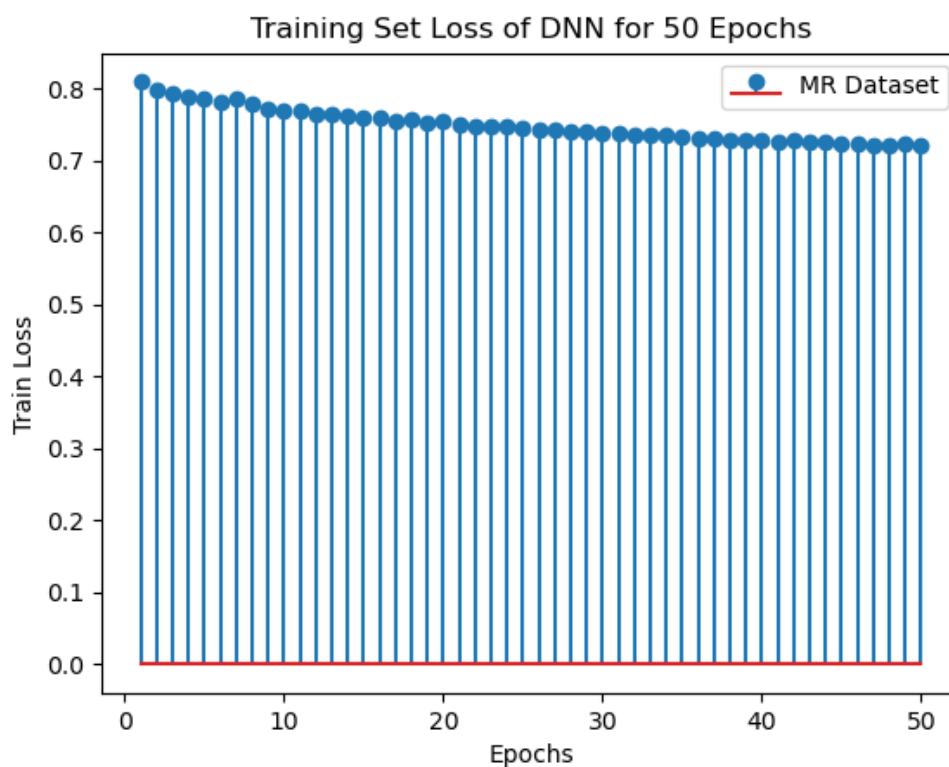
Epoch Loss = 0.8209592653438449

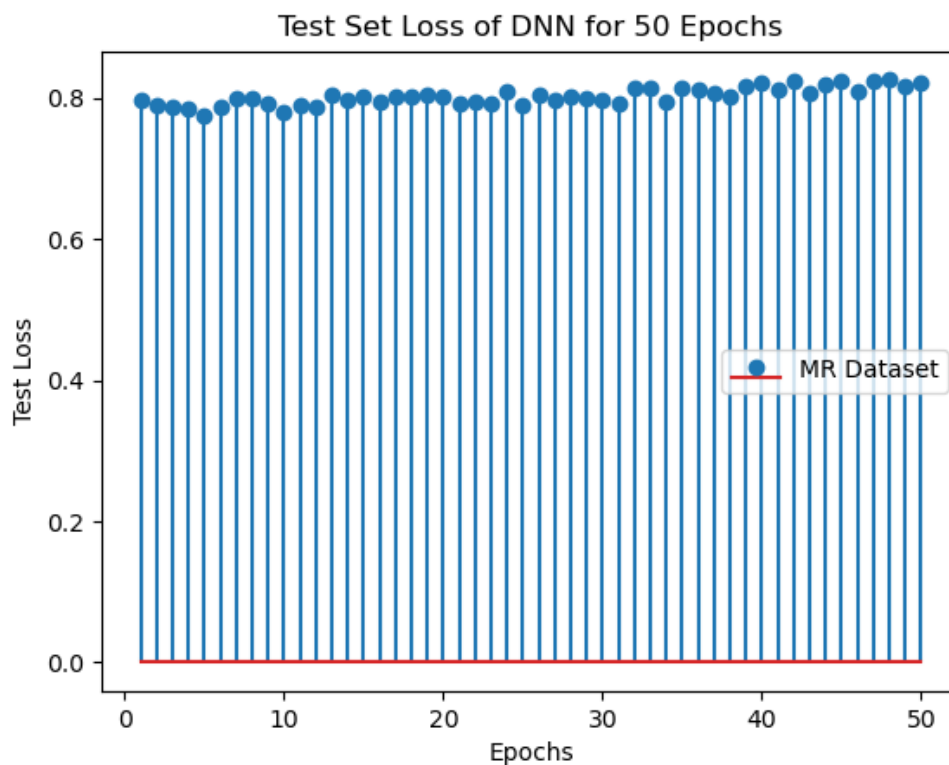
Accuracy = 0.6149956597222223

Recall = 0.6050296437102639

F1 Score = 0.5980204452941111

Παρακάτω φαίνονται οι γραφικές παραστάσεις των training Loss και test Loss, ανά τις 100 εποχές:





Παρατηρούμε ότι οι επιδόσεις του εν λόγω μοντέλου **δεν είναι αρκετά ικανοποιητικές**, καθώς επιτυγχάνεται μετά βίας accuracy 61%. Αυτό μπορεί να οφείλεται στο επιθετικό pre-processing που ακολουθήσαμε. Σίγουρα πάντως το μοντέλο επιδέχεται βελτιώσεις, κάτι το οποίο θα επιδιώξουμε και στο κύριο μέρος της εργαστηριακής άσκησης με τη χρήση των RNNs.

References:

- (1) <https://machinelearningmastery.com/how-to-control-the-speed-and-stability-of-training-neural-networks-with-gradient-descent-batch-size/>
- (2) <https://medium.com/mini-distill/effect-of-batch-size-on-training-dynamics-21c14f7a716e>
- (3) https://en.wikipedia.org/wiki/Universal_approximation_theorem