ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

A coin with a head on it

Description automatically generated

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

«ΑΞΙΟΠΟΙΗΣΗ ΜΕΓΑΛΩΝ ΓΛΩΣΣΙΚΩΝ ΜΟΝΤΕΛΩΝ ΓΙΑ ΤΗΝ ΑΝΑΠΤΥΞΗ ΑΥΤΟΜΑΤΩΝ ΕΛΕΓΧΩΝ ΑΠΟΔΟΧΗΣ ΛΟΓΙΣΜΙΚΟΥ»

ΠΛΑΤΙΑΣ ΚΩΝΣΤΑΝΤΙΝΟΣ

ΑΡΙΘΜΟΣ ΜΗΤΡΩΟΥ: P3200157

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ:  
ΔΙΑΜΑΝΤΙΔΗΣ ΝΙΚΟΛΑΟΣ

ΑΘΗΝΑ, ΑΥΓΟΥΣΤΟΣ 2024

© Copyright

Κωνσταντίνος Πλατιάς

Σημείωμα Συγγραφέα

Το δοκίμιο αυτό αποτελεί πτυχιακή εργασία που συντάχθηκε για το Προπτυχιακό Πρόγραμμα Σπουδών του τμήματος Πληροφορικής του ΟΙΚΟΝΟΜΙΚΟΥ ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΑΘΗΝΩΝ και υποβλήθηκε τον Αύγουστο του 2024.

Ο συγγραφέας βεβαιώνει ότι το περιεχόμενο του παρόντος έργου είναι αποτέλεσμα προσωπικής εργασίας και ότι έχει γίνει η κατάλληλη αναφορά στην εργασία τρίτων -όπου κάτι τέτοιο ήταν απαραίτητο-, σύμφωνα με τους κανόνες της ακαδημαϊκής δεοντολογίας.

Πινακασ περιεχομενων

[1. ΕΙΣΑΓΩΓΗ 2](#_Toc172911308)

[2. ΜΕΓΑΛΑ ΓΛΩΣΣΙΚΑ ΜΟΝΤΕΛΑ 3](#_Toc172911309)

[2.1 Τι είναι ένα νευρωνικό δίκτυο και πως λειτουργεί 3](#_Toc172911310)

[2.1.1 Ορισμός νευρωνικών δικτύων 3](#_Toc172911311)

[2.1.2 Ορισμός της μείωσης βαρών με καθοδική κλίση 4](#_Toc172911312)

[2.1.3 Ορισμός του αλγόριθμου ανάστροφης μετάδοσης 6](#_Toc172911313)

[2.2 Τι είναι ένα μεγάλο γλωσσικό μοντέλο 6](#_Toc172911314)

[2.2.1 Ορισμός μεγάλου γλωσσικού μοντέλου 6](#_Toc172911315)

[2.2.2 Τι κάνει ένα transformer 7](#_Toc172911316)

[2.2.3 Τι είναι το attention layer 8](#_Toc172911317)

[2.2.4 Τι είναι το feed forward step 8](#_Toc172911318)

[2.3 8](#_Toc172911319)

[3. ΒΙΒΛΙΟΓΡΑΦΙΑ 9](#_Toc172911320)

# ΕΙΣΑΓΩΓΗ

# ΜΕΓΑΛΑ ΓΛΩΣΣΙΚΑ ΜΟΝΤΕΛΑ

## Τι είναι ένα νευρωνικό δίκτυο και πως λειτουργεί

### Ορισμός νευρωνικών δικτύων

A diagram of a network

Description automatically generatedΞεκινώντας την συζήτηση για τα μεγάλα γλωσσικά μοντέλα και την χρήση τους, δεν θα μπορούσε να παραληφθεί η αναφορά στην βασική συνιστώσα που τα αποτελεί, το ονομαζόμενο νευρωνικό δίκτυο αλλά και τους τρόπους με τους οποίους αυτό λειτουργεί. Ένα νευρωνικό δίκτυο, όπως γίνεται αντιληπτό και από το ίδιο το όνομα, είναι ένα δίκτυο πολλών συνδεδεμένων νευρώνων, χωρισμένων σε διαφορετικά «στρώματα» όπου κάθε νευρώνας μπορεί να συσχετιστεί ως ένα αντικείμενο το οποίο παίρνει κάποιες εισόδους, εκτελεί κάποιους πολύ απλούς υπολογισμούς με αυτές και στην συνέχεια παράγει κάποια έξοδο, ένα νούμερο, το οποίο με την σειρά του περνάει σαν είσοδο στους επόμενους. Συγκεκριμένα, οι απλοί αυτοί υπολογισμοί είναι το άθροισμα των εισόδων που δέχθηκε, πολλαπλασιασμένοι με ένα νούμερο, το ονομαζόμενο «βάρος», το οποίο μαθαίνει το νευρωνικό δίκτυο από χιλιάδες δεδομένα κατά την εκπαίδευσή του. Στην συνέχεια, αφού εφαρμοστεί ένας αριθμητικός μετασχηματισμός, το αποτέλεσμα αυτό, όπως αναφέρθηκε και προηγουμένως, μεταφέρεται στους νευρώνες του επόμενου «στρώματος» σαν είσοδο και η διαδικασία αυτή σταματά στο τελευταίο «στρώμα» νευρώνων του νευρωνικού δικτύου. Για παράδειγμα, μπορεί το αποτέλεσμα που προωθεί νευρώνας στο επόμενο στρώμα να είναι μηδέν εάν το άθροισμα που υπολόγισε μέσω των εισόδων είναι αρνητικός αριθμός, διαφορετικά εάν είναι θετικός αριθμός ή μηδέν, να προωθεί απλά το ίδιο το άθροισμα χωρίς αλλαγές (ονομαζόμενη «συνάρτηση ενεργοποίησης»). Συνεχίζοντας, τα μοντέλα που χρησιμοποιούν βαθιά μάθηση, χρησιμοποιούνε πολλά εκατομμύρια νευρώνες, χωρισμένα σε πολυάριθμα στρώματα με δισεκατομμύρια βάρη και πολύπλοκες διατάξεις νευρώνων, όπου όμως η βασική ιδέα παραμένει η παραπάνω. (Ανδρουτσόπουλος, 2024)

Εικόνα 1. Παράδειγμα απλού νευρωνικού δικτύου

### Ορισμός της μείωσης βαρών με καθοδική κλίση

Αυτό που κάνει τα νευρωνικά δίκτυα τόσο ενδιαφέροντα, και ως αποτέλεσμα, και την μηχανική μάθηση, είναι ότι στην ουσία, ποτέ δεν γράφεται κάποιος αλγόριθμος ο οποίος ορίζει συγκεκριμένα τι θα πρέπει να κάνει ένα νευρωνικό δίκτυο και πως θα πρέπει να παράγει αποτελέσματα, αλλά από την άλλη, γράφεται ένας αλγόριθμος ο οποίος θα μπορεί μέσω της εισαγωγής εκατομμυρίων παραδειγμάτων και δεδομένων με τις ορθές «ετικέτες» τους (το επιθυμητό δηλαδή αποτέλεσμα), να μεταβάλλει με τέτοιο τρόπο τα εκατομμύρια βάρη από τα οποία αποτελείται, ώστε να αποδίδει καλύτερα στα παραδείγματα αυτά. Τα δεδομένα αυτά ονομάζονται «δεδομένα εκπαίδευσης», τα οποία παράλληλα με τα «δεδομένα δοκιμής», τα οποία είναι παραδείγματα όπου το νευρωνικό δίκτυο δεν έχει «ξαναδεί» και τα οποία χρησιμοποιούμε για να αξιολογήσουμε την αποτελεσματικότητα του νευρωνικού μας δικτύου, αποτελούν την συνολική αξιολόγηση του συστήματός μας. Το πρόβλημα αυτό της μεταβολής των βαρών που καλείται να λύσει ένας αλγόριθμος, καταλήγει τελικά να είναι η εύρεση του ελαχίστου μιας «συνάρτησης κόστους». Η συνάρτηση αυτή στην γενική εικόνας της υπολογίζεται με βάση τα αποτελέσματα που παράγει το νευρωνικό δίκτυο και τα επιθυμητά αποτελέσματα που έχουμε για κάθε παράδειγμα, και συνεπώς έχει μεγάλη τιμή εάν τα αποτελέσματα διαφέρουν σε μεγάλο βαθμό από τα επιθυμητά και μικρό στην αντίθετη περίπτωση. Σκοπός του αλγορίθμου είναι με βάση τον μέσο όρο όλων αυτών των τιμών κόστους για κάθε παράδειγμα, να προσπαθήσει να μεταβάλλει τις τιμές των βαρών της συνάρτησης κόστους, ώστε αυτή να φτάσει σε ένα τοπικό ελάχιστο. Η τεχνική αυτή ονομάζεται «**καθοδική κλίση»** ή όπως είναι γνωστό, gradient descent, αφού προσπαθεί να βρει ένα τοπικό ελάχιστο μίας συνάρτησης με πολλές χιλιάδες μεταβλητές (τις εισόδους και τα βάρη), η οποία εάν A computer graphics of a red and yellow wavy surface

Description automatically generated with medium confidenceαναπαρασταθεί σε ένα διανυσματικό χώρο, έχει την εικόνα ενός «γεωγραφικού τοπίου» στο οποίο πρέπει να βρεθεί ένα τοπικό «χαμηλότερο σημείο», όπως φαίνεται και στην *Εικόνα 2*. Χρησιμοποιώντας μαθηματικές έννοιες, όλα τα παραπάνω καταλήγουν τελικά στην εύρεση του αρνητικού του «ανάδελτα ή ∇» της συνάρτησης κόστους, το οποίο δείχνει προς την πιο απότομη μείωση μίας συνάρτησης (Sanderson, 2017).A black and white text with numbers and symbols

Description automatically generated with medium confidence

Εικόνα 2. Γεωγραφική αναπαράσταση ενός παραδείγματος Gradient descent

Εικόνα 3. Παράδειγμα μαθηματικής αναπαράστασης του ανάδελτα

Κατά συνέπεια, σε αυτό το σημείο θα αναφερθεί περιληπτικά ο ορισμός του αλγορίθμου που προσπαθεί να επιτύχει όλα τα παραπάνω, δηλαδή να υπολογίσει αυτό το τοπικό ελάχιστο της συνάρτησης κόστους, ο οποίος αναφέρεται ως «**αλγόριθμος ανάστροφης μετάδοσης**»

### Ορισμός του αλγόριθμου ανάστροφης μετάδοσης

Όπως παρουσιάστηκε παραπάνω, ο αλγόριθμος ανάστροφης μετάδοσης είναι ένας αλγόριθμος εύρεσης ενός τοπικού ελαχίστου μέσω του υπολογισμού του αρνητικού ανάδελτα μίας συνάρτησης κόστους. Αρχικά, ο αλγόριθμος αρχικοποιεί όλα τα βάρη του νευρωνικού δικτύου με τυχαίες μικρές τιμές και για μία δεδομένη είσοδο/παράδειγμα εκπαίδευσης, υπολογίζει το συνολικό σφάλμα στην τελική έξοδο, συγκρίνοντας την πραγματική έξοδο με την επιθυμητή έξοδο. Στην συνέχεια, το σφάλμα μεταδίδεται από την έξοδο προς την είσοδο υπολογίζοντας παράλληλα τους παραγώγους ως προς κάθε ξεχωριστό βάρος, με τον κανόνα της αλυσίδας, και κάθε βάρος ενημερώνεται ώστε να δείχνει (με την τεχνική που αναφέραμε νωρίτερα της καθοδικής κλίσης) προς την κατεύθυνση που μειώνεται το σφάλμα. Η παραπάνω διαδικασία γίνεται για κάθε παράδειγμα εκπαίδευσης που δίνεται στο νευρωνικό δίκτυο, οι οποίες ονομάζονται και «εποχές» και τελειώνει είτε μόλις το σύστημα ξεπεράσει έναν μέγιστο αριθμό εποχών, είτε εάν το συνολικό σφάλμα μειωθεί σε επιθυμητό αριθμό ( Ανδρουτσόπουλος, 2023).

## Τι είναι ένα μεγάλο γλωσσικό μοντέλο

### Ορισμός μεγάλου γλωσσικού μοντέλου

Όπως αναφέρει και ο Ανδρουτσόπουλος με απλά λόγια, ένα μεγάλο γλωσσικό μοντέλο είναι και αυτό ένα νευρωνικό δίκτυο, το οποίο όμως παίρνει σαν εισόδους λέξεις, ή καλύτερα tokens, σε μορφή αριθμών, τα οποία αποτελούν ίσως ένα ημιτελές κείμενο, και παράγει ως εξόδους όλες τις πιθανές λέξεις που θα μπορούσαν να είναι η επόμενη λέξη, επιλέγοντας αυτήν με την μεγαλύτερη πιθανότητα ορθότητας. Έτσι, βασιζόμενοι στο μεγάλο γλωσσικό μοντέλο, μπορούμε να πάρουμε την νέα αυτή πρόταση που παρήγαγε, να την ξαναδώσουμε σαν είσοδο και να παράγει ξανά άλλη μία επόμενη πιθανή λέξη και έτσι μετά από έναν αριθμό επαναλήψεων της διαδικασίας αυτής να παραχθεί μια ολοκληρωμένη απάντηση η οποία να βγάζει ένα νόημα (2024, σελίδα 9).

### Τι κάνει ένα transformer

Πιο συγκεκριμένα, τα μεγάλα γλωσσικά μοντέλα σπάνε το κείμενο που δέχονται σαν είσοδο σε διαφορετικά tokens, τα οποία δεν είναι απαραίτητα λέξεις, αλλά μπορεί να αποτελούν και κομμάτια λέξεων, και στην συνέχεια τα αναπαριστά σε διανύσματα. Τα διανύσματα αυτά αποτελούνται από πολλές χιλιάδες μεταβλητές και αριθμούς και τοποθετούνται σε έναν πολυδιάστατο χώρο, όπου λέξεις με παρόμοια σημασία όπως για παράδειγμα «γάτα» και «σκύλος» βρίσκονται πολύ κοντά. Η χρήση διανυσμάτων επιτρέπει στα μεγάλα γλωσσικά μοντέλα να πραγματοποιούν μαθηματικές πράξεις που αποκαλύπτουν σχέσεις μεταξύ λέξεων. Για παράδειγμα έχει διαπιστωθεί ότι αν από το διάνυσμα της λέξης "μεγαλύτερος" αφαιρεθεί το διάνυσμα της λέξης "μεγάλος" και προστεθεί το διάνυσμα της λέξης "μικρός", το αποτέλεσμα θα είναι το διάνυσμα της λέξης "μικρότερος". Τα μεγάλα γλωσσικά μοντέλα μπορούν να αντιπροσωπεύουν τις λέξεις με διαφορετικά διανύσματα ανάλογα με τα συμφραζόμενα. Αυτό επιτυγχάνεται με τη χρήση ενός αρχιτεκτονικού μοντέλου νευρωνικού δικτύου, γνωστού ως «**transformer»,** που ενημερώνει τα διανύσματα των λέξεων μέσω πολλαπλών επιπέδων. A close-up of words

Description automatically generated Κάθε μεγάλο γλωσσικό μοντέλο αποτελείται από πολλά στρώματα transformers συνδεδεμένα μεταξύ τους, όπου σκοπός του κάθε ενός είναι να εμπλουτίσει με πληροφορία κάθε ένα token με βάση τα γύρω και έτσι να αλλάξει ως προς την σωστή «κατεύθυνση» το διάνυσμα του για να έχει την σωστή σημασία που αποκαλύπτεται από τα συμφραζόμενα. Η διαδικασία αυτή επαναλαμβάνεται για κάθε transformer το οποίο προσθέτει και βελτιώνει τα διανύσματα έως το τελικό στρώμα.

Εικόνα 4. Παράδειγμα επικοινωνίας των transformers σε ένα μεγάλο γλωσσικό μοντέλο

### Τι είναι το attention layer

Ένα κύριο συστατικό το οποίο αποτελεί το transformer και του επιτρέπει να εμπλουτίζει τα διανύσματα, είναι το «attention layer». Συγκεκριμένα, το συστατικό αυτό επιτρέπει στα tokens να ανταλλάσσουν πληροφορίες μεταξύ τους και να εμπλουτίζουν το ένα το άλλο, ώστε να προκύπτει η σωστή σημασία κάθε ενός. Για παράδειγμα, η λέξη «μοντέλο» έχει διαφορετική σημασία ανάλογα τα συμφραζόμενα, όπως για παράδειγμα στην πρόταση «μοντέλο μαθηματικών» και «μοντέλο του Χόλυγουντ». Η λειτουργία του attention layer είναι να γίνει ο σωστός αυτός διαχωρισμός για κάθε διαφορετικό token με βάση τα γύρω του. Η διαδικασία αυτή επιτρέπει στα μεγάλα γλωσσικά μοντέλα να μαντεύουν σωστά την επόμενη λέξη σε κάθε κείμενο.

### Τι είναι το feed forward step

Μετά τη μεταφορά πληροφοριών ανάμεσα σε διανύσματα λέξεων από τα attention heads, to transformer αποτελείται και από ένα ακόμη συστατικό, ένα στρώμα που ονομάζεται feed forward, το οποίο «σκέφτεται» κάθε διάνυσμα λέξης και προσπαθεί να προβλέψει την επόμενη λέξη. Σε αυτό το στάδιο δεν γίνεται ανταλλαγή πληροφοριών μεταξύ των λέξεων, το στρώμα feed forward αναλύει κάθε λέξη μεμονωμένα. Ωστόσο, το στρώμα feed forward έχει πρόσβαση σε οποιαδήποτε πληροφορία αντιγράφηκε προηγουμένως από όλα τα προηγούμενα attention heads ώστε να μπορέσει αποτελεσματικότερα να προβλέψει την επόμενη λέξη. (Lee, 2023)

## 

# ΒΙΒΛΙΟΓΡΑΦΙΑ

Lee, T. B. (2023, Ιούλιος 31). Ανάκτηση από arstechnica.com: https://arstechnica.com/science/2023/07/a-jargon-free-explanation-of-how-ai-large-language-models-work/6/

Sanderson, G. (2017, Οκτώβριος 16). *3blue1brown.* Ανάκτηση από 3blue1brown: https://www.3blue1brown.com/lessons/gradient-descent#another-way-to-think-about-the-gradient

Ανδρουτσόπουλος, Ί. (2024, Φεβρουάριος). Τεχνητή Νοημοσύνη και Μεγάλα Γλωσσικά Μοντέλα. *ΟΠΑ News Εφημερίδα Οικονομικού Πανεπιστημίου Αθηνών*, σσ. 8-9.