

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Η/Υ

ΜΗ ΓΡΑΜΜΙΚΗ ΜΕΙΩΣΗ ΔΙΑΣΤΑΣΕΩΝ
ΜΕ ΤΗ ΧΡΗΣΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ
LOCALLY LINEAR EMBEDDINGS ΚΑΙ
ΣΤΟΧΟ ΤΗ ΒΕΛΤΙΩΣΗ ΤΟΥ
ΠΟΣΟΣΤΟΥ ΤΑΞΙΝΟΜΗΣΗΣ
ΣΕ ΕΦΑΡΜΟΓΕΣ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ
ΚΑΙ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Εκπόνηση:
Πέτρος Κατσιλέρος

Επίβλεψη:
Νικόλαος Πιτσιάνης
Νίκος Σισμάνης

Με την εκπόνηση της εν λόγω διπλωματικής εργασίας ολοκληρώνεται ο κύκλος των
προπτυχιακών μου σπουδών αποκτώντας δίπλωμα Ηλεκτρολόγου Μηχανικού και Μηχανικού
Η/Υ απο το Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης.

Περίληψη

Η τεχνητή νοημοσύνη μέσω της μηχανικής μάθησης είναι αναμφισβήτητα ένας επιστημονικός κλάδος ο οποίος επικεντρώνει το ενδιαφέρον ολοένα και περισσότερων μηχανικών-ερευνητών. Το γεγονός αυτό οφείλεται στην επιτυχία τέτοιου είδους εφαρμογών σε διάφορους κλάδους της καθημερινότητάς μας όπως αυτός της ρομποτικής, της υγείας, της εξόρυξης γνώσης κλπ. Επίσης οι σημερινοί υπολογιστές λόγω της ραγδαίας ανάπτυξης της τεχνολογίας παρέχουν τους απαραίτητους πόρους ώστε να μπορέσουν να αναπτυχθούν και να διερευνηθούν τέτοιου είδους προβλήματα. Παρ' όλα αυτά, όσους πόρους και αν διαθέσουμε δεν μπορούμε σε καμιά περίπτωση να δημιουργήσουμε κάτι αντίστοιχο με τον ανθρώπινο εγκέφαλο.

Γνωρίζουμε ότι ο ανθρώπινος εγκέφαλος είναι ένα τρομερά περίπλοκο σύστημα εκατομμυρίων νευρώνων συνδεδεμένων μεταξύ τους οι οποίοι είναι σε θέση να εκτελούν σε κλάσματα του δευτερολέπτου έναν τεράστιο αριθμό λογικών πράξεων. Το μοντέλο αυτό είναι αδύνατον να προσομοιωθεί με οποιοδήποτε υπολογιστικό σύστημα διαθέτει ο άνθρωπος σήμερα. Στην προσπάθεια των Μηχανικών να μοντελοποιήσουν τις λειτουργίες του λαμβάνοντας φυσικά υπόψιν ευρήματα και αποτελέσματα των επιστημόνων της Ιατρικής σημαντικές λύσεις και βελτιστοποιήσεις έρχονται να δώσουν αλγόριθμοι οι οποίοι έχουν ως στόχο να μειώσουν τις παραμέτρους τις οποίες πρέπει να εκτιμήσει κάποιο υπολογιστικό σύστημα ώστε τελικά να μπορέσει να εξάγει συμπεράσματα ανάλογα με αυτά ενός ανθρώπου.

Χαρακτηριστικά παραδείγματα τέτοιων εφαρμογών με τα οποία καταπιάνεται και η εργασία αυτή είναι η χαρακτηριστικών σε μοτίβα εικόνων ή άλλων δεδομένων με στόχο την εξαγωγή συμπεράσματος για την ταξινόμηση των δεδομένων σε κλάσεις. Συγκεκριμένα γίνεται εφαρμογή του αλγορίθμου Locally Linear Embedding σε δύο σετ δεδομένων με εικόνες ψηφία αριθμών και σε ένα με ιατρικά δεδομένα απο καρκινοπαθείς και μη ασθενείς. Αφού γίνει μείωση των διαστάσεων που πρέπει να ληφθούν υπόψιν για την εξαγωγή συμπεράσματος εφαρμόζεται ο ταξινομητής κοντινότερων γειτόνων ο οποίος εξάγει και το τελικό συμπέρασμα για την ταξινόμηση των δεδομένων στις κατάλληλες κλάσεις.

Ευχαριστίες

Με την ολοκλήρωση αυτής της διπλωματικής εργασίας Θα ήθελα καταρχήν να ευχαριστήσω τον κ.Νικόλαο Πιτσιάνη επίκουρο καθηγητή του τμήματός μου ο οποίος μου έδωσε το ερέθισμα καθώς και χρήσιμες συμβουλές αλλά και πόρους ώστε να μπορέσω να ολοκληρώσω την έρευνα για το συγκεκριμένο θέμα. Επίσης ένα μεγάλο ευχαριστώ στον υποψήφιο διδάκτορα του τμήματος Νίκο Σισμάνη για την καθοδήγηση του καθ'όλη την διάρκεια εκπλήρωσης της εργασίας μου αυτής.

Τέλος, ένα πολύ θερμό και μεγάλο ευχαριστώ στους γονείς μου οι οποίοι με στήριξαν τόσο οικονομικά όσο και ψυχολογικά όλα αυτά τα χρόνια ώστε να μπορέσω να αποκτήσω το δίπλωμά μου. Στο σημείο αυτό δεν θα μπορούσα να παραλείψω τον σκύλο μου, τους φίλους και την κοπέλα μου διότι ο καθένας με τον τρόπο του βοήθησαν στην αντιμετώπιση των δυσκολιών που συνάντησα καθ'όλη την διάρκεια των σπουδών μου.

Κατσιλέρος Πέτρος
Θεσσαλονίκη, Μάρτιος 2016

Αφιέρωση

Αφιερώνω την διπλωματική αυτή εργασία πρωτίστως στον εαυτό μου για τον κόπο μου όλα αυτά τα χρόνια ώστε να μπορέσω να αποκτήσω το δίπλωμα Ηλεκτρολόγου Μηχανικού και Μηχανικού Η/Υ και κατά δεύτερον στους γονείς μου οι οποίοι με στήριξαν ανελλιπώς και με κάθε τρόπο σε όλη αυτή την πορεία.

Κατσιλέρος Πέτρος
Θεσσαλονίκη, Μάρτιος 2016

Τι φέρνει μια χαλώνα απο την Κίνα ...

Το πτυχίο μου !!!

Περιεχόμενα

1	Εισαγωγή	13
1.1	Αναγνώριση προτύπων και μηχανική μάθηση	13
1.2	Ερεθίσματα απο τον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου	14
1.3	Μάθηση με επίβλεψη	15
1.4	Μείωση της διάστασης των δεδομένων	17
2	Αλγόριθμοι μείωσης διαστάσεων	19
2.1	Μαθηματικό και θεωρητικό υπόβαθρο	19
2.2	Αλγόριθμοι για γραμμική μείωση διαστάσεων	20
2.2.1	PCA	20
2.2.2	MDS	20
2.3	Αλγόριθμοι για μη γραμμική μείωση διαστάσεων	20
2.3.1	ISOMAP	20
2.3.2	Laplacian Eigenmaps	20
2.3.3	LLE	20
3	Ανάλυση του αλγορίθμου Locally Linear Embeddings	21
3.1	Βήμα-1: Εύρεση του πίνακα γειτνίασης	21

3.2	Βήμα-2: Κατασκευή του Laplacian	21
3.3	Βήμα-3: Επίλυση του προβλήματος εύρεσης ιδιτιμών και ιδιοδιανυσμάτων	21
3.4	Βήμα-4: Επιλογή των τελικών διαστάσεων	21
4	Πειράματα	22
4.1	Σετ δεδομένων	22
4.2	Σχεδιασμός και οργάνωση των πειραμάτων	22
4.3	Αποτελέσματα	22
5	Συμπεράσματα	23
5.1	Συμπεράσματα για τα πειράματα	23
5.2	Συμπεράσματα για τον αλγόριθμο Locally Linear Embeddings	23
5.3	Συμπεράσματα για την μείωση διαστάσεων σε αφαρμογές μηχανικής μάθησης και εξόρυξης γνώσης	23
	Βιβλιογραφία	23

Κατάλογος Πινάκων

Κατάλογος Σχημάτων

Κεφάλαιο 1

Εισαγωγή

1.1 Αναγνώριση προτύπων και μηχανική μάθηση

Αναγνώριση προτύπων καλείται η επιστημονική περιοχή που έχει στόχο την ταξινόμηση αντικειμένων σε κατηγορίες ή κλάσεις. Ανάλογα με την κάθε εφαρμογή τα δεδομένα μπορεί να είναι είτε εικόνες, είτε σήματα είτε οποιοδήποτε άλλο σκετ δεδομένων χρειάζεται για κάποιο λόγο να ταξινομηθεί. Στις μέρες μας η ανάγκη διαχείρισης αλλά και ανάκτησης πληροφοριών μέσω ηλεκτρονικών υπολογιστών αποκτά τεράστια σπουδαιότητα καταρχήν διότι ο όγκος των πληροφοριών αυξάνεται ραγδαία με ρυθμό αδύνατο να διαχειριστεί ο άνθρωπος και επίσης διότι η ανάπτυξη της τεχνολογίας μας παρέχει πολύ ισχυρά υπολογιστικά συστήματα με τη χρήση των οποίων μπορούμε να δημιουργήσουμε πολύπλοκα μοντέλα εξόρυξης γνώσης .

Αντίστοιχοι κλάδοι στους οποίους έχει τεράστια σημασία η αναγνώριση προτύπων είναι οι επιστημονικοί κλάδοι της Ιατρικής, της Βιολογίας, ο χώρος των αγορών και των επιχειρήσεων και τέλος η διαχείριση και η εξόρυξη γνώσης απο τον τεράστιο όγκο της πληροφορίας που είναι διαθέσιμος στο διαδίκτυο. Φυσικά η αναγνώριση προτύπων είναι ένα πολύ σημαντικό μέρος του κλάδου της μηχανικής μάθησης σε ρομποτικά/υπολογιστικά συστήματα.

Η υπολογιστική όραση για παράδειγμα είναι αντικείμενο ιδιαίτερα χρήσιμο τόσο στον χώρο της ρομποτικής όσο σε αυτόν της ιατρικής αλλά και προφανώς της βιομηχανίας. Τέτοιου είδους εφαρμογές έχουν εισέλθει πολύ δυναμικά στην καθημερινότητά μας τα τελευταία χρόνια. Συγκεκριμένα

στον χώρο της βιομηχανίας υπάρχουν συστήματα τα οποία επιβλέπουν μέσω μια κάμερας την γραμμή παραγωγής καθώς και ρομπότ τα οποία μεταφέρουν και συναρμολογούν αντικείμενα. Επίσης υπάρχουν εφαρμογές οι οποίες αναγνωρίζουν για παράδειγμα πρόσωπα τραβώντας μια εικόνα με το κινητό μας τηλέφωνο. Τέλος στον χώρο της αυτοκινητοβιομηχανίας δεν είναι λίγες αντίστοιχες εφαρμογές οι οποίες έχουν συμβάλει δυναμικά στην αυτόνομη οδήγηση αλλά και στην προειδοποίηση για εμπόδια κλπ.

Ιδιαίτερη έμφαση αξίζει να δωθεί στην εξόρυξη γνώσης σε κλάδους όπως στη βιολογία αλλά και στην ιατρική. Για παράδειγμα η πρόβλεψη εμφάνισης ασθενειών όπως ο καρκίνος μέσω αναγνώρισης συγκεκριμένων μοτίβων σε εικόνες απο μαγνητικό τομογράφο, η μελέτη της αλυσίδας του γενετικού υλικού αλλά και ο χώρος των εγχειρίσεων υψηλής ακρίβειας με τη χρήση της ρομποτικής.

1.2 Ερεθίσματα απο τον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου

Απο μελέτες που έχουν γίνει για την λειτουργία του ανθρώπινου εγκεφάλου γνωρίζουμε ότι για οποιοδήποτε σύνολο μετρήσεων προέρχεται για παράδειγμα είτε απο την όραση μας είτε απο την ακοή μας ο εγκέφαλός μας μετασχηματίζει το σύνολο των δεδομένων αυτών σε ένα νέο σύνολο χαρακτηριστικών. Με τον τρόπο αυτό, επιλέγοντας προφανώς κάθε φορά τα κατάλληλα χαρακτηριστικά, επιτυγχάνεται τεράστια συμπίεση του όγκου της πληροφορίας σε σύγκριση με τα αρχικά δεδομένα εισόδου. Αυτό σημαίνει λοιπόν ότι το μεγαλύτερο μέρος της πληροφορίας για παράδειγμα μια σκηνής που βλέπουμε και στην οποία θέλουμε να αναγνωρίσουμε τα αντικείμενα που περιέχονται, συμπιέζεται σε έναν πολύ μικρό αριθμό χαρακτηριστικών. Η παραπάνω διαδικασία χαρακτηρίζεται ως τεχνική μείωσης διάστασης γνωστή στην βιβλιογραφία με τον όρο Dimensionality Reduction.

Ας πάρουμε για παράδειγμα τον κλάδο της υπολογιστικής όρασης ο οποίος αποτελεί και αντικείμενο μελέτης της εν λόγω εργασίας και ας αναρωτηθούμε το εξής: Πόσο δύσκολο είναι για κάποιον απο εμάς να αναγνωρίσει κάποιο νούμερο αποτυπωμένο σε μια εικόνα. Η προφανής απάντηση είναι καθόλου. Και αυτή είναι μια πολύ σωστή απάντηση, διότι για τον ανθρώπινο εγκέφαλο το να

καταλάβει ότι το ψηφίο το οποίο βρίσκεται στην εικόνα είναι για παράδειγμα το 1 και όχι το 9 είναι ένα πολύ απλό πρόβλημα.

Πιο συγκεκριμένα βλέποντας μια οποιαδήποτε σκηνή ο ανθρώπινος εγκέφαλος προσπαθεί να εντοπίσει σημεία ενδιαφέροντος τα οποία αποτελούν χαρακτηριστικά σημεία της. Τέτοια μπορεί να είναι πολύ έντονες αλλαγές στην φωτεινότητα όπως για παράδειγμα γωνίες, κενά ή τρύπες. Στην συνέχεια εντοπίζει πιο σύνθετες γεωμετρίες όπως ευθείες ή καμπύλες γραμμές και τέλος προσδιορίζει πιο ολοκληρωμένες δομές τρισδιάστατων αντικειμένων. Το ίδιο ακριβώς γίνεται και στην παραπάνω περίπτωση με το ψηφίο. Εντοπίζουμε αρχικά ότι το μοτίβο του ψηφίου 1 είναι πολύ κοντά σε αυτά των ψηφίων επτά και τέσσερα αλλά σε καμιά περίπτωση δεν θα λέγαμε ότι έχει τρομερές ομοιότητες με αυτό του δύο ή του οχτώ για παράδειγμα.

Το παραπάνω παράδειγμα είναι ένα πολύ απλό δείγμα του τρόπου με τον οποίο ο ανθρώπινος εγκέφαλος προσπαθεί με κάθε τρόπο να ελαχιστοποιήσει τις παραμέτρους που πρέπει να εκτιμήσει. Φυσικά αν αναλογιστούμε ένα ρεαλιστικό περίπλοκο πρόβλημα της καθημερινότητάς μας θα δούμε ότι απαιτούνται πολύ πιο σύνθετοι υπολογισμοί και θα πρέπει να συνδιάσουμε ένα πλήθος παραμέτρους ώστε τελικά να καταλήξουμε στο τελικό συμπέρασμα για κάποια απόφαση. Σε κάθε περίπτωση όμως γίνεται τεράστια συμπίεση της αρχικής πληροφορίας μέσω τεχνικών μείωσης διαστάσεων ώστε να ελαχιστοποιηθούν οι παράμετροι που πρέπει να υπολογιστούν και προφανώς να επιταχυνθεί η διαδικασία εξαγωγής της τελικής μας απόφασης.

Το γεγονός αυτό και δεδομένου ότι το όραμα της επιστημονικής κοινότητας των μηχανικών που ασχολούνται με την μηχανική μάθηση και την εξόρυξη γνώσης είναι να δημιουργηθεί ένα μοντέλο αντίστοιχο με αυτό του ανθρώπινου εγκεφάλου δεν θα μπορούσε να τους αφήσει αδιάφορους ώστε να μελετήσουν και να αναπτύξουν αντίστοιχους αλγόριθμους με σκοπό να εφαρμοστούν σε μοντέλα εξόρυξης γνώσης.

1.3 Μάθηση με επίβλεψη

Ένα πολύ εύλογο ερώτημα το οποίο προκύπτει από την παραπάνω ανάλυση είναι πως ο ανθρώπινος εγκέφαλος έχει μάθει και τελικώς έχει αποθηκεύσει το σύνολο αυτών των μοντέλων για τον

κάθε αριθμό ή για οποιοδήποτε άλλο αντικείμενο ή μοτίβο μπορεί να αναγνωρίσει με τόσο μεγάλη ταχύτητα και ευκολία. Η απάντηση είναι προφανώς η συνεχής εκπαίδευση και η διαρκής υπενθύμιση των συγκεκριμένων προτύπων.

Πιο συγκεκριμένα ο άνθρωπος από την μέρα που αρχίζει να αλληλεπιδρά με το περιβάλλον παίρνει διάφορα ερεθίσματα τα οποία καιρό με τον καιρό μαθαίνει να τα ταξινομεί κατάλληλα και να τα χρησιμοποιεί όποτε ξαναεμφανιστούν μπροστά του. Τα ερεθίσματα αυτά είναι είτε εικόνες, είτε ήχοι είτε ερεθίσματα τα οποία μπορεί να προέρχονται από τις υπόλοιπες αισθήσεις του.

Ο τρόπος με τον οποίο καταφέρνουμε να συγκρατούμε και να μπορούμε να διαχειριστούμε ανα πάσα στιγμή τον τεράστιο όγκο πληροφοριών που βρίσκονται καταχωρημένες στον εγκέφαλό μας είναι ένας συνδυασμός τεχνικών μάθησης και συνεχούς εκπαίδευσης. Οι τεχνικές αυτές στον χώρο της τεχνητής νοημοσύνης αναφέρονται ως τεχνικές μάθησης με επίβλεψη, χωρίς επίβλεψη και με ημιεπίβλεψη. Θα μπορούσε κάποιος αρχικά να υποστηρίξει ότι ο ανθρώπινος εγκέφαλος χρησιμοποιεί κατεξοχήν τεχνικές μάθησης χωρίς επίβλεψη διότι μπορεί να μαθαίνει μόνος του νέα πράγματα. Είναι όμως πραγματικά αυτό το οποίο συμβαίνει; Η απάντηση είναι μάλλον όχι, και αυτό διότι από την πολύ νεαρή του ηλικία ο καθένας μας έχει γύρω του ανθρώπους οι οποίοι προσπαθούν συνεχώς να μας μεταφέρουν γνώση και να μας μάθουν τι βρίσκεται γύρω μας και πως να αλληλεπιδρούμε μεταξύ του. Παρόλα αυτά μετά από κάποιο σημείο ο ανθρώπινος εγκέφαλος αποκτά δυνατότητες με τις οποίες μπορεί αξιολογεί και να μαθαίνει μόνος του πολύ σύνθετα πράγματα αναλύοντάς τα σε απλούστερα προβλήματα τα οποία γνωρίζει ήδη πως να τα διαχειριστεί. Επίσης είναι στην φύση του ανθρώπου να εξερευνεί συνεχώς άγνωστα μονοπάτια και να αναζητεί απαντήσεις σε άγνωστα προβλήματα επιτυγχάνοντας αξιοθαύμαστα αποτελέσματα.

Από τα παραπάνω καταλήγουμε στο συμπέρασμα ότι ο άνθρωπος χρησιμοποιεί τεχνικές ημιεπίβλεψης για την εκπαίδευση του εγκεφάλου του γεγονός το οποίο του δίνει την δυνατότητα να μπορεί να διαχειριστεί αλλά και να μάθει πολύ σύνθετα μοντέλα. Μέσα από αυτή την διαδικασία είναι σε θέση με το πέρασμα του χρόνου να δημιουργήσει ένα τεράστιο και πανίσχυρο δίκτυο πληροφοριών, ταξινομημένο με τρόπο τον οποίο δεν μπορούμε ακόμα να εξηγήσουμε και να κατανοήσουμε. Με αυτό το μοντέλο είναι σε θέση ταχύτατα να αποφασίζει που βρίσκεται ο ευρύτερος χώρος της πληροφορίας που θέλει να αντλήσει και στην συνέχεια να αποφασίζει με τεράστια ακρίβεια και

ταχύτητα την τελική του απόφαση.

Το μοντέλο αυτό με το οποίο λειτουργεί ο ανθρώπινος εγκέφαλος είναι αν μη τι άλλο αξιοθαύμαστο και ανεξήγητο. Παρόλα είναι πολύ δύσκολο να εφαρμοστεί στον τομέα της τεχνητής νοημοσύνης και αυτό διότι ακόμα δεν είμαστε σε θέση να δώσουμε εξηγήσεις για τον τρόπο λειτουργίας του. Το συνηθέστερο και πιο αποτελεσματικό μέχρι στιγμής μοντέλο το οποίο χρησιμοποιείται στην εξόρυξη γνώσης μέσω ηλεκτρονικών υπολογιστών είναι αυτό της μάθησης με επίβλεψη. Σύμφωνα με το μοντέλο αυτό θα πρέπει αν συλλέξουμε ένα μεγάλο συνήθως όγκο δεδομένων τον οποίο να τροφοδοτήσουμε στην συνέχεια ως είσοδο στο σύστημά μας και με την κατάλληλη μεθοδολογία να το οδηγήσουμε να μάθει συγκεκριμένα μοντέλα τα οποία να μπορεί να χρησιμοποιήσει στην συνέχεια με σκοπό της εξαγωγή κάποιου συμπεράσματος.

1.4 Μείωση της διάστασης των δεδομένων

Στην παραπάνω διαδικασία δεδομένου ότι στις περισσότερες περιπτώσεις έχουμε να αντιμετωπίσουμε πολύ σύνθετα υπολογιστικά προβλήματα ο αριθμός των παραμέτρων που πρέπει να υπολογιστούν είναι σε συγκεκριμένες εφαρμογές απαγορευτικά μεγάλος. Σε κάποιες εφαρμογές το πρόβλημα είναι θέμα χρόνου όπου πρέπει να γίνει μείωση των παραμέτρων ώστε να ελαχιστοποιηθεί ο χρόνος εξαγωγής συμπεράσματος. Σε άλλες είναι θέμα χώρου διότι ένας μεγάλος αριθμός πολυδιάστατων δεδομένων μπορεί να αποτελεί πρόβλημα σε συγκεκριμένες εφαρμογές. Τέλος υπάρχουν περιπτώσεις στις οποίες χρειαζόμαστε την μείωση των διαστάσεων ώστε να διώξουμε εντελώς παραμέτρους οι οποίες επιδρούν σαν θόρυβος και επηρεάζουν αρνητικά την εξαγωγή ορθού συμπεράσματος ταξινόμησης. Προφανώς σε πολλές πρακτικές εφαρμογές επικρατεί ένας συνδυασμός των παραπάνω προβλημάτων.

Αντικείμενο λοιπόν της εν λόγω διπλωματικής εργασίας είναι η διερεύνηση και η χρήση του αλγορίθμου Locally Linear Embeddings για την μείωση των διαστάσεων σε πρακτικά προβλήματα όπως η αναγνώριση ψηφίων αλλά και η ταξινόμηση ασθενών με βάση το αν πρόκειται να εμφανίζουν κάποιας μορφής καρκίνου ή όχι. Τα αποτελέσματα των πειραμάτων είναι ιδιαίτερα ενθαρυντικά και δείχνουν σε όλες τις περιπτώσεις ότι η μείωση των διαστάσεων επιδρά δραματικά στην μείωση του

κόστους των υπολογισμών αλλά και στην αύξηση της σωστής πρόβλεψης λόγω απομάχρυνσης του θορύβου. Επίσης παρουσιάζονται δύο πρακτικές και ρεαλιστικές μέθοδοι εφαρμογής του αλγορίθμου σε πραγματικά προβλήματα από τις οποίες η πρώτη έρχεται να αντιμετωπίσει το πρόβλημα της πολύ μεγάλης μνήμης που απαιτεί η εκτέλεση του αλγορίθμου και η δεύτερη παρέχει την δυνατότητα για την ταξινόμηση των αποτελεσμάτων και την εξαγωγή συμπεράσματος σε πραγματικό χρόνο.

Κεφάλαιο 2

Αλγόριθμοι μείωσης διαστάσεων

2.1 Μαθηματικό και θεωρητικό υπόβαθρο

Έστω για παράδειγμα ότι έχουμε ένα σύνολο δειγμάτων εισόδου με αντίστοιχο διάνυσμα \mathbf{x} διάστασης $N \times 1$,

$$x^T = [x(0), \dots, x(N-1)]$$

2.2 Αλγόριθμοι για γραμμική μείωση διαστάσεων

2.2.1 PCA

2.2.2 MDS

2.3 Αλγόριθμοι για μη γραμμική μείωση διαστάσεων

2.3.1 ISOMAP

2.3.2 Laplassian Eigenmaps

2.3.3 LLE

Κεφάλαιο 3

Ανάλυση του αλγορίθμου Locally Linear Embeddings

- 3.1 Βήμα-1: Εύρεση του πίνακα γειτνίασης
- 3.2 Βήμα-2: Κατασκευή του Laplacian
- 3.3 Βήμα-3: Επίλυση του προβλήματος εύρεσης ιδιτιμών και ιδιοδιανυσμάτων
- 3.4 Βήμα-4: Επιλογή των τελικών διαστάσεων

Κεφάλαιο 4

Πειράματα

4.1 Σετ δεδομένων

4.2 Σχεδιασμός και οργάνωση των πειραμάτων

4.3 Αποτελέσματα

Κεφάλαιο 5

Συμπεράσματα

- 5.1 Συμπεράσματα για τα πειράματα
- 5.2 Συμπεράσματα για τον αλγόριθμο Locally Linear Embeddings
- 5.3 Συμπεράσματα για την μείωση διαστάσεων σε εφαρμογές μηχανικής μάθησης και εξόρυξης γνώσης