



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή
Τμήμα Ηλεκτρολόγων Μηχανικών &
Μηχανικών Υπολογιστών

Διπλωματική Εργασία

*Μελέτη manifold-based dimensionality
reduction τεχνικών για την ταξινόμηση
αντικειμένων*

Εκπόνηση:
Παπακωνσταντίνου Αντώνης
ΑΕΜ: 8977

Επίβλεψη:
Καθηγητής Α.Π.Θ.
Ιωάννης Θεοχάρης

Θεσσαλονίκη, Σεπτέμβριος 2023

Ευχαριστίες

Σε αυτό το σημείο, θα ήθελα να ευχαριστήσω ιδιαίτερα τον επιβλέποντα καθηγητή της διπλωματικής μου εργασίας κ. Ιωάννη Θεοχάρη για την εμπιστοσύνη που μου επέδειξε κατά την επιλογή του θέματος, τις πολύτιμες συμβουλές και τις ουσιαστικές παρατηρήσεις για την ολοκλήρωση της εργασίας. Ακόμη, τον ευχαριστώ για την ευκαιρία που μου έδωσε να ανακαλύψω νέες πτυχές της επιστήμης της τεχνητής νοημοσύνης. Ιδιαίτερες ευχαριστίες θα ήθελα να δώσω και στον υποψήφιο διδάκτορα Χρήστο Χαδουλό για τη συνεργασία, την υπομονή και την καθοδήγησή του κατά τη διάρκεια της εργασίας μου, καθώς και για τη βοήθεια και τις παρατηρήσεις του. Δεν μπορώ να παραλείψω να ευχαριστήσω τους φίλους και τους συμφοιτητές μου για τη στήριξή τους κατά τη διάρκεια των φοιτητικών μου χρόνων, καθώς και για τα υπέροχα φοιτητικά χρόνια που ζήσαμε. Τέλος, θα ήθελα να ευχαριστήσω ιδιαίτερα την οικογένειά μου για την αμέριστη υποστήριξη που μου έδειξαν κατά τη διάρκεια των σπουδών μου, η οποία αποτέλεσε ισχυρό κίνητρο για την επιτυχή ολοκλήρωσή τους.

Περίληψη

Οι μέθοδοι μείωσης διάστασης στη μηχανική μάθηση αποτελούν ισχυρά εργαλεία συμπίεσης δεδομένων υψηλής διάστασης σε απλούστερες μορφές, χαμηλότερης διάστασης, ενισχύοντας με αυτόν τον τρόπο τις δυνατότητες των συστημάτων αναγνώρισης προσώπων και μοτίβων. Αυτές οι μέθοδοι είναι κρίσιμες για την επίλυση της "κατάρας της διάστασης", μιας πρόκλησης διαχείρισης πολύπλοκων δεδομένων υψηλής διάστασης που είναι δυνατόν να κατακλύσει τους υπολογιστικούς πόρους και να αποκρύψει σημαντικά μοτίβα.

Συγκεκριμένα, οι τεχνικές αυτές χωρίζονται σε τρεις κατηγορίες: γραμμικές μέθοδοι μείωσης των διαστάσεων, μη γραμμικές ή μέθοδοι βασισμένες σε πολυεπίπεδα, και αραιές μέθοδοι αναπαράστασης. Η παρούσα διπλωματική εργασία επικεντρώνεται στην αναδόμηση δύο προηγμένων τεχνικών μεθόδων μείωσης της διάστασης, και συγκεκριμένα, στις προβολές διατήρησης της αραιότητας (SPP) και στη διακριτική αραιή ενσωμάτωση που διατηρεί τη γειτονιά (DSNPE). Αμφότερες, εμπίπτουν στην κατηγορία των αραιών μεθόδων αναπαράστασης. Οι μέθοδοι αυτές αξιολογούνται και συγκρίνονται με γνωστές γραμμικές και μη γραμμικές τεχνικές, όπως είναι η ανάλυση κύριων συνιστωσών (PCA), η γραμμική ανάλυση διακρίσεων (LDA), οι προβολές διατήρησης τοπικότητας (LPP) και η τοπικά γραμμική ενσωμάτωση (LLE), χρησιμοποιώντας τα σύνολα δεδομένων εικόνων προσώπων Yale, Extended Yale B και AR. Η συγκεκριμένη ανάλυση βαθαίνει στην κατανόηση της μείωσης διαστάσεων στην έρευνα της μηχανικής μάθησης, με ιδιαίτερη έμφαση στην εξερεύνηση των τεχνικών αναπαράστασης λεπτομερειών.

Abstract

Dimensionality reduction methods in machine learning are powerful tools for condensing high-dimensional data into simpler, lower-dimensional formats, enhancing the capabilities of face and pattern recognition systems. These methods are pivotal in resolving the 'curse of dimensionality', a challenge of managing complex, high-dimensional data that can overwhelm computational resources and obfuscate meaningful patterns.

Broadly, these techniques fall into three categories: linear dimensionality reduction methods, nonlinear or manifold-based dimensionality reduction methods and sparse representation methods. This thesis focuses on reconstructing two advanced techniques for dimensionality reduction, sparsity preserving projections (SPP) and discriminant sparse neighborhood preserving embedding (DSNPE), both of which fall within the sparse representation category. These methods are evaluated and compared against well-known linear and nonlinear techniques, namely principal component analysis (PCA), linear discriminant analysis (LDA), locality preserving projections (LPP) and locally linear embedding (LLE), using the Yale, Extended Yale B and AR face image datasets. This analysis deepens the understanding of dimensionality reduction in machine learning research, with a special emphasis on the exploration of sparse representation techniques.

Study of manifold-based dimensionality reduction techniques for object classification

Αντώνης Παπακωνσταντίνου
tonypap@ece.auth.gr

Θεσσαλονίκη, Σεπτέμβριος 2023

Περιεχόμενα

1. Κεφάλαιο Εισαγωγή	13
1.1 Κίνητρο	13
1.1.1 Γραμμικές μέθοδοι	13
1.1.2 Μη γραμμικές μέθοδοι	14
1.2 Περιγραφή προβλήματος	14
1.3 Σκοπός της διπλωματικής	15
1.4 Διάρθρωση της διπλωματικής	15
2. Κεφάλαιο Μέθοδοι μείωσης διαστάσεων	17
2.1 Εισαγωγή	17
2.2 Γραμμικές μέθοδοι	17
2.2.1 Principal Component Analysis	17
2.2.2 Locality Preserving Projections	18
2.2.3 Linear Discriminant Analysis	19
2.3 Μη γραμμικές μέθοδοι	21
2.3.1 Locally Linear Embedding	21
3. Κεφάλαιο Αραιή αναπαράσταση	25
3.1 Εισαγωγή	25
3.2 Αραιή αναπαράσταση (Sparse representation)	25
4. Κεφάλαιο Επισκόπηση ερευνητικής περιοχής – (Υλοποίηση)	29
4.1 Προβολές Διατήρησης Αραιότητας (SPP)	29
4.1.1 Ανακατασκευή αραιών βαρών	30
4.1.2 Διατήρηση αραιών βαρών ανακατασκευής	32
4.1.3 SPP αλγόριθμος	34
4.2 Διακριτική αραιή ενσωμάτωση που διατηρεί τη γειτονιά (DSNPE)	34
4.2.1 Κριτήριο μεγιστοποίησης περιθωρίου (MMC)	35
4.2.2 DSNPE αλγόριθμος	36
4.2.3 Ανάλυση πολυπλοκότητας χρόνου	38
4.2.4 Θεωρητική ανάλυση του DSNPE αλγορίθμου	38
5. Κεφάλαιο Πειραματικά αποτελέσματα	41
5.1 Το πρόβλημα	41
5.2 Σύνολο δεδομένων	41
5.2.1 Yale	41
5.2.2 Extended Yale B	42
5.2.3 AR	42

5.2.4	Διαχωρισμός δεδομένων σε εκπαίδευση και δοκιμή	43
5.3	Ταξινόμηση	43
5.3.1	Ταξινομητής 1-NN (1-πλησιέστερος γείτονας)	44
5.3.2	Ταξινομητής k-NN (k-πλησιέστεροι γείτονες)	45
5.3.3	Ταξινομητής SVM (Μηχανή διανυσμάτων στήριξης)	45
5.4	Μοντέλα επίλυσης	46
5.4.1	Επιλογή παραμέτρων	48
5.4.2	Λεπτομέρειες υλοποίησης – Λογισμικό και βιβλιοθήκες	49
5.4.3	Εκπαίδευση μοντέλων	49
5.5	Αποτελέσματα	50
5.5.1	Βάση (baseline)	51
5.5.2	Ανάλυση κύριων συνιστωσών (PCA)	53
5.5.3	Γραμμική διακριτική ανάλυση (LDA)	56
5.5.4	Προβολές διατήρησης τοπικότητας (LPP)	59
5.5.5	Τοπικά γραμμική ενσωμάτωση (LLE)	61
5.5.6	Προβολές διατήρησης αραιότητας (SPP)	66
5.5.7	Διακριτική αραιή ενσωμάτωση που διατηρεί τη γειτονιά (DSNPE)	68
5.5.8	Σύγκριση των μεθόδων	71
5.5.9	Συνολικές παρατηρήσεις και συζήτηση των παραπάνω πειραματικών αποτελεσμάτων	75
6.	Κεφάλαιο Συμπεράσματα και μελλοντική εργασία	77
6.1	Γενικά συμπεράσματα	77
6.2	Προβλήματα	77
6.3	Μελλοντικές επεκτάσεις	78
7.	Βιβλιογραφία	80

Λίστα εικόνων

Εικόνα 1.	Απεικονίζονται δεδομένα τριών διαστάσεων που με χρήση PCA μειώνονται οι διαστάσεις στις δυο.	17
Εικόνα 2.	Απεικόνιση Swiss roll dataset και το αποτέλεσμα μετά από εφαρμογή του LPP. .	18
Εικόνα 3.	Απεικόνιση του LDA.	20
Εικόνα 4.	Το LLE εφαρμοσμένο στο σύνολο δεδομένων S curve.	21
Εικόνα 5.	Απεικόνιση του αλγορίθμου LLE.	22
Εικόνα 6.	Τα βάρη ανακατασκευής του LLE.	23
Εικόνα 7.	Ένα 2D παράδειγμα βελτιστοποίησης υπό ℓ_2 ελαχιστοποίηση (αριστερά) και ℓ_1 ελαχιστοποίηση (δεξιά).	26
Εικόνα 8.	Παράδειγμα δεδομένων που έχουν κρυμμένα χαρακτηριστικά και που είναι κατεστραμμένα.	27
Εικόνα 9.	Ένα υπο-μπλοκ του πίνακα βαρών S για τον SPP.	31
Εικόνα 10.	Τριάντα δύο περικομμένα δείγματα εικόνων ενός ατόμου, Yale.	42

Εικόνα 11. Τριάντα δύο περικομμένα δείγματα εικόνων ενός ατόμου, Extended Yale B.	42
Εικόνα 12. Ορισμένα χαρακτηριστικά δείγματα των περικομμένων εικόνων που βρίσκονται στη βάση δεδομένων εικόνων προσώπων AR.	43
Εικόνα 13. Παράδειγμα χρησιμοποιώντας τον 1-NN και τον k-NN ταξινομητή.....	44
Εικόνα 14. Τρόπος λειτουργίας SVM.	46
Εικόνα 15. Απόδοση SPP και DSNPE στη βάση δεδομένων Yale.....	50
Εικόνα 16. Απόδοση baseline στη βάση δεδομένων Yale.	52
Εικόνα 17. Απόδοση baseline στη βάση δεδομένων Extended Yale B.	52
Εικόνα 18. Απόδοση baseline στη βάση δεδομένων AR fixed.....	53
Εικόνα 19. Απόδοση baseline στη βάση δεδομένων AR random.	53
Εικόνα 20. Γράφημα απόδοσης – διάστασης, στο Yale χρησιμοποιώντας την μέθοδο PCA.	54
Εικόνα 21. Γράφημα απόδοσης – διάστασης, στο Extended Yale B χρησιμοποιώντας την μέθοδο PCA.....	55
Εικόνα 22. Γράφημα απόδοσης – διάστασης, στο AR fixed χρησιμοποιώντας την μέθοδο PCA.....	55
Εικόνα 23. Γράφημα απόδοσης – διάστασης, στο AR random χρησιμοποιώντας την μέθοδο PCA.	56
Εικόνα 24. Γράφημα απόδοσης – διάστασης, στο Yale χρησιμοποιώντας την μέθοδο LDA.	57
Εικόνα 25. Γράφημα απόδοσης – διάστασης, στο Extended Yale B χρησιμοποιώντας την μέθοδο LDA.....	57
Εικόνα 26. Γράφημα απόδοσης – διάστασης, στο AR fixed χρησιμοποιώντας την μέθοδο LDA.	58
Εικόνα 27. Γράφημα απόδοσης – διάστασης, στο AR random χρησιμοποιώντας την μέθοδο PCA.	58
Εικόνα 28. Γράφημα απόδοσης – διάστασης, στο Yale χρησιμοποιώντας την μέθοδο LPP..	59
Εικόνα 29. Γράφημα απόδοσης – διάστασης, στο Extended Yale B χρησιμοποιώντας την μέθοδο LPP.	60
Εικόνα 30. Γράφημα απόδοσης – διάστασης, στο AR fixed χρησιμοποιώντας την μέθοδο LPP.....	61
Εικόνα 31. Γράφημα απόδοσης – διάστασης, στο AR random χρησιμοποιώντας την μέθοδο LPP.....	61
Εικόνα 32. Γράφημα απόδοσης – γείτονα, στο Yale χρησιμοποιώντας την μέθοδο LLE για διάσταση 50.....	62
Εικόνα 33. Γράφημα απόδοσης – διάστασης, στο Yale χρησιμοποιώντας την μέθοδο LLE για 10 γείτονες.	63
Εικόνα 34. Γράφημα απόδοσης – διάστασης, στο Yale χρησιμοποιώντας την μέθοδο LLE για 14 γείτονες.	63
Εικόνα 35. Γράφημα απόδοσης – διάστασης, στο Extended Yale B χρησιμοποιώντας την μέθοδο LLE για 15 γείτονες.	64
Εικόνα 36. Γράφημα απόδοσης – διάστασης, στο AR fixed χρησιμοποιώντας την μέθοδο LLE για 15 γείτονες.	65
Εικόνα 37. Γράφημα απόδοσης – διάστασης, στο AR random χρησιμοποιώντας την μέθοδο LLE για 15 γείτονες.....	65
Εικόνα 38. Γράφημα απόδοσης – διάστασης, στο Yale χρησιμοποιώντας την μέθοδο SPP.	66
Εικόνα 39. Γράφημα απόδοσης – διάστασης, στο Extended Yale B χρησιμοποιώντας την μέθοδο SPP.	67

Εικόνα 40. Γράφημα απόδοσης – διάστασης, στο AR fixed χρησιμοποιώντας την μέθοδο SPP.....	67
Εικόνα 41. Γράφημα απόδοσης – διάστασης, στο AR random χρησιμοποιώντας την μέθοδο SPP.....	68
Εικόνα 42. Γράφημα απόδοσης – διάστασης, στο Yale χρησιμοποιώντας την μέθοδο DSNPE.	69
Εικόνα 43. Γράφημα απόδοσης – διάστασης, στο Extended Yale B χρησιμοποιώντας την μέθοδο DSNPE.	69
Εικόνα 44. Γράφημα απόδοσης – διάστασης, στο AR fixed χρησιμοποιώντας την μέθοδο DSNPE.....	70
Εικόνα 45. Γράφημα απόδοσης – διάστασης, στο AR random χρησιμοποιώντας την μέθοδο DSNPE.....	70
Εικόνα 46. Γραφήματα απόδοσης – διάστασης, σύγκρισης όλων των μεθόδων στη βάση δεδομένων Yale.	71
Εικόνα 47. Γραφήματα απόδοσης – διάστασης, σύγκρισης όλων των μεθόδων στη βάση δεδομένων Extended Yale B.	72
Εικόνα 48. Γραφήματα απόδοσης – διάστασης, σύγκρισης όλων των μεθόδων στη βάση δεδομένων AR fixed.....	74
Εικόνα 49. Γραφήματα απόδοσης – διάστασης, σύγκρισης όλων των μεθόδων στη βάση δεδομένων AR random.....	75

Λίστα πινάκων

Πίνακας 1. Μέγιστα ποσοστά αναγνώρισης στο Yale για κάθε ταξινομητή και οι αντίστοιχες διαστάσεις.	72
Πίνακας 2. Μέγιστα ποσοστά αναγνώρισης στο Extended Yale B για κάθε ταξινομητή και οι αντίστοιχες διαστάσεις.....	73
Πίνακας 3. Μέγιστα ποσοστά αναγνώρισης στο AR fixed για κάθε ταξινομητή και οι αντίστοιχες διαστάσεις.....	73
Πίνακας 4. Μέγιστα ποσοστά αναγνώρισης στο AR random για κάθε ταξινομητή και οι αντίστοιχες διαστάσεις.....	74

1. Κεφάλαιο

Εισαγωγή

1.1 Κίνητρο

Το κίνητρο πίσω από τις μεθόδους μείωσης διαστάσεων είναι να κάνει τα δεδομένα υψηλής διάστασης πιο διαχειρίσιμα για επεξεργασία και ανάλυση. Με τη μείωση του αριθμού των χαρακτηριστικών ή διαστάσεων, αυτές οι μέθοδοι επιτρέπουν πιο αποδοτικό υπολογισμό, παρέχουν τη δυνατότητα για σημασιολογική οπτικοποίηση, διευκολύνουν την αποθήκευση δεδομένων και βελτιώνουν την ερμηνευσιμότητα χωρίς να χάνουν σημαντικές πληροφορίες. Κατ' ουσίαν, η μείωση των διαστάσεων λειτουργεί ως μέσο για τη μετάφραση πολύπλοκων, πολυδιάστατων δεδομένων σε μια μορφή που διατηρεί τα βασικά χαρακτηριστικά, καθιστώντας τα πιο προσβάσιμα και εφαρμόσιμα σε διάφορους τομείς, όπως τη μηχανική μάθηση (machine learning), την ανάλυση δεδομένων (data analysis) και την οπτικοποίηση (visualization). Για να κατανοήσουμε πώς επιτυγχάνονται αυτοί οι στόχοι, είναι απαραίτητο να διερευνήσουμε τις δύο κύριες κατηγορίες μεθόδων μείωσης διαστάσεων:

- τις γραμμικές μεθόδους (linear methods)
- τις μη γραμμικές ή τις μεθόδους βασισμένες σε πολυεπίπεδα (nonlinear methods or manifold-based methods)

1.1.1 Γραμμικές μέθοδοι

Οι γραμμικές μέθοδοι εφαρμόζουν γραμμικούς μετασχηματισμούς για να μειώσουν τις διαστάσεις των δεδομένων. Οι τεχνικές αυτές προϋποθέτουν ότι η βασική δομή των δεδομένων μπορεί να αποτυπωθεί χρησιμοποιώντας γραμμικές σχέσεις μεταξύ των χαρακτηριστικών.

Γραμμικές μέθοδοι είναι:

- **PCA (Ανάλυση κύριων συνιστωσών):** Μεγιστοποιεί τη διασπορά και μειώνει τις διαστάσεις βρίσκοντας τις ορθογώνιες κατευθύνσεις υψηλότερης διασποράς.
- **LDA (Γραμμική διακριτική ανάλυση):** Επικεντρώνεται στη διαχωριστικότητα των κλάσεων, βρίσκοντας το γραμμικό συνδυασμό των χαρακτηριστικών που διαχωρίζει καλύτερα τις διαφορετικές κλάσεις.
- **LPP (Προβολές διατήρησης τοπικότητας):** Διατηρεί την τοπική δομή των δεδομένων, διατηρώντας τις σχέσεις μεταξύ γειτονικών σημείων δεδομένων.

1.1.2 Μη γραμμικές μέθοδοι

Οι μη γραμμικές μέθοδοι χρησιμοποιούνται όταν οι σχέσεις μεταξύ των μεταβλητών είναι σύνθετες και δεν μπορούν να αποτυπωθούν με ακρίβεια με γραμμικούς μετασχηματισμούς.

Μη γραμμικές μέθοδοι είναι:

- **LLE (Τοπικά γραμμική ενσωμάτωση):** Επικεντρώνεται στη διατήρηση των τοπικών σχέσεων μεταξύ των σημείων δεδομένων, επιτρέποντας την αναπαράσταση των μη γραμμικών δομών.
- **SPP (Προβολές διατήρησης αραιότητας):** Χρησιμοποιεί ιδιότητες αραιότητας για να διατηρήσει τις σχετικές πληροφορίες κατά τον μετασχηματισμό.
- **DSNPE (Διακριτική αραιή ενσωμάτωση που διατηρεί τη γειτονιά):** Συνδυάζει τη διακριτική ανάλυση με τη διατήρηση αραιής γειτονιάς, αποτυπώνοντας σύνθετες σχέσεις και βελτιώνοντας τη διαχωριστικότητα των κλάσεων.

Με την κατηγοριοποίηση των τεχνικών μείωσης διάστασης σε γραμμικές και μη γραμμικές μεθόδους, γίνεται δυνατή η επιλογή της πλέον κατάλληλης προσέγγισης με βάση τη κύρια δομή και τις σχέσεις εντός των δεδομένων, επιτρέποντας μια πιο αποτελεσματική και διορατική ανάλυση.

1.2 Περιγραφή προβλήματος

Η μείωση των διαστάσεων αναφέρεται στη διαδικασία της μείωσης του αριθμού των τυχαίων μεταβλητών που εξετάζονται, μέσω της απόκτησης ενός συνόλου κύριων μεταβλητών. Μπορεί να θεωρηθεί ως απλοποίηση ενός συνόλου δεδομένων υψηλής διαστατικότητας σε μια μορφή χαμηλότερης διαστατικότητας, διατηρώντας όσο το δυνατόν περισσότερες σχετικές πληροφορίες.

Τα δεδομένα υψηλής διαστατικότητας, τα οποία αναφέρονται συχνά ως "η κατάρα της διαστατικότητας (curse of dimensionality)", μπορεί να οδηγήσουν σε αρκετά ζητήματα:

1. **Υψηλό κόστος υπολογισμού (Computationally Expensive):** Όσες περισσότερες διαστάσεις έχει ένα σύνολο δεδομένων, τόσο περισσότεροι υπολογιστικοί πόροι απαιτούνται για την επεξεργασία του.
2. **Δυσκολία στην οπτικοποίηση (Difficult to Visualize):** Τα δεδομένα υψηλής διαστατικότητας δεν μπορούν να οπτικοποιηθούν εύκολα, με αποτέλεσμα να μην είναι δυνατή η κατανόηση και η ερμηνεία.
3. **Υπερεκπαίδευση (Overfitting):** Η υψηλή διαστατικότητα μπορεί να οδηγήσει σε υπερεκπαίδευση, όπου το μοντέλο εκπαιδεύεται καλά στα δεδομένα εκπαίδευσης αλλά είναι κακό στα αόρατα δεδομένα.

4. **Αραιότητα δεδομένων** (Sparsity of Data): Καθώς η διαστατικότητα αυξάνεται, ο όγκος του χώρου αυξάνεται εκθετικά, και τα δεδομένα γίνονται αραιά, με αποτέλεσμα να οδηγούμαστε σε μη αξιόπιστα μοντέλα.

Αθροιστικά τα προβλήματα αυτά δημιουργούν ένα πολυεπίπεδο ζήτημα στη διαχείριση των δεδομένων υψηλής διάστασης κι αυτό το ουσιώδες θέμα στοχεύουν να αντιμετωπίσουν οι τεχνικές μείωσης διαστάσεων, καθιστώντας τέτοια δεδομένα πιο ευέλικτα σε αποτελεσματική ανάλυση, μοντελοποίηση και ερμηνεία.

1.3 Σκοπός της διπλωματικής

Η παρούσα διπλωματική εργασία στοχεύει να αναδομήσει και να εξετάσει κριτικά δύο συγκεκριμένους αλγορίθμους, το SPP και το DSNPE, και να τους συγκρίνει με άλλους καθιερωμένους αλγορίθμους όπως το PCA, το LDA, το LPP και το LLE. Η σύγκριση πραγματοποιείται εφαρμόζοντας αυτούς τους αλγορίθμους σε τρία διαφορετικά σύνολα δεδομένων προσώπων, τα YALE, το Extended YALE B και το AR. Μέσω της ακριβής αυτής ανάλυσης και του πειραματισμού, η διπλωματική στοχεύει να παράσχει ενδείξεις για τη σχετική αποτελεσματικότητα και την αποδοτικότητα αυτών των μεθόδων, συμβάλλοντας στην ευρύτερη κατανόηση των τεχνικών μείωσης διάστασης, στο πλαίσιο της αναγνώρισης και ανάλυσης προσώπων. Στο πλαίσιο της εργασίας αυτής αναπτύσσονται τόσο οι κώδικες των αλγορίθμων SPP και DSNPE, όσο και των υπόλοιπων απαραίτητων αλγορίθμων.

1.4 Διάρθρωση της διπλωματικής

Η διπλωματική εργασία ακολουθεί την εξής δομή :

- **Κεφάλαιο 2:** Στο κεφάλαιο αυτό γίνεται η περιγραφή των πιο γνωστών μεθόδων για τη μείωση των διαστάσεων, οι οποίες θα χρησιμοποιηθούν για τη σύγκριση με τα μοντέλα που κατασκευάσαμε σε αυτή την εργασία.
- **Κεφάλαιο 3:** Στο κεφάλαιο αυτό γίνεται η περιγραφή της αραιής αναπαράστασης που είναι ένα από τα εργαλεία που χρησιμοποιούμε για τη δημιουργία των δυο μεθόδων μας.
- **Κεφάλαιο 4:** Στο κεφάλαιο αυτό αναπτύσσεται όλο το θεωρητικό υπόβαθρο για την κατανόηση των αλγορίθμων που αναπτύχθηκαν στα πλαίσια της εργασίας.
- **Κεφάλαιο 5:** Στο κεφάλαιο αυτό παρουσιάζονται τα αποτελέσματα και επιβεβαιώνεται η ορθή λειτουργία του προτεινόμενου μοντέλου.
- **Κεφάλαιο 6:** Στο κεφάλαιο αυτό παρουσιάζονται τα συμπεράσματα, καθώς και κάποιες μελλοντικές επεκτάσεις για τη συνέχιση της παρούσας εργασίας.

2. Κεφάλαιο

Μέθοδοι μείωσης διαστάσεων

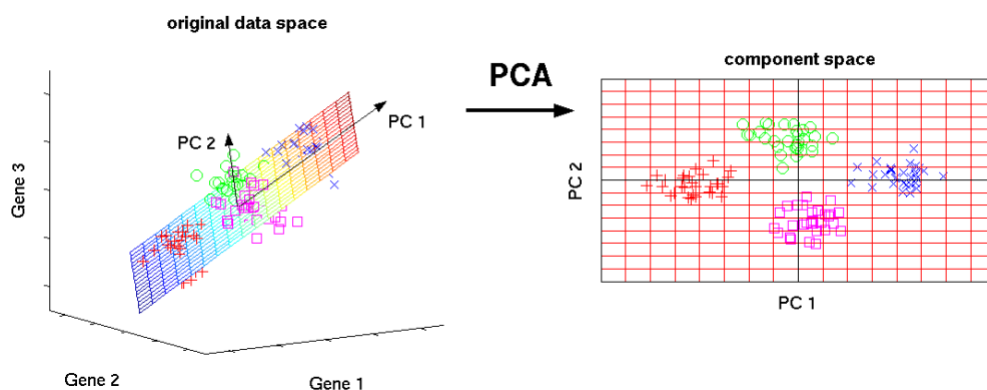
2.1 Εισαγωγή

Οι τεχνικές μείωσης διάστασης συμβάλλουν στην απλούστευση της πολυπλοκότητας των δεδομένων υψηλής διάστασης και είναι ζωτικά εργαλεία στην αναγνώριση προτύπων και στη μηχανική μάθηση. Αν και υπάρχουν πολλοί τρόποι μείωσης της διάστασης, αυτή η εργασία επικεντρώνεται κυρίως στις γραμμικές τεχνικές. Η μαθηματική τους απλότητα, αποτελεσματικότητα και αποδοτικότητα τις καθιστούν κατάλληλες για την επίλυση πολλών προβλημάτων του πραγματικού κόσμου, όπως είναι η αναγνώριση προσώπων. Αυτή η ενότητα θα εξερευνήσει μερικές από τις πιο ευρέως χρησιμοποιούμενες γραμμικές τεχνικές όπως το PCA, το LDA, και το LPP, μαζί με τη μη γραμμική μέθοδο LLE.

2.2 Γραμμικές μέθοδοι

Οι γραμμικές μέθοδοι μείωσης διάστασης προβάλλουν τα δεδομένα σε έναν χώρο μειωμένων διαστάσεων χρησιμοποιώντας ένα γραμμικό μετασχηματισμό. Η υπόθεση εδώ είναι ότι τα δεδομένα βρίσκονται, περίπου, σε έναν γραμμικό υπόχωρο.

2.2.1 Principal Component Analysis



Εικόνα 1. Απεικονίζονται δεδομένα τριών διαστάσεων που με χρήση PCA μειώνονται οι διαστάσεις στις δυο.

Το PCA αναζητά μια αναπαράσταση χαμηλής διάστασης των δεδομένων για να διατηρήσει όσο το δυνατόν περισσότερη από τη διακύμανση των δεδομένων. Δεδομένου ενός συνόλου σημείων δεδομένων $\{\mathbf{x}_i\}_{i=1}^n$, όπου $\mathbf{x}_i \in \mathbb{R}^m$ είναι ένας διάνυσμα στήλης m -διάστασης, περιμένουμε να πάρουμε τις εικόνες χαμηλής διάστασης τους $\{y_i\}_{i=1}^n$ προβάλλοντας κάθε \mathbf{x}_i στο διάνυσμα κατεύθυνσης $\mathbf{w} \in \mathbb{R}^m$. Η αντικειμενική συνάρτηση του PCA ορίζεται ως εξής:

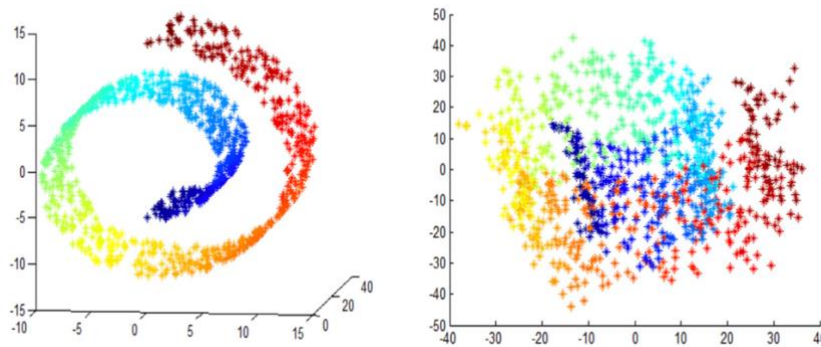
$$\max_{\|\mathbf{w}\|=1} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (1)$$

Όπου $y_i = \mathbf{w}^T \mathbf{x}_i$, και \bar{y} είναι ο μέσος όρος του $\{y_i\}_{i=1}^n$. Η εξίσωση (1) μπορεί να γραφτεί ξανά ως:

$$\max_{\|\mathbf{w}\|=1} \mathbf{w}^T \Sigma \mathbf{w} \quad (2)$$

όπου Σ είναι το δείγμα του πίνακα συσχέτισης. Τα ιδιοδιανύσματα του Σ που αντιστοιχούν στις μεγαλύτερες d ιδιοτιμές καλύπτουν τον βέλτιστο υπόχωρο του PCA. Στην αναγνώριση προσώπου, το \mathbf{x}_i αντιπροσωπεύει μια εικόνα προσώπου, και τα ιδιοδιανύσματα ονομάζονται ιδιοπρόσωπα (Eigenfaces).

2.2.2 Locality Preserving Projections



Εικόνα 2. Απεικόνιση του Swiss roll dataset και το αποτέλεσμα μετά από εφαρμογή του LPP.

Ενώ το PCA στοχεύει στη διατήρηση της συνολικής δομής των δεδομένων, το LPP στοχεύει στη διατήρηση της τοπικής (δηλαδή, γειτονικής) δομής των δεδομένων. Διαισθητικά, το LPP μπορεί να διατηρήσει περισσότερες πληροφορίες διάκρισης από το PCA, υποθέτοντας ότι τα δείγματα από την ίδια κλάση είναι πιθανώς κοντά το ένα στο άλλο στο χώρο εισόδου.

Χρησιμοποιώντας τα ίδια μαθηματικά, όπως στο PCA, η αντικειμενική συνάρτηση του LPP ορίζεται ως εξής:

$$\min_{\mathbf{w}} \sum_{ij} (y_i - y_j)^2 p_{ij} \quad (3)$$

Όπου $y_i = \mathbf{w}^T \mathbf{x}_i$, $i = 1, 2, \dots, n$, και $\mathbf{P} = (p_{ij})_{n \times n}$ είναι ένας πίνακας ομοιότητας που ορίζεται ως εξής:

$$p_{ij} = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / t), & \text{εάν το } \mathbf{x}_i \text{ βρίσκεται μεταξύ των } kNN \text{ του } \mathbf{x}_j \text{ ή } \mathbf{x}_i \\ 0, & \text{σε διαφορετική περίπτωση} \end{cases}$$

Η ελαχιστοποίηση της (3) στοχεύει στην ενθάρρυνση του γεγονότος ότι εάν δύο σημεία \mathbf{x}_i και \mathbf{x}_j είναι κοντά το ένα στο άλλο στον χώρο εισόδου, τότε θα πρέπει να είναι και στον αντίστοιχο χώρο εξόδου. Με απλή διατύπωση, η αντικειμενική συνάρτηση είναι ισοδύναμη με την ελαχιστοποίηση του:

$$\begin{aligned} \frac{1}{2} \sum_{ij} (y_i - y_j)^2 p_{ij} &= \frac{1}{2} \sum_{ij} (\mathbf{w}^T \mathbf{x}_i - \mathbf{w}^T \mathbf{x}_j)^2 p_{ij} \\ &= \mathbf{w}^T \mathbf{X}(\mathbf{D} - \mathbf{P})\mathbf{X}^T \mathbf{w} = \mathbf{w}^T \mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{w} \end{aligned} \quad (4)$$

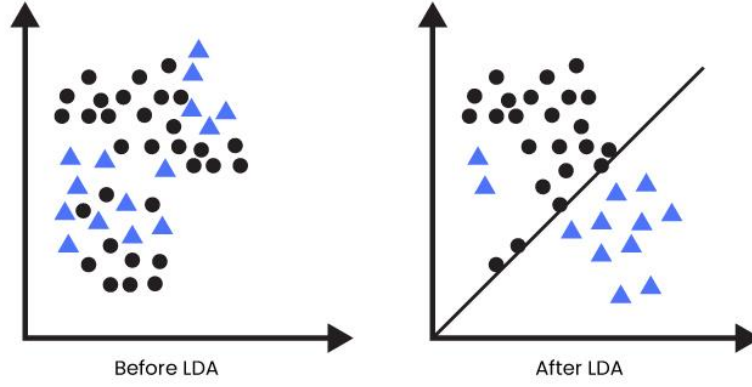
όπου \mathbf{D} είναι ένας διαγώνιος πίνακας με τις καταχωρήσεις του να είναι τα αθροίσματα της γραμμής (ή της στήλης, εφόσον το \mathbf{P} είναι συμμετρικό) του \mathbf{P} , δηλαδή $d_{ii} = \sum_j p_{ij}$, και $\mathbf{L} = \mathbf{D} - \mathbf{P}$ είναι ο λαπλασιανός πίνακας. Επιβάλλοντας τον περιορισμό $\mathbf{w}^T \mathbf{X}\mathbf{D}\mathbf{X}^T \mathbf{w} = 1$, η LPP μειώνεται σε:

$$\min_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{w}}{\mathbf{w}^T \mathbf{X}\mathbf{D}\mathbf{X}^T \mathbf{w}} \quad (5)$$

Το βέλτιστο \mathbf{w} δίνεται από την ελάχιστη ιδιοτιμή της λύσης στο ακόλουθο γενικευμένο πρόβλημα ιδιοτιμών:

$$\mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{w} = \lambda \mathbf{X}\mathbf{D}\mathbf{X}^T \mathbf{w}$$

2.2.3 Linear Discriminant Analysis



Εικόνα 3. Απεικόνιση του LDA.

Σε αυτή την ενότητα, θα δώσω μια σύντομη επισκόπηση του κλασικού LDA. Δεδομένου ενός συνόλου δεδομένων $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}^{D \times n}$, όπου το $\{\mathbf{x}_i\}_{i=1}^n$ αντιπροσωπεύει τα επισημασμένα δεδομένα. Το \mathbf{X} μπορεί να χωριστεί σε C ασύμφωρες κλάσεις και ο μέσος όρος του είναι το μηδενικό διάνυσμα. Το $y_i (y_i \in \{1, 2, \dots, C\})$ υποδηλώνει την ετικέτα του $\mathbf{x}_i (i = 1, \dots, n)$. Χωρίς απώλεια γενικότητας, χρησιμοποιούμε το σχήμα 1-από-Κ για να αντιπροσωπεύσουμε την ιδιότητα κλάσης των δειγμάτων, όπως, το $\mathbf{Y}_{ij} = 1$ υποδηλώνει το $y_i = j$, διαφορετικά, $\mathbf{Y}_{ij} = 0$. Επιπλέον, το \mathbf{Y} μπορεί να κλιμακωθεί από το $\mathbf{Y} = \mathbf{Y}(\mathbf{Y}^T \mathbf{Y})^{\frac{1}{2}}$. Το κλασικό LDA υπολογίζει μια γραμμική προβολή που χαρτογραφεί το \mathbf{x}_i σε έναν χώρο χαμηλής διάστασης: $\mathbf{W} : \tilde{\mathbf{x}}_i = \mathbf{W}^T \mathbf{x}_i$. Στο LDA, τρεις πίνακες διασποράς, που ονομάζονται εντός-κλάσης, μεταξύ-κλάσης και συνολικοί πίνακες διασποράς, ορίζονται ως:

$$\mathbf{S}_w = \frac{1}{n} \sum_{i=1}^C \sum_{y_j=i} (\mathbf{x}_j - \mathbf{m}_i)(\mathbf{x}_j - \mathbf{m}_i)^T \quad (6)$$

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{y_j=i} \mathbf{x}_j$$

$$\mathbf{S}_b = \sum_{i=1}^C \frac{n_i}{n} (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \quad (7)$$

$$\mathbf{S}_t = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T \quad (8)$$

όπου το n_i υποδηλώνει τον αριθμό των δειγμάτων στην κλάση i .

Από τον παραπάνω ορισμό, γνωρίζουμε ότι $\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_w$. Εάν τα δεδομένα έχουν κεντραριστεί αφαιρώντας το μέσο όρο τους, οι μεταξύ-κλάσης και συνολικοί πίνακες διασποράς μπορούν να κατασκευαστούν ως εξής:

$$\mathbf{S}_b = \sum_{i=1}^c \frac{n_i}{n} \mathbf{m}_i \mathbf{m}_i^T = \frac{1}{n} \mathbf{X} \mathbf{Y} \mathbf{Y}^T \mathbf{X}^T \quad (9)$$

$$\mathbf{S}_t = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T = \frac{1}{n} \mathbf{X} \mathbf{X}^T \quad (10)$$

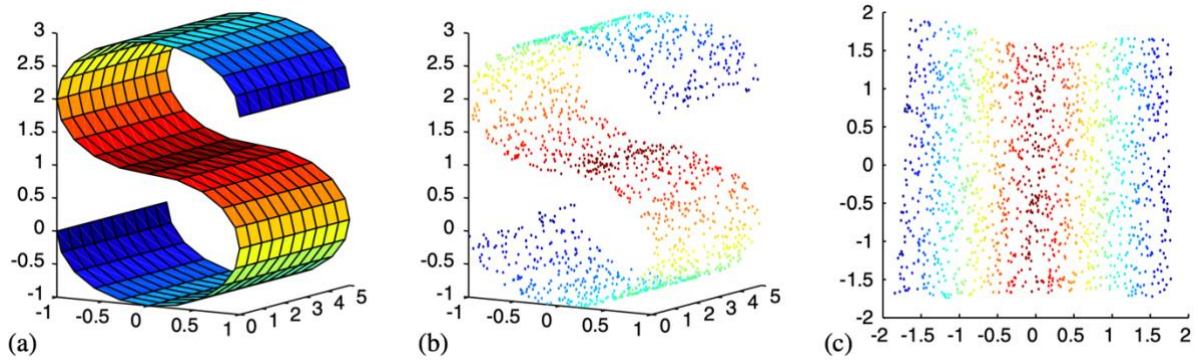
Ο βέλτιστος μετασχηματισμός του LDA υπολογίζεται λύνοντας το ακόλουθο πρόβλημα:

$$\mathbf{W} = \max_{\mathbf{W}} \frac{\text{tr}\{\mathbf{W}^T \mathbf{S}_b \mathbf{W}\}}{\text{tr}\{\mathbf{W}^T \mathbf{S}_t \mathbf{W}\}} \quad (11)$$

2.3 Μη γραμμικές μέθοδοι

Οι μη γραμμικές τεχνικές μείωσης διάστασης, ή αλλιώς μέθοδοι βασισμένες σε πολυεπίπεδα, από την άλλη πλευρά, μπορούν να αντιμετωπίσουν δεδομένα που δεν βρίσκονται σε γραμμικό υπόχωρο. Αυτές οι μέθοδοι μπορούν να αποτυπώσουν πιο πολύπλοκες σχέσεις εντός των δεδομένων και να αναπαραστήσουν μη γραμμικά πολυεπίπεδα.

2.3.1 Locally Linear Embedding

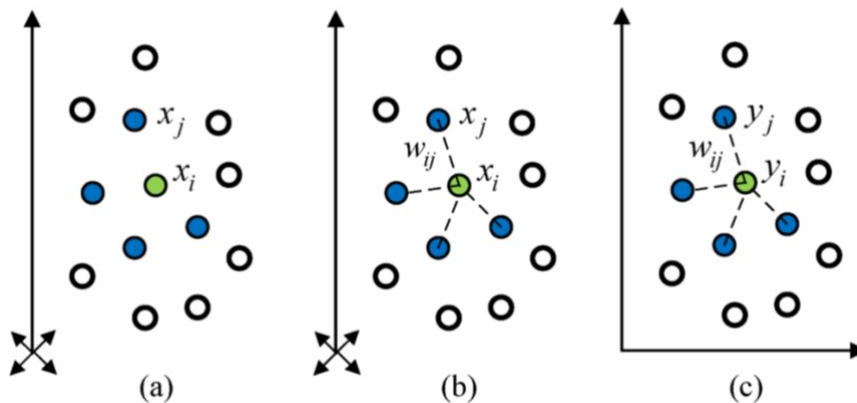


Εικόνα 4. Το LLE εφαρμοσμένο στο σύνολο δεδομένων S curve: (a) Το manifold S curve στο \mathbb{R}^3 ; (b) 1.500 σημεία δεδομένων που επιλέγονται τυχαία από το manifold; (c) Το αποτέλεσμα του LLE για τον χώρο ενσωμάτωσης στο \mathbb{R}^2 .

Ας έχουμε $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ να είναι ένα σύνολο N σημείων σε έναν χώρο δεδομένων υψηλής διάστασης R^D . Τα σημεία δεδομένων υποθέτεται ότι βρίσκονται πάνω ή κοντά σε ένα μη γραμμικό πολυεπίπεδο ενδογενούς διαστασιμότητας $d < D$ (συνήθως $d \ll D$). Υπό την προϋπόθεση ότι υπάρχουν επαρκή δεδομένα διαθέσιμα με καλή δειγματοληψία από το πολυεπίπεδο, ο στόχος της LLE είναι να βρει μια ενσωμάτωση χαμηλής διάστασης του \mathbf{X} απεικονίζοντας τα δεδομένα D -διάστασης σε ένα μοναδικό παγκόσμιο σύστημα συντεταγμένων στο R^d . Ας υποδηλώσουμε το αντίστοιχο σύνολο των N σημείων στον χώρο ενσωμάτωσης R^d με $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$.

Ο αλγόριθμος LLE μπορεί να συνοψιστεί ως εξής:

- 1) Για κάθε σημείο δεδομένων $\mathbf{x}_i \in \mathbf{X}$:
 - a) Βρίσκει το σύνολο N_i των K πλησιέστερων γειτόνων του \mathbf{x}_i .
 - b) Υπολογίζει τα βάρη ανακατασκευής των γειτόνων που ελαχιστοποιούν το σφάλμα ανακατασκευής του \mathbf{x}_i .
- 2) Υπολογίζει την ενσωμάτωση χαμηλής διάστασης \mathbf{Y} για το \mathbf{X} που διατηρεί καλύτερα την τοπική γεωμετρία που αντιπροσωπεύεται από τα βάρη ανακατασκευής.



Εικόνα 5. Απεικόνιση του αλγορίθμου LLE: (a) Επιλογή γειτόνων, (b) Ανακατασκευή με ενσωματωμένα γραμμικά βάρη, (c) Χαρτογράφηση σε συντεταγμένες.

Το βήμα (1)(a) γίνεται συνήθως χρησιμοποιώντας την ευκλείδεια απόσταση για να ορίσει τη γειτονιά, αν και μπορεί να χρησιμοποιηθούν και πιο περίπλοκα κριτήρια.

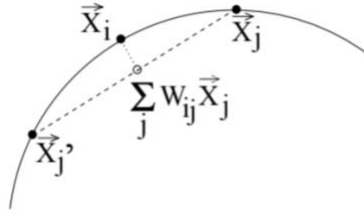
Βασισμένο στους K πλησιέστερους γείτονες που προσδιορίστηκαν, το βήμα (1)(b) αναζητά να βρει τα καλύτερα βάρη ανακατασκευής.

Η βελτιστοποίηση επιτυγχάνεται ελαχιστοποιώντας το τοπικό σφάλμα ανακατασκευής για το \mathbf{x}_i :

$$\mathbf{E}_i = \left\| \mathbf{x}_i - \sum_{\mathbf{x}_j \in N_i} \mathbf{w}_{ij} \mathbf{x}_j \right\|^2 \quad (12)$$

το οποίο είναι η τετραγωνική απόσταση μεταξύ του \mathbf{x}_i και της ανακατασκευής του, υπό τους περιορισμούς:

$$\sum_{\mathbf{x}_j \in N_i} \mathbf{w}_{ij} = 1 \text{ και } \mathbf{w}_{ij} = 0 \text{ για οποιοδήποτε } \mathbf{x}_j \notin N_i$$



Εικόνα 6. Ένα σημείο δεδομένων \mathbf{x}_i , οι γείτονές του \mathbf{x}_j , και η τοπικά γραμμική ανακατασκευή του $\sum_j w_{ij} \mathbf{x}_j$. Τα βάρη ανακατασκευής περιορίζονται ώστε να ικανοποιούν τη σχέση $\sum_j w_{ij} = 1$.

Η ελαχιστοποίηση του \mathbf{E}_i υπό τους περιορισμούς είναι ένα πρόβλημα ελαχιστοποίησης ελαχίστων τετραγώνων με περιορισμούς. Μετά την επανάληψη των βημάτων (1)(a) και (1)(b) για όλα τα N σημεία δεδομένων στο \mathbf{X} , τα βάρη ανακατασκευής που προκύπτουν σχηματίζουν έναν πίνακα βαρών $\mathbf{W} = [\mathbf{w}_{ij}]_{N \times N}$. Το βήμα (2) του αλγορίθμου LLE είναι να υπολογίσει την καλύτερη ενσωμάτωση χαμηλής διάστασης \mathbf{Y} βασισμένη στον πίνακα βαρών \mathbf{W} που προκύπτει. Αυτό αντιστοιχεί στην ελαχιστοποίηση της ακόλουθης συνάρτησης κόστους:

$$\Phi = \sum_{i=1}^N \left\| \mathbf{y}_i - \sum_{\mathbf{x}_j \in N_i} \mathbf{w}_{ij} \mathbf{y}_j \right\|^2 \quad (13)$$

υπό τους περιορισμούς:

$$\sum_{i=1}^N \mathbf{y}_i = \mathbf{0}$$

και:

$$\frac{1}{N} \sum_{i=1}^N \mathbf{y}_i \mathbf{y}_i^T = \mathbf{I}$$

όπου $\mathbf{0}$ είναι ένα μηδενικό διάνυσμα στήλης και \mathbf{I} είναι ο μοναδιαίος πίνακας, παρατηρείται μια ομοιότητα μεταξύ των εξισώσεων (12) και (13). Βασισμένοι στο \mathbf{W} ,

μπορούμε να ορίσουμε έναν αραιό, συμμετρικό και θετικά ημικαθορισμένο πίνακα \mathbf{M} ως εξής:

$$\mathbf{M} = (\mathbf{I} - \mathbf{W})^T(\mathbf{I} - \mathbf{W})$$

Είναι σημαντικό να σημειωθεί ότι το (13) μπορεί να εκφραστεί και στην τετραγωνική μορφή, $\Phi = \sum_{i,j} \mathbf{M}_{ij} \mathbf{y}_i^T \mathbf{y}_j$, βασισμένο στο $\mathbf{M} = [\mathbf{M}_{ij}]_{N \times N}$. Με την χρήση του θεωρήματος Rayleigh–Ritz, μια μέθοδο για την επίλυση του προβλήματος της ιδιοτιμής, η ελαχιστοποίηση του (13) ως προς τα \mathbf{y}_i στο \mathbf{Y} μπορεί να γίνει βρίσκοντας τα ιδιοδιανύσματα με τις μικρότερες (μη μηδενικές) ιδιοτιμές.

3. Κεφάλαιο

Αραιή αναπαράσταση

3.1 Εισαγωγή

Τα τελευταία χρόνια, ορισμένοι ερευνητές έχουν καταδείξει ότι οι περισσότερες από τις υπάρχουσες μεθόδους μείωσης διαστάσεων μπορούν να εξηγηθούν από την οπτική του πυρήνα (kernel)[27] και να ενοποιηθούν κάτω από ένα framework γράφου [25], όπου η κατασκευή ενός συγκεκριμένου γράφου και του πίνακα βάρους συναφείας (affinity weight matrix) παίζει καίριο ρόλο. Ωστόσο, σύμφωνα με το διάσημο θεώρημα "No Free Lunch" [29], δεν υπάρχει σαφής απόδειξη ότι οποιοσδήποτε πίνακας βάρους συναφείας είναι πάντα ανώτερος από τους άλλους, εξάλλου, οι πίνακες βάρους στις περισσότερες μεθόδους μείωσης διαστάσεων που βασίζονται στην τοπικότητα, όπως το LPP, έχουν ένα κοινό χαρακτηριστικό: την αραιότητα (sparsity)[17]. Η αραιότητα είναι ένας σημαντικός τρόπος κωδικοποίησης της γνώσης του πεδίου (domain knowledge) και βοηθά στη βελτίωση της ικανότητας γενίκευσης του μοντέλου. Με βάση αυτό, εδώ παρουσιάζουμε δυο μεθόδους για τον σχεδιασμό του πίνακα βάρους απευθείας βασισμένη στη θεωρία αραιής αναπαράστασης [35,36,41], μέσω της οποίας η αραιότητα μπορεί να προκύψει βέλτιστα και φυσικά. Συνοπτικά εξετάζουμε πρώτα την έννοια της αραιής αναπαράστασης και στο επόμενο κεφάλαιο θα εξετάσουμε τις λεπτομέρειες των δύο μεθόδων SPP και DSNPE.

3.2 Αραιή αναπαράσταση (Sparse representation)

Η αραιή αναπαράσταση (SR) προτείνεται αρχικά ως επέκταση των παραδοσιακών αναπαραστάσεων σήματος, όπως οι αναπαραστάσεις Fourier και wavelet. Τα τελευταία χρόνια, η αραιή αναπαράσταση έχει εφαρμοστεί με επιτυχία για την επίλυση πολλών πρακτικών προβλημάτων στην επεξεργασία σημάτων, τη στατιστική και την αναγνώριση προτύπων. Για παράδειγμα, στους τομείς της επεξεργασίας σημάτων και εικόνων, η SR χρησιμοποιείται για τη συμπίεση και την κωδικοποίηση σημάτων [30], την αποθορυβοποίηση εικόνων [31] και την υπερ-ανάλυση εικόνων [32]. Στη στατιστική, η SR είναι ένα αποτελεσματικό εργαλείο για την επιλογή μεταβλητών και πολλές φορές συνδέεται με το δημοφιλές LASSO [33,34]. Στην μηχανική μάθηση και στην αναγνώριση προτύπων, η SR χρησιμοποιείται για την ανίχνευση και ταξινόμηση αντικειμένων [22,23]. Στον αναδυόμενο τομέα του compressive sensing [35,36], και συγκρίνοντας το θεώρημα δειγματοληψίας Shannon-Nyquist, η SR επιδιώκει να ανακτήσει το σήμα με τον πιο αξιόπιστο τρόπο από άποψη οικονομίας χρόνου. Ειδικά, έρευνες [21,37] κατέδειξαν ότι ο

ταξινομητής βασισμένος στη SR είναι εξαιρετικά αποτελεσματικός και επιτυγχάνει το καλύτερο ποσοστό αναγνώρισης σε ορισμένες βάσεις δεδομένων προσώπων.

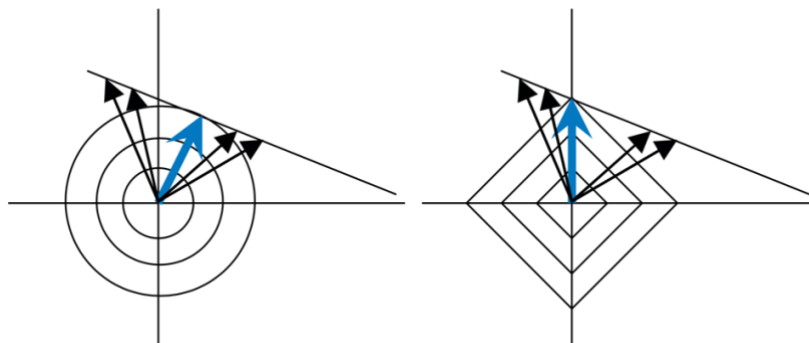
Η SR έχει συμπυκνωμένη μαθηματική έκφραση. Δεδομένου ενός σήματος (ή μιας εικόνας με διάταξη διανύσματος) $\mathbf{x} \in R^m$, και ενός πίνακα $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in R^{m \times n}$ που περιέχει τα στοιχεία ενός υπερπλήρους λεξικού [28] στις στήλες του, ο στόχος της SR είναι να αναπαραστήσει το \mathbf{x} χρησιμοποιώντας όσο το δυνατόν λιγότερες καταχωρήσεις του \mathbf{X} . Αυτό μπορεί να εκφραστεί επίσης ως εξής:

$$\min_{\mathbf{s}} \|\mathbf{s}\|_0 \text{ subject to } \mathbf{x} = \mathbf{X}\mathbf{s} \quad (14)$$

όπου $\mathbf{s} \in R^n$ είναι το διάνυσμα συντελεστών, και $\|\mathbf{s}\|_0$ είναι η ψευδό- ℓ_0 νόρμα, η οποία είναι ίση με τον αριθμό των μη μηδενικών συνιστωσών στο \mathbf{s} . Δυστυχώς, αυτό το κριτήριο δεν είναι κυρτό, και η εύρεση της πιο αραιής λύσης της εξίσωσης (14) είναι NP-hard. Το NP-hard αναφέρεται σε προβλήματα που είναι τουλάχιστον τόσο δύσκολα όσο τα δυσκολότερα προβλήματα στην κλάση NP, και δεν μπορούν να λυθούν σε πολυωνυμικό χρόνο. Αυτή η δυσκολία μπορεί να παρακαμφθεί με την μετατροπή του προβλήματος σε κυρτό και με την επίλυση:

$$\min_{\mathbf{s}} \|\mathbf{s}\|_1 \text{ subject to } \mathbf{x} = \mathbf{X}\mathbf{s} \quad (15)$$

όπου χρησιμοποιείται το ℓ_1 αντί για το ℓ_0 . Μπορεί να αποδειχθεί ότι, εάν η λύση \mathbf{s}^0 που αναζητείται είναι αρκετά αραιή, η λύση του προβλήματος ελαχιστοποίησης ℓ_0 είναι ίση με τη λύση του προβλήματος ελαχιστοποίησης ℓ_1 [35,36]. Η Εικόνα 7 δείχνει ότι η αραιή λύση μπορεί να βρεθεί επιλύοντας ένα πρόβλημα ελαχιστοποίησης ℓ_1 , αλλά μπορεί να μην βρεθεί με άλλες παραδοσιακές στρατηγικές όπως η ελαχιστοποίηση ℓ_2 .



Εικόνα 7. Ένα 2D παράδειγμα βελτιστοποίησης υπό ℓ_2 ελαχιστοποίηση (αριστερά) και ℓ_1 ελαχιστοποίηση (δεξιά). Η λοξή γραμμή δηλώνει τον εφικτό χώρο λύσης, δηλαδή $\{\mathbf{s} \in R^2 \mid \mathbf{x} = \mathbf{X}\mathbf{s}\}$ στην περίπτωση 2D. Τα δύο έντονα βέλη δηλώνουν τις βέλτιστες λύσεις των προβλημάτων ελαχιστοποίησης ℓ_2 και ℓ_1 , αντίστοιχα.

Στην πραγματικότητα, μη βέλτιστες λύσεις μπορούν να βρεθούν με μια ποικιλία προσεγγίσεων, όπως είναι οι άπληστοι αλγόριθμοι (greedy algorithms)[38] και οι Bayesian στρατηγικές[39]. Ωστόσο, η ισοδυναμία του προβλήματος ℓ_0 και ℓ_1 έχει μελετηθεί βαθιά από μαθηματική οπτική γωνία, και καθιστά την ℓ_1 προσέγγιση πιο αξιόπιστη από άλλες για πρακτικές εφαρμογές. Γενικά, το πρόβλημα ελαχιστοποίησης ℓ_1 μπορεί να επιλυθεί με τυπικό γραμμικό προγραμματισμό (Linear Programming)[41].

Σε πολλά πρακτικά προβλήματα, το σήμα \mathbf{x} έχει γενικά αρκετό θόρυβο, οπότε ο περιορισμός $\mathbf{x} = \mathbf{X}\mathbf{s}$ στην εξίσωση (15) δεν ισχύει πάντα. Σύμφωνα με το [21], τουλάχιστον δύο ανθεκτικές επεκτάσεις μπορούν να χρησιμοποιηθούν για να χειριστούν αυτό το πρόβλημα:

- 1) χαλαρώνουμε τον περιορισμό σε $\|\mathbf{x} - \mathbf{X}\mathbf{s}\| < \varepsilon$, όπου ε μπορεί να θεωρηθεί ως η ανοχή σφάλματος,
- 2) απλώς αντικαθιστούμε το \mathbf{X} με $[\mathbf{X}, \mathbf{I}]$, όπου \mathbf{I} είναι ο μοναδιαίος πίνακας m -τάξης.

Και οι δύο στρατηγικές λαμβάνονται υπόψη σε αυτήν τη διπλωματική. Η δεύτερη στρατηγική χρησιμοποιείται κυρίως για την αντιμετώπιση δεδομένων που είτε είναι κατεστραμμένα, είτε κάποια χαρακτηριστικά είναι κρυμμένα (corrupted - occluded)[21].



Εικόνα 8. Παράδειγμα δεδομένων που έχουν κρυμμένα χαρακτηριστικά και που είναι κατεστραμμένα. Στην πρώτη εικόνα είναι κρυμμένα τα μάτια με γυαλιά και στην δεύτερη εικόνα έχουμε θόρυβο που έχει καταστρέψει εντελώς τα χαρακτηριστικά.

4. Κεφάλαιο

Επισκόπηση ερευνητικής περιοχής – (Υλοποίηση)

4.1 Προβολές Διατήρησης Αραιότητας (SPP)

Με βάση την προηγούμενη αναφορά στην αραιή αναπαράσταση (SR)[21-24], αναλύουμε μια απλή μέθοδο μείωσης διάστασης που ονομάζεται προβολές διατήρησης αραιότητας (SPP). Συγκεκριμένα, στον προτεινόμενο αλγόριθμο, ένας πίνακας βάρους "γεινιάσης" του συνόλου δεδομένων κατασκευάζεται πρώτα βάσει ενός τροποποιημένου framework αραιής αναπαράστασης (MSR), και στη συνέχεια αξιολογείται η ενσωμάτωση χαμηλής διάστασης των δεδομένων έτσι ώστε να διατηρήσει καλύτερα τον πίνακα βάρους. Παρόλο που δεν χρειάζεται εποπτευόμενη (supervised) πληροφορία, ο SPP τείνει να βρει την διακριτική χαρτογράφηση, καθώς η πιο αραιή αναπαράσταση έχει από τη φύση της διακριτική ισχύ: λαμβάνοντας υπόψη τις εικόνες προσώπου, η πιο συμπαγής έκφραση μιας συγκεκριμένης εικόνας προσώπου δίνεται γενικά από τις εικόνες προσώπου της ίδιας κατηγορίας [21]. Μερικά χαρακτηριστικά του προτεινόμενου αλγορίθμου έχουν ως εξής:

- 1) Ο SPP μοιράζεται ορισμένα πλεονεκτήματα του LPP, αλλά και πολλών άλλων γραμμικών μεθόδων μείωσης διαστάσεων. Για παράδειγμα, είναι γραμμικός και ορίζεται παντού, έτσι το πρόβλημα "εκτός δείγματος" είναι λυμένο. Επιπλέον, ο πίνακας βάρους διατηρείται αραιός, όπως στους περισσότερους αλγορίθμους διατήρησης τοπικότητας, πράγμα που είναι ωφέλιμο για την επίλυση του προβλήματος.
- 2) Ο SPP δεν χρειάζεται παραμέτρους, που είναι γενικά δύσκολο να οριστούν στην πράξη, όπως το μέγεθος της γειτονιάς που προκύπτει στο LPP και άλλους αλγορίθμους. Αν και η τεχνική διασταυρούμενης επικύρωσης (cross-validation)[10,40] μπορεί να χρησιμοποιηθεί σε αυτές τις περιπτώσεις, είναι πολύ χρονοβόρα και τείνει να σπαταλά τα περιορισμένα δεδομένα εκπαίδευσης. Αντίθετα, ο SPP δεν χρειάζεται να χρησιμοποιήσει τέτοιες παραμέτρους, με αυτόν τον τρόπο καθίσταται πολύ απλός στη χρήση του στην πράξη.
- 3) Αν και ο SPP εκ φυσικού ανήκει στις παγκόσμιες(global) μεθόδους, κατέχει ορισμένες τοπικές ιδιότητες λόγω της διαδικασίας αραιής αναπαράστασης. Στο επόμενο κεφάλαιο, θα δούμε ότι ο SPP έχει κάποια σύνδεση με αρκετούς δημοφιλείς αλγορίθμους διατήρησης τοπικότητας υπό ορισμένες συνθήκες.

4.1.1 Ανακατασκευή αραιών βαρών

Εφόσον η μέθοδος μείωσης διαστάσεων χαρακτηρίζεται κυρίως από συγκεκριμένο πίνακα βάρους συναφείας των δεδομένων, προσπαθούμε να κατασκευάσουμε τον πίνακα βάσει ενός τροποποιημένου framework αραιής αναπαράστασης, και στη συνέχεια να εξηγήσουμε γιατί είναι τόσο χρήσιμο για τη συμπαγή αναπαράσταση των δεδομένων.

Δεδομένου ενός συνόλου δειγμάτων εκπαίδευσης $\{\mathbf{x}_i\}_{i=1}^n$, όπου $\mathbf{x}_i \in R^m$, ας ορίσουμε $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in R^{m \times n}$ ως τον πίνακα δεδομένων που περιλαμβάνει όλα τα δείγματα εκπαίδευσης στις στήλες του. Σκοπός είναι να ανακατασκευάσουμε κάθε δείγμα \mathbf{x}_i , όπως μια εικόνα προσώπου, χρησιμοποιώντας όσο το δυνατόν λιγότερα δείγματα. Επομένως, αρχικά αναζητούμε ένα διάνυσμα αραιού ανακατασκευαστικού βάρους \mathbf{s}_i για κάθε \mathbf{x}_i μέσω του ακόλουθου τροποποιημένου προβλήματος ελαχιστοποίησης ℓ_1 :

$$\begin{aligned} \min_{\mathbf{s}_i} & \|\mathbf{s}_i\|_1 \\ \text{subject to } & \mathbf{x}_i = \mathbf{X}\mathbf{s}_i \\ \text{and } & \mathbf{1} = \mathbf{1}^T \mathbf{s}_i \end{aligned} \quad (16)$$

όπου $\mathbf{s}_i = [s_{i1}, \dots, s_{i,i-1}, 0, s_{i,i+1}, \dots, s_{in}]^T$ είναι ένα διάνυσμα n -διαστάσεων στο οποίο το i -οστό στοιχείο είναι ίσο με μηδέν, υποδηλώνοντας ότι το \mathbf{x}_i αφαιρείται από το \mathbf{X} , τα στοιχεία $s_{ij}, j \neq i$ υποδηλώνουν τη συνεισφορά κάθε \mathbf{x}_j στην ανακατασκευή του \mathbf{x}_i , $\mathbf{1} \in R^n$ είναι ένα διάνυσμα με κάθε στοιχείο την μονάδα.

Το πρόβλημα MSR μπορεί να λυθεί με τυπικό γραμμικό προγραμματισμό, όπως το αρχικό πρόβλημα ελαχιστοποίησης ℓ_1 εξίσωση (15), καθώς ο περιορισμός άθροισης-σε-ένα $\mathbf{1} = \mathbf{1}^T \mathbf{s}_i$ είναι επίσης γραμμικός. Θα εξηγήσουμε τον λόγο για αυτόν τον περιορισμό σύντομα. Μετά τον υπολογισμό του διανύσματος βάρους \mathbf{s}_i για κάθε $\mathbf{x}_i, i = 1, 2, \dots, n$, μπορούμε να ορίσουμε τον αραιό πίνακα βάρους ανακατασκευής $\mathbf{S} = (\tilde{\mathbf{s}}_{ij})_{n \times n}$ ως εξής:

$$\mathbf{S} = [\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \dots, \tilde{\mathbf{s}}_n]^T \quad (17)$$

όπου $\tilde{\mathbf{s}}_i$ είναι η βέλτιστη λύση της εξίσωσης (16). Το στοιχείο \tilde{s}_{ij} στο \mathbf{S} δεν είναι μια απλή μέτρηση ομοιότητας μεταξύ των δειγμάτων \mathbf{x}_i και \mathbf{x}_j , και σε αυτήν την περίπτωση το \mathbf{S} διαφέρει ουσιαστικά από τον πίνακα βάρους γειτνίασης του LPP.

Παραθέτουμε μερικές ενδείξεις για την αποτελεσματικότητα του \mathbf{S} ως πίνακα βάρους για τη μείωση της διάστασης και το ακόλουθο έργο αναγνώρισης.

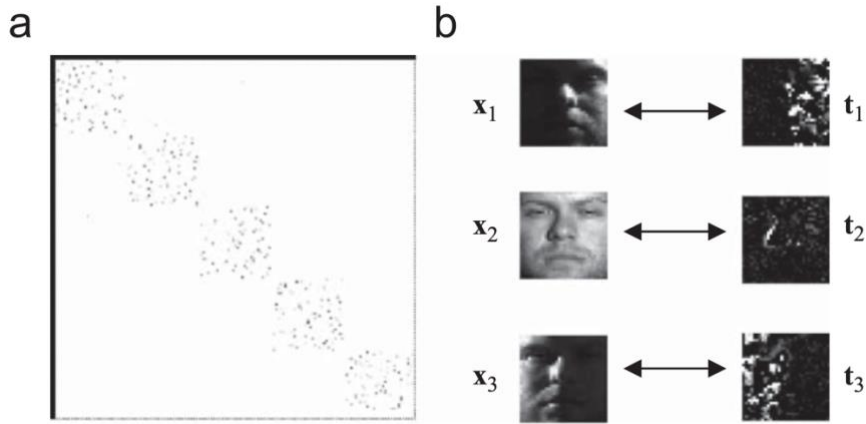
- 1) Κάθε διάνυσμα βάρους \mathbf{s}_i υπακούει σε μια σημαντική συμμετρία: είναι αμετάβλητο σε περιστροφές (rotation) και επανακλιμακώσεις (rescale), λόγω του πρώτου περιορισμού στην εξίσωση (16), και αμετάβλητο σε μετατοπίσεις, λόγω του

περιορισμού άθροισης-σε-ένα $1 = \mathbf{1}^T \mathbf{s}_i$. Ως αποτέλεσμα, ο πίνακας βάρους \mathbf{S} αντικατοπτρίζει κάποιες εσωτερικές γεωμετρικές ιδιότητες των δεδομένων.

- 2) Πληροφορίες διάκρισης μπορούν να διατηρηθούν στον πίνακα βάρους \mathbf{S} , ακόμη και αν δεν παρέχονται ετικέτες κλάσης. Ας πάρουμε την αναγνώριση προσώπου ως παράδειγμα. Μια ιδιαίτερα απλή αλλά αποτελεσματική υπόθεση στην αναγνώριση προσώπου είναι ότι τα δείγματα από την ίδια κλάση βρίσκονται σε ένα γραμμικό υποχώρο (face subspace). Δεδομένης μιας εικόνας προσώπου \mathbf{x}_i^j από την j -οστή κλάση, το \mathbf{x}_i^j μπορεί θεωρητικά να αναπαρασταθεί χρησιμοποιώντας τα δείγματα από την j -οστή κλάση, σύμφωνα με την υπόθεση του υποχώρου. Δηλαδή:

$$\mathbf{x}_i^j = 0 \cdot \mathbf{x}_1^1 + \dots + \alpha_{i,i-1} \mathbf{x}_{i-1}^j + s_{i,i+1} \mathbf{x}_{i+1}^j + \dots + 0 \cdot \mathbf{x}_n^c \quad (18)$$

όπου $j = 1, \dots, c$ υποδηλώνει την ετικέτα κλάσης. Το διάνυσμα βάρους $\mathbf{s}_i^0 = [0, \dots, \alpha_{i,i-1}, 0, \alpha_{i,i+1}, \dots, 0]^T$ είναι αραιό, καθώς ο αριθμός κλάσης είναι γενικά μεγάλος στα περισσότερα προβλήματα αναγνώρισης προσώπου. Αν και η εξίσωση (18) δεν ισχύει πάντα λόγω ανεπαρκούς δειγματοληψίας, τα πειράματά μας δείχνουν ότι το αραιό \mathbf{s}_i^0 μπορεί πραγματικά να προσεγγιστεί με τις βέλτιστες λύσεις $\tilde{\mathbf{s}}_i$ (Εικόνα 9). Με άλλα λόγια, οι μη μηδενικές καταχωρήσεις στο $\tilde{\mathbf{s}}_i$ αντιστοιχούν κυρίως στα δείγματα από την j -οστή κλάση, που υποδηλώνει ότι το $\tilde{\mathbf{s}}_i$ μπορεί να βοηθήσει στη διάκριση αυτής της κλάσης από τις υπόλοιπες. Επομένως, το διάνυσμα βάρους $\tilde{\mathbf{s}}_i$, που κατασκευάζεται χρησιμοποιώντας όλα τα δείγματα με περιορισμό να είναι αραιά, τείνει να περιλαμβάνει πιθανές πληροφορίες διάκρισης.



Εικόνα 9. (α) Ένα υπο-μπλοκ του πίνακα βαρών \mathbf{S} που κατασκευάστηκε από την εξίσωση (20). (β) Τα βέλτιστα \mathbf{t}_i για τρία διαφορετικά δείγματα.

Όπως περιγράφηκε παραπάνω, σε κάποιες πραγματικές εφαρμογές, ο περιορισμός $\mathbf{x}_i = \mathbf{X} \mathbf{s}_i$ στην εξίσωση (16) δεν ισχύει πάντα. Λαμβάνοντας υπόψη τα δυο μοντέλα που αναφέρθηκαν στο Κεφάλαιο 3.2, το πρόβλημα MSR μπορεί να επεκταθεί στις ακόλουθες δύο σταθερές εκδόσεις, (19) και (20). Η πρώτη επέκταση ορίζεται ως:

$$\begin{aligned}
& \min_{\mathbf{s}_i, \mathbf{t}} \|\mathbf{s}_i\|_1 \\
& \text{subject to } \|\mathbf{x}_i - \mathbf{X}\mathbf{s}_i\| < \varepsilon \\
& \text{and } \mathbf{1} = \mathbf{1}^T \mathbf{s}_i
\end{aligned} \tag{19}$$

όπου ε είναι η ανοχή σφάλματος και γενικά είναι σταθερή καθ' όλη την διάρκεια του προβλήματος [21]. Είναι εύκολο να επιβεβαιώσουμε ότι η βέλτιστη λύση αντικατοπτρίζει ακόμη κάποιες εσωτερικές γεωμετρικές ιδιότητες των αρχικών δεδομένων, όπως η αμεταβλητότητα σε μετατοπίσεις και περιστροφές. Μια άλλη επέκταση του MSR μπορεί να εκφραστεί ως:

$$\begin{aligned}
& \min_{[\mathbf{s}_i^T \mathbf{t}_i^T]^T} \|[\mathbf{s}_i^T \mathbf{t}_i^T]^T\|_1 \\
& \text{subject to } \begin{bmatrix} \mathbf{x}_i \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{I} \\ \mathbf{1}^T & \mathbf{0}^T \end{bmatrix} \begin{bmatrix} \mathbf{s}_i \\ \mathbf{t}_i \end{bmatrix}
\end{aligned} \tag{20}$$

όπου \mathbf{t}_i είναι ένα διάνυσμα m -διαστάσεων και $\mathbf{0}$ είναι ένα διάνυσμα m -διαστάσεων με όλα τα στοιχεία μηδενικά. Η βέλτιστη λύση του (20) είναι επίσης αμετάβλητη σε μετατοπίσεις, αλλά η αμεταβλητότητα σε περιστροφές και επανακλιμακώσεις δεν ισχύει αυστηρά πλέον.

4.1.2 Διατήρηση αραιών βαρών ανακατασκευής

Με τον παραπάνω σχεδιασμό, ο αραιός πίνακας βάρους \mathbf{S} μπορεί να αντικατοπτρίσει τις εσωτερικές γεωμετρικές ιδιότητες των δεδομένων σε κάποιο βαθμό και να περιέχει φυσικές πληροφορίες διάκρισης. Επομένως, περιμένουμε ότι τα επιθυμητά χαρακτηριστικά στον αρχικό χώρο υψηλής διάστασης μπορούν να διατηρηθούν στον υποχώρο χαμηλής διάστασης. Παρόμοια με το LLE (Locally Linear Embedding), ορίζουμε την ακόλουθη αντικειμενική συνάρτηση για να αναζητήσουμε τις προβολές που διατηρούν καλύτερα το βέλτιστο διάνυσμα βάρους $\tilde{\mathbf{s}}_i$:

$$\min_{\mathbf{w}} \sum_{i=1}^n \|\mathbf{w}^T \mathbf{x}_i - \mathbf{w}^T \mathbf{X} \tilde{\mathbf{s}}_i\|^2 \tag{21}$$

Με απλή άλγεβρα, μπορούμε να πάρουμε:

$$\sum_{i=1}^n \|\mathbf{w}^T \mathbf{x}_i - \mathbf{w}^T \mathbf{X} \tilde{\mathbf{s}}_i\|^2 = \mathbf{w}^T \left(\sum_{i=1}^n (\mathbf{x}_i - \mathbf{X} \tilde{\mathbf{s}}_i)(\mathbf{x}_i - \mathbf{X} \tilde{\mathbf{s}}_i)^T \right) \mathbf{w} \quad (22)$$

Ας είναι το \mathbf{e}_i ένα n -διάστατο μοναδιαίο διάνυσμα, με το i -οστό στοιχείο να είναι το 1, διαφορετικά το 0, τότε η εξίσωση (18) είναι ίση με:

$$\begin{aligned} & \mathbf{w}^T \left(\sum_{i=1}^n (\mathbf{X} \mathbf{e}_i - \mathbf{X} \tilde{\mathbf{s}}_i)(\mathbf{X} \mathbf{e}_i - \mathbf{X} \tilde{\mathbf{s}}_i)^T \right) \mathbf{w} \\ &= \mathbf{w}^T \mathbf{X} \left(\sum_{i=1}^n (\mathbf{e}_i - \tilde{\mathbf{s}}_i)(\mathbf{e}_i - \tilde{\mathbf{s}}_i)^T \right) \mathbf{X}^T \mathbf{w} \\ &= \mathbf{w}^T \mathbf{X} \left(\sum_{i=1}^n \mathbf{e}_i \mathbf{e}_i^T - \tilde{\mathbf{s}}_i \mathbf{e}_i^T - \mathbf{e}_i \tilde{\mathbf{s}}_i^T - \tilde{\mathbf{s}}_i \tilde{\mathbf{s}}_i^T \right) \mathbf{X}^T \mathbf{w} \\ &= \mathbf{w}^T \mathbf{X} (\mathbf{I} - \mathbf{S} - \mathbf{S}^T + \mathbf{S}^T \mathbf{S}) \mathbf{X}^T \mathbf{w} \end{aligned} \quad (23)$$

Για να αποφύγουμε εκφυλισμένες λύσεις, επιβάλλουμε $\mathbf{w}^T \mathbf{X} \mathbf{X}^T \mathbf{w} = 1$. Έτσι, η αντικειμενική συνάρτηση μπορεί να αναδιατυπωθεί ως το ακόλουθο πρόβλημα βελτιστοποίησης:

$$\min_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{X} (\mathbf{I} - \mathbf{S} - \mathbf{S}^T + \mathbf{S}^T \mathbf{S}) \mathbf{X}^T \mathbf{w}}{\mathbf{w}^T \mathbf{X} \mathbf{X}^T \mathbf{w}} \quad (24)$$

Για συμπαγή έκφραση, το πρόβλημα ελαχιστοποίησης μπορεί να μετασχηματιστεί περαιτέρω σε ένα ισοδύναμο πρόβλημα μεγιστοποίησης ως εξής:

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{X} \mathbf{S}_\beta \mathbf{X}^T \mathbf{w}}{\mathbf{w}^T \mathbf{X} \mathbf{X}^T \mathbf{w}} \quad (25)$$

όπου $\mathbf{S}_\beta = \mathbf{S} + \mathbf{S}^T - \mathbf{S}^T \mathbf{S}$. Ένα άλλο πλεονέκτημα αυτού του μετασχηματισμού είναι ότι το πρόβλημα μεγιστοποίησης σε κάποιες περιπτώσεις μπορεί να πάρει μια πιο αριθμητικά σταθερή λύση [17]. Στη συνέχεια, τα βέλτιστα \mathbf{w} είναι τα ιδιοδιανύσματα που αντιστοιχούν στις μεγαλύτερες d ιδιοτιμές του ακόλουθου γενικευμένου προβλήματος ιδιοτιμών:

$$\mathbf{XS}_\beta \mathbf{X}^T \mathbf{w} = \lambda \mathbf{XX}^T \mathbf{w} \quad (26)$$

4.1.3 SPP αλγόριθμος

Βάσει των παραπάνω, συνοψίζουμε τον SPP αλγόριθμο ως εξής:

- 1) Δημιουργία πίνακα βάρους \mathbf{S} χρησιμοποιώντας MSR(16) ή το σταθερό MSR(19), (20).
- 2) Υπολογισμός των διανυσμάτων προβολής χρησιμοποιώντας την εξίσωση (26), και τα ιδιοδιανύσματα που αντιστοιχούν στις μεγαλύτερες d ιδιοτιμές που καλύπτουν τον βέλτιστο υποχώρο.

Ο αλγόριθμος είναι απλός, καθώς δεν περιλαμβάνει καμία υπερ-παραμέτρο εκτός από τη διάσταση του υποχώρου d . Για κάποια δεδομένα υψηλής διάστασης, ο πίνακας \mathbf{XX}^T είναι μοναδιαίος, καθώς το μέγεθος του δείγματος εκπαίδευσης είναι πολύ μικρότερο από τις διαστάσεις των χαρακτηριστικών. Για να αντιμετωπιστεί αυτό το πρόβλημα, το σύνολο εκπαίδευσης μπορεί πρώτα να προβληθεί σε έναν υποχώρο PCA που καλύπτεται από τα κύρια ιδιοδιανύσματά του: $\mathbf{W}_{pca} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_d]$. Ο πίνακας \mathbf{XX}^T στη συνέχεια προσεγγίζεται από τον $\hat{\mathbf{X}}\hat{\mathbf{X}}^T$ ($\hat{\mathbf{X}} = \mathbf{W}_{pca}^T \mathbf{X}$), που δεν είναι μοναδιαίος πλέον. Στο επόμενο κεφάλαιο θα δούμε πιο αναλυτικά την υλοποίηση, χρησιμοποιώντας την python.

4.2 Διακριτική αραιή ενσωμάτωση που διατηρεί τη γειτονιά (DSNPE)

Σε αυτό το κεφάλαιο, εξετάζουμε έναν αλγόριθμο διακριτικής αραιής διατήρησης γειτονιάς (DSNPE) συνδυάζοντας τις μεθόδους SPP και το κριτήριο μέγιστου περιθωρίου (MMC), το οποίο μπορεί να θεωρηθεί ως ένας νέος αλγόριθμος που ενσωματώνει το κριτήριο Φίσερ και το κριτήριο αραιότητας. Το κριτήριο μέγιστου περιθωρίου είναι μια μέθοδος που προτάθηκε για τη μεγιστοποίηση του ίχνους της διαφοράς του μεταξύ-κλάσης πίνακα διασποράς και του εντός-κλάσης πίνακα διασποράς, από το οποίο μπορεί να προκύψει το LDA, ενσωματώνοντας κάποιους περιορισμούς. Έτσι, το DSNPE προτείνεται με την εισαγωγή του κριτηρίου μέγιστου περιθωρίου στην αντικειμενική συνάρτηση του SPP, το οποίο έχει δύο πλεονεκτήματα:

- 1) Διατηρεί το χαρακτηριστικό αραιότητας του SPP.
- 2) Τονίζει τη διακριτική πληροφορία, ενσωματώνοντας το κριτήριο μέγιστου περιθωρίου (MMC), το οποίο μπορεί να κάνει τα διανύσματα κλάσης μέσης τιμής να έχουν μια ευρεία διασπορά και να κάνει κάθε κλάση να διασκορπίζεται σε έναν μικρό χώρο.

Επιπρόσθετα, για να αυξήσουμε περαιτέρω τη διακριτική ισχύ του DSNPE, ενσωματώνουμε επιπλέον διακριτική πληροφορία. Συγκεκριμένα, για τον υπολογισμό της αραιής ανακατασκευαστικής σχέσης, χρησιμοποιούμε μόνο τα δείγματα εκπαίδευσης με την ίδια ετικέτα με το τρέχον δείγμα, αντί να χρησιμοποιούμε όλα τα δείγματα εκπαίδευσης. Ο λόγος πίσω από αυτήν την απόφαση βασίζεται στην ακόλουθη παρατήρηση: λαμβάνοντας υπόψη τις εικόνες προσώπου, η πιο συμπαγής έκφραση μιας συγκεκριμένης εικόνας προσώπου δίνεται γενικά από τις εικόνες προσώπου της ίδιας κλάσης [21]. Παρακάτω θα δούμε αναλυτικά την μέθοδο, ξεκινώντας με μια σύντομη ανασκόπηση του κριτηρίου μέγιστου περιθωρίου.

4.2.1 Κριτήριο μεγιστοποίησης περιθωρίου (MMC)

Το κριτήριο μέγιστου περιθωρίου (MMC) [4,5] προτείνεται για τη μεγιστοποίηση του (μέσου) περιθωρίου μεταξύ των κλάσεων μετά τη μείωση της διάστασης. Το MMC μπορεί να αναπαραστήσει τον διαχωρισμό της κλάσης καλύτερα από το PCA. Επιπλέον, όπως αναφέρθηκε (και είδαμε στο Κεφάλαιο 2.2.3) προηγουμένως το LDA μπορεί να προκύψει από το MMC με την ενσωμάτωση ορισμένων περιορισμών. Ωστόσο, το MMC δεν επηρεάζεται αρνητικά από το πρόβλημα του μικρού μεγέθους δείγματος, το οποίο είναι γνωστό ότι προκαλεί σοβαρά προβλήματα σταθερότητας στο LDA. Στο [4,5], η συνάρτηση στόχου του MMC γράφεται ως:

$$J_1 = \max \left\{ \sum_{ij} p_i p_j (d(\mathbf{m}_i, \mathbf{m}_j) - \mathbf{s}(\mathbf{m}_i) - \mathbf{s}(\mathbf{m}_j)) \right\} \quad (27)$$

όπου p_i και p_j είναι η προηγούμενη πιθανότητα της κλάσης i και της κλάσης j , αντίστοιχα, ενώ m_i και m_j είναι τα διανύσματα μέσης τιμής της κλάσης i και της κλάσης j . Εδώ, τα $d(\mathbf{m}_i, \mathbf{m}_j)$, $\mathbf{s}(\mathbf{m}_i)$, και $\mathbf{s}(\mathbf{m}_j)$ ορίζονται ως:

$$d(\mathbf{m}_i, \mathbf{m}_j) = \|\mathbf{m}_i - \mathbf{m}_j\| \quad (28)$$

$$\mathbf{s}(\mathbf{m}_i) = \text{tr}(\mathbf{S}_i) \quad (29)$$

$$\mathbf{s}(\mathbf{m}_j) = \text{tr}(\mathbf{S}_j) \quad (30)$$

όπου \mathbf{S}_j είναι ο πίνακας διασποράς της κλάσης j .

Έτσι, η βελτιστοποιημένη συνάρτηση μπορεί να προκύψει ως εξής:

$$J_2 = \max \text{tr}(\mathbf{S}_b - \mathbf{S}_w) \quad (31)$$

Ο πίνακας \mathbf{S}_b ονομάζεται πίνακας διασποράς μεταξύ των κλάσεων και ο \mathbf{S}_w ονομάζεται πίνακας διασποράς εντός της κλάσης.

4.2.2 DSNPE αλγόριθμος

Σε αυτό το κεφάλαιο, θα συζητήσουμε τη λύση του DSNPE. Εμπνευσμένοι από την παρατήρηση ότι η πιο συμπαγής έκφραση ενός συγκεκριμένου προσώπου γενικά δίνεται από τις εικόνες προσώπου από την ίδια κλάση [21], τροποποιούμε την αρχική αραιή αναπαράσταση ως εξής:

$$\min_{\mathbf{s}_i} \|\mathbf{s}_i\|_1 \text{ subject to } \mathbf{x}_i = \mathbf{X}_k \mathbf{s}_i \text{ και } \text{label}(\mathbf{x}_i) = k \quad (32)$$

ή

$$\min_{\mathbf{s}_i} \|\mathbf{s}_i\|_1 \text{ subject to } \|\mathbf{x}_i - \mathbf{X}_k \mathbf{s}_i\| < \varepsilon \text{ και } \text{label}(\mathbf{x}_i) = k \quad (33)$$

όπου \mathbf{X}_k δηλώνει το σύνολο των δειγμάτων εκπαίδευσης των οποίων η ετικέτα είναι η ίδια με το \mathbf{x}_i . Δηλαδή, για να υπολογίσουμε την αραιή ανακατασκευαστική σχέση, χρησιμοποιούμε μόνο τα δείγματα εκπαίδευσης με την ίδια ετικέτα, όπως το τρέχον δείγμα, αντί να χρησιμοποιούμε όλα τα δείγματα εκπαίδευσης.

Επιπλέον, αν ένας γραμμικός μετασχηματισμός $\mathbf{Y} = \mathbf{w}^T \mathbf{X}$ μπορεί να μεγιστοποιήσει το J_2 , θα ερευνηθεί ένας βέλτιστος υπόχωρος για την ταξινόμηση προτύπων. Αυτό συμβαίνει επειδή ο γραμμικός μετασχηματισμός στοχεύει να προβάλει ένα πρότυπο πιο κοντά στα πρότυπα της ίδιας κλάσης αλλά πιο μακριά από αυτά με διαφορετικές κλάσεις, το οποίο είναι ακριβώς ο στόχος για την ταξινόμηση. Δηλαδή, για να βρεθεί ένας βέλτιστος γραμμικός υπόχωρος για την ταξινόμηση θα πρέπει να μεγιστοποιηθεί η ακόλουθη συνάρτηση βελτιστοποίησης:

$$J_3 = \max \text{tr}(\mathbf{w}^T (\mathbf{S}_b - \mathbf{S}_w) \mathbf{w}) \quad (34)$$

Εάν ο γραμμικός μετασχηματισμός που προκύπτει από τον SPP μπορεί να ικανοποιήσει ταυτόχρονα το J_3 , η διακριτικότητα των δεδομένων θα βελτιωθεί σημαντικά. Έτσι, η λύση

για το DSNPE μπορεί να αναπαρασταθεί ως το ακόλουθο πρόβλημα βελτιστοποίησης πολλαπλών αντικειμένων:

$$\begin{cases} \min tr(\mathbf{w}^T \mathbf{X} \mathbf{S}_\alpha \mathbf{X}^T \mathbf{w}) \\ \max tr(\mathbf{w}^T (\mathbf{S}_b - \mathbf{S}_w) \mathbf{w}) \end{cases} \text{ subject to } \mathbf{w}^T \mathbf{X} \mathbf{X}^T \mathbf{w} = \mathbf{I} \quad (35)$$

όπου $\mathbf{S}_\alpha = \mathbf{I} - \mathbf{S} - \mathbf{S}^T + \mathbf{S}^T \mathbf{S}$. Η λύση στο πρόβλημα βελτιστοποίησης με περιορισμούς πολλαπλών αντικειμένων είναι να βρεθεί ένας υπόχωρος που να διατηρεί την ιδιότητα της αραιότητας και ταυτόχρονα να μεγιστοποιεί το περιθώριο μεταξύ διαφορετικών κλάσεων. Οπότε το πρόβλημα μπορεί να μετατραπεί στο ακόλουθο:

$$\begin{aligned} \min tr(\mathbf{w}^T (\mathbf{X} \mathbf{S}_\alpha \mathbf{X}^T - \gamma (\mathbf{S}_b - \mathbf{S}_w)) \mathbf{w}) \\ \text{subject to } \mathbf{w}^T \mathbf{X} \mathbf{X}^T \mathbf{w} = \mathbf{I} \end{aligned} \quad (36)$$

όπου το γ είναι μια παράμετρος για την ισορροπία μεταξύ της αραιότητας και της διακριτικής πληροφορίας.

Η εξίσωση (36) μπορεί να λυθεί με τη μέθοδο του πολλαπλασιαστή Lagrangian:

$$\frac{\partial}{\partial \mathbf{w}} tr(\mathbf{w}^T (\mathbf{X} \mathbf{S}_\alpha \mathbf{X}^T - \gamma (\mathbf{S}_b - \mathbf{S}_w)) \mathbf{w} - \lambda_i (\mathbf{w}^T \mathbf{X} \mathbf{X}^T \mathbf{w} - \mathbf{I})) = 0$$

όπου το λ_i είναι ο πολλαπλασιαστής Lagrangian. Στη συνέχεια, μπορούμε να πάρουμε:

$$(\mathbf{X} \mathbf{S}_\alpha \mathbf{X}^T - \gamma (\mathbf{S}_b - \mathbf{S}_w)) \mathbf{w}_i = \lambda_i \mathbf{X} \mathbf{X}^T \mathbf{w}_i \quad (37)$$

όπου το \mathbf{w}_i είναι το γενικευμένο ιδιοδιάνυσμα του $\mathbf{X} \mathbf{S}_\alpha \mathbf{X}^T - \gamma (\mathbf{S}_b - \mathbf{S}_w)$ και $\mathbf{X} \mathbf{X}^T$ και το λ_i είναι η αντίστοιχη ιδιοτιμή.

Τα διανύσματα στηλών $\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_{d-1}$ είναι οι λύσεις της εξίσωσης (37), διατεταγμένα σύμφωνα με τις πρώτες d μικρότερες ιδιοτιμές $\lambda_0, \lambda_1, \dots, \lambda_{d-1}$. Έτσι, η ενσωμάτωση γράφεται ως εξής:

$$\mathbf{x}_i \rightarrow \mathbf{y}_i = \mathbf{w}^T \mathbf{x}_i, \quad \mathbf{w} = [\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_{d-1}]$$

όπου \mathbf{y}_i είναι ένα διάνυσμα d -διαστάσεων και \mathbf{w} είναι ένας πίνακας $D \times d$.

Ο DSNPE αλγόριθμος ακολουθεί τα ίδια βήματα, όπως ο SPP και συνοψίζεται ως εξής:

- 1) Όπως και με τον SPP έτσι και με τον DSNPE προβάλλεται πρώτα το σύνολο εικόνων $\{\mathbf{x}_i\}$ στον υποχώρο PCA, απορρίπτοντας έτσι τα μικρότερα κύρια συστατικά (principal components).

- 2) Δημιουργία του πίνακα βάρους S χρησιμοποιώντας τις εξισώσεις (32) ή (33).
- 3) Λύση του γενικευμένου προβλήματος ιδιοτιμών χρησιμοποιώντας την εξίσωση (37), δημιουργία του $D \times d$ πίνακα χαρακτηριστικών w του οποίου οι στήλες αποτελούνται από τα ιδιοδιανύσματα που αντιστοιχούν στις d –μικρότερες ιδιοτιμές του.

4.2.3 Ανάλυση πολυπλοκότητας χρόνου

Σε αυτή την ενότητα, αναλύουμε σύντομα και θεωρητικά την πολυπλοκότητα χρόνου του αλγορίθμου DSNPE. Παραλείπουμε την ανάλυση πολυπλοκότητας χρόνου της αραιής μάθησης, επειδή υπάρχουν πολλά πακέτα λογισμικού που υλοποιούν τον αλγόριθμο της αραιής μάθησης και κάθε πακέτο έχει διαφορετικές πολυπλοκότητες χρόνου. Για διευκόλυνση, δίνουμε μια σημείωση ότι ο αριθμός των κύριων συστατικών στο βήμα PCA του DSNPE είναι q . Το DSNPE περιλαμβάνει το βήμα PCA και το βήμα ιδιοαποσύνθεσης (eigendecomposition) χρησιμοποιώντας την εξίσωση (37). Εφόσον το βήμα PCA του DSNPE είναι το ίδιο με αυτό που χρησιμοποιείται συχνά σε άλλους αλγορίθμους, όπως στο SPP (δηλαδή, PCA + SPP) και στο LPP, εστιάζουμε στην πολυπλοκότητα χρόνου του βήματος ιδιοαποσύνθεσης, χρησιμοποιώντας την εξίσωση (37), που είναι $O(q^3)$. Επομένως, η πολυπλοκότητα χρόνου του DSNPE είναι $O(q^3)$.

4.2.4 Θεωρητική ανάλυση του DSNPE αλγορίθμου

Σε αυτή την ενότητα, παρέχουμε κάποιες θεωρητικές αναλύσεις για να κατανοήσουμε καλύτερα τα χαρακτηριστικά του DSNPE. Αρχικά, παρουσιάζεται ένα λήμμα ως εξής:

Λήμμα 1. [57] Για συμμετρικούς πίνακες $A \in R^{n \times n}$, $E \in R^{n \times n}$, έστω $A = QLQ^T$ η ιδιοαποσύνθεση (eigendecomposition) του A και $A + E = B = PL_1P^T$ η ιδιοαποσύνθεση του B . Γράψε $Q = [q_1, q_2, \dots, q_n]$, $P = [p_1, p_2, \dots, p_n]$, όπου q_i και p_i είναι τα κανονικοποιημένα ιδιοδιανύσματα του A και B αντίστοιχα. Έστω θ η οξεία γωνία μεταξύ των q_i και p_i , τότε $\sin(\theta) \leq a\|E\|_2$, όπου a είναι μια σταθερά που εξαρτάται μόνο από το A .

Το ακόλουθο Θεώρημα 1 χαρακτηρίζει τη λύση του DSNPE όταν η παράμετρος γ πλησιάζει το άπειρο. Το Θεώρημα 1 απαιτεί το XX^T να είναι θετικό, που πάντα ισχύει για τον αλγόριθμό μας, εφόσον χρησιμοποιούμε το PCA για την προεπεξεργασία των δεδομένων. Επιπλέον, υποθέτουμε ότι ο πίνακας δεδομένων X έχει κεντραριστεί.

Θεώρημα 1. Όταν $\gamma \rightarrow \infty$, το w_i που προκύπτει από την προτεινόμενη μέθοδο DSNPE συγκλίνει στο γενικευμένο ιδιοδιάνυσμα m_i του μεταξύ κλάσεων πίνακα διασποράς S_b και του εντός κλάσης πίνακα διασποράς S_w , δηλαδή, όταν $\gamma \rightarrow \infty$, υπάρχει μια σταθερά β τέτοια ώστε:

$$\mathbf{w}_i - \beta \mathbf{m}_i = 0$$

Απόδειξη 1. Εάν και οι δύο πλευρές της εξίσωσης (37) διαιρεθούν με γ , τότε έχουμε:

$$\left(\frac{1}{\gamma} \mathbf{X} \mathbf{S}_\alpha \mathbf{X}^T - (\mathbf{S}_b - \mathbf{S}_w) \right) \mathbf{w}_i = \frac{\lambda_i}{\gamma} \mathbf{X} \mathbf{X}^T \mathbf{w}_i \quad (38)$$

Η εξίσωση (38) είναι ισοδύναμη με:

$$(\mathbf{X} \mathbf{X}^T)^{-0.5} \left(\frac{1}{\gamma} \mathbf{X} \mathbf{S}_\alpha \mathbf{X}^T - (\mathbf{S}_b - \mathbf{S}_w) \right) (\mathbf{X} \mathbf{X}^T)^{-0.5} (\mathbf{X} \mathbf{X}^T)^{0.5} \mathbf{w}_i = \frac{\lambda_i}{\gamma} (\mathbf{X} \mathbf{X}^T)^{0.5} \mathbf{w}_i \quad (39)$$

Επομένως το $(\mathbf{X} \mathbf{X}^T)^{0.5} \mathbf{w}_i$ είναι το ιδιοδιάνυσμα του $(\mathbf{X} \mathbf{X}^T)^{-0.5} (\frac{1}{\gamma} \mathbf{X} \mathbf{S}_\alpha \mathbf{X}^T - (\mathbf{S}_b - \mathbf{S}_w)) (\mathbf{X} \mathbf{X}^T)^{-0.5}$.

Από την άλλη πλευρά, από τον ορισμό του \mathbf{m}_i , παίρνουμε:

$$\mathbf{S}_b \mathbf{m}_i = c_i \mathbf{S}_w \mathbf{m}_i \quad (40)$$

όπου c_i είναι η αντίστοιχη ιδιοτιμή του \mathbf{m}_i .

Εφόσον το \mathbf{X} έχει κεντραριστεί, τότε έχουμε $\mathbf{X} \mathbf{X}^T = n \mathbf{S}_t = n(\mathbf{S}_b + \mathbf{S}_w)$.

Με την (40), έχουμε στη συνέχεια:

$$-(\mathbf{X} \mathbf{X}^T)^{-0.5} (\mathbf{S}_b + \mathbf{S}_w) (\mathbf{X} \mathbf{X}^T)^{-0.5} (\mathbf{X} \mathbf{X}^T)^{0.5} \mathbf{m}_i = d_i (\mathbf{X} \mathbf{X}^T)^{0.5} \mathbf{m}_i \quad (41)$$

όπου $d_i = -1 + c_i/n + n c_i$. Επομένως το $(\mathbf{X} \mathbf{X}^T)^{0.5} \mathbf{m}_i$ είναι το ιδιοδιάνυσμα του $-(\mathbf{X} \mathbf{X}^T)^{-0.5} (\mathbf{S}_b + \mathbf{S}_w) (\mathbf{X} \mathbf{X}^T)^{-0.5}$.

Ας υποθέσουμε ότι το θ δηλώνει την οξεία γωνία μεταξύ $(\mathbf{X} \mathbf{X}^T)^{0.5} \mathbf{m}_i$ και $(\mathbf{X} \mathbf{X}^T)^{0.5} \mathbf{w}_i$.

Εφαρμόζοντας άμεσα το **Λήμμα 1**, έχουμε:

$$\sin(\theta) \leq \frac{\alpha}{\gamma} \|(\mathbf{X} \mathbf{X}^T)^{-0.5} (\mathbf{X} \mathbf{S}_\alpha \mathbf{X}^T) (\mathbf{X} \mathbf{X}^T)^{-0.5}\|_2$$

όταν το γ προσεγγίζει το άπειρο, είναι εύκολο να δούμε ότι $\sin(\theta) \rightarrow 0$. Επομένως, υπάρχει μια σταθερά β τέτοια ώστε:

$$(\mathbf{X} \mathbf{X}^T)^{0.5} (\mathbf{w}_i - \beta \mathbf{m}_i) = 0 \quad (42)$$

Εφόσον υποθέτουμε ότι $\mathbf{X}\mathbf{X}^T > 0$, έχουμε:

$$\mathbf{w}_i - \beta \mathbf{m}_i = 0 \quad (43)$$

Αυτό ολοκληρώνει την απόδειξη του **θεωρήματος 1**.

Όταν η παράμετρος γ είναι ίση με το μηδέν, το DSNPE μετατρέπεται στον SPP. Από αυτό μπορούμε να συμπεράνουμε ότι, το SPP είναι μια ειδική περίπτωση του DSNPE.

5. Κεφάλαιο

Πειραματικά αποτελέσματα

5.1 Το πρόβλημα

Η αναγνώριση προσώπων αποτελεί ένα από τα πιο σύνθετα και καίρια προβλήματα στις επιστήμες της μηχανικής μάθησης και της αναγνώρισης προτύπων. Στην ουσία, το πρόβλημα περιστρέφεται γύρω από το διαχωρισμό και την ταυτοποίηση ατόμων, με βάση αποκλειστικά τα χαρακτηριστικά του προσώπου τους. Αυτό απαιτεί την εξαγωγή μοναδικών και συνεπών μοτίβων από εικόνες προσώπων, παρά τις πολυάριθμες πιθανές παραλλαγές. Αυτές οι παραλλαγές μπορεί να προκύψουν από αλλαγές στον φωτισμό, τις προσωπικές εκφράσεις, τη γήρανση και τα εμπόδια, όπως είναι τα γυαλιά ή καπέλα. Επιπλέον, με την τεράστια ποικιλία στα ανθρώπινα πρόσωπα διαφορετικών εθνοτήτων, ηλικιών και άλλων στοιχείων, η κατασκευή ενός παγκοσμίως αποτελεσματικού συστήματος αναγνώρισης προσώπων γίνεται ακόμη πιο δύσκολη. Στόχος είναι, λοιπόν, να σχεδιαστούν αλγόριθμοι που μπορούν να αναγνωρίζουν συνεχώς πρόσωπα με υψηλή ακρίβεια, ανεξάρτητα από τις προαναφερθείσες προκλήσεις. Αυτό προσπαθήσαμε να πετύχουμε και εμείς με πολύ θετικά αποτελέσματα.

5.2 Σύνολο δεδομένων

Αρχικά παρέχουμε απλές περιγραφές των συνόλων δεδομένων που θα χρησιμοποιηθούν αργότερα.

5.2.1 Yale

Μία από τις πιο γνωστές βάσεις δεδομένων για πειραματικούς σκοπούς είναι η Yale του πανεπιστημίου του Γέιλ, στο Νιου Χέιβεν του Κονέκτικατ των Ηνωμένων Πολιτειών της Αμερικής. Αυτή η βάση δεδομένων [2] περιέχει 165 εικόνες προσώπων από 15 άτομα. Υπάρχουν 11 εικόνες ανά άτομο, και αυτές οι 11 εικόνες είναι, αντίστοιχα, υπό διαφορετική έκφραση ή κατάσταση. Οι εικόνες δείχνουν παραλλαγές στις συνθήκες φωτισμού, δηλαδή μπορούμε να έχουμε το φως να χτυπάει από αριστερά, από το κέντρο και από τα δεξιά. Στην έκφραση του προσώπου, έχουμε εικόνες με κανονική έκφραση, χαρούμενη, λυπημένη, νυσταγμένη, έκπληκτη και με κλειστό το ένα μάτι, και τέλος, τα άτομα μπορεί να είναι με ή χωρίς γυαλιά. Στο πείραμά μας, όλες οι εικόνες ήταν σε

ασπρόμαυρο, περικόπηκαν και αναδιαμορφώθηκαν στην ανάλυση των 32x32 εικονοστοιχείων. Προεπεξεργαστήκαμε τα δεδομένα κανονικοποιώντας κάθε διάνυσμα προσώπου στη μονάδα. Στην εικόνα 10 παρουσιάζονται έντεκα περικομμένα και αναδιαμορφωμένα δείγματα ενός ατόμου στη βάση δεδομένων προσώπων του Yale.



Εικόνα 10. Τριάντα δύο περικομμένα δείγματα εικόνων ενός ατόμου από την βάση δεδομένων Yale.

5.2.2 Extended Yale B

Εξίσου γνωστή βάση δεδομένων που έχει πάλι προέλευση το πανεπιστήμιο του Γέιλ είναι η Extended Yale B. Αυτή η βάση δεδομένων [50] περιέχει 2414 εικόνες προσώπου από μπροστινή όψη, 38 ατόμων. Για κάθε άτομο, περίπου 64 φωτογραφίες ελήφθησαν υπό διάφορες συνθήκες φωτισμού, ελεγχόμενες από το εργαστήριο. Στα πειράματά μας, χρησιμοποιούμε τις περικομμένες εικόνες με ανάλυση 32x32 εικονοστοιχεία. Προεπεξεργαστήκαμε τα δεδομένα κανονικοποιώντας κάθε διάνυσμα προσώπου στη μονάδα. Τριάντα δύο περικομμένα δείγματα εικόνων ενός ατόμου από την βάση δεδομένων Extended Yale B μετά την κανονικοποίηση κλίμακας παρουσιάζονται στην Εικόνα 11.

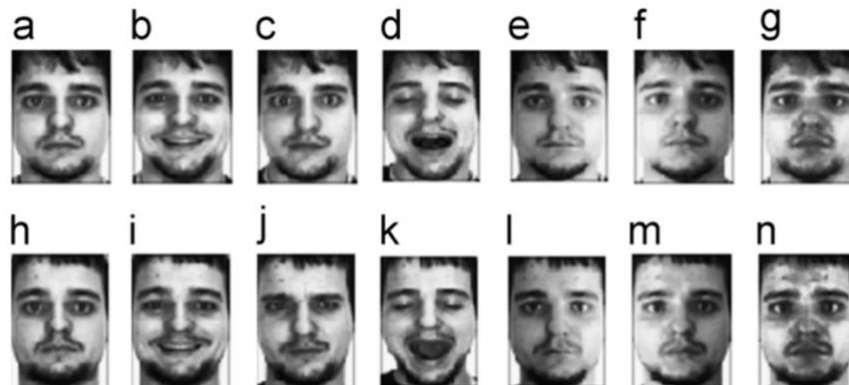


Εικόνα 11. Τριάντα δύο περικομμένα δείγματα εικόνων ενός ατόμου από την βάση δεδομένων Extended Yale B.

5.2.3 AR

Η τελευταία βάση δεδομένων που θα χρησιμοποιήσουμε για να εξετάσουμε την απόδοση των αλγορίθμων μείωσης διάστασης είναι αυτή του AR που κατασκευάστηκε στο Computer Vision Center, καθαρά για ερευνητικούς σκοπούς. Αυτή η βάση δεδομένων περιλαμβάνει πάνω από 4000 εικόνες προσώπων, 126 ατόμων. Για κάθε άτομο, λήφθηκαν 26 φωτογραφίες, σε δύο συνεδρίες, χωρισμένες με δύο εβδομάδες διαφορά, και κάθε τμήμα περιέχει 13 εικόνες. Αυτές οι εικόνες περιλαμβάνουν μπροστινή όψη των προσώπων με διαφορετικές εκφράσεις, φωτισμό και εμπόδια. Στα πειράματά μας εδώ, χρησιμοποιούμε ένα υποσύνολο της βάσης δεδομένων προσώπων AR που παρέχεται από τον Martinez [49], έναν από τους δημιουργούς της. Αυτό το υποσύνολο περιέχει 1400 εικόνες προσώπων που αντιστοιχούν σε 100 άτομα (50 άνδρες και 50 γυναίκες), όπου κάθε άτομο έχει 14

διαφορετικές εικόνες με αλλαγή φωτισμού και εκφράσεων. Η αρχική ανάλυση αυτών των εικόνων προσώπου είναι 165×120 εικονοστοιχεία. Εδώ, για υπολογιστική ευκολία, τις μετατρέπουμε σε 66×48 εικονοστοιχεία, και στη συνέχεια τις κανονικοποιούμε. Οι κανονικοποιημένες εικόνες ενός ατόμου φαίνονται στην Εικόνα 12, όπου οι εικόνες στην πρώτη σειρά προέρχονται από την πρώτη συνεδρία, και οι εικόνες στη δεύτερη σειρά προέρχονται από την δεύτερη συνεδρία. Οι λεπτομέρειες των εικόνων είναι: Εικόνα 12.a ουδέτερη έκφραση, Εικόνα 12.b χαμόγελο, Εικόνα 12.c θυμός, Εικόνα 12.d φωνή, Εικόνα 12.e φως από τα δεξιά, Εικόνα 12.f φως από τα αριστερά, Εικόνα 12.g φως από όλες τις πλευρές, και οι Εικόνες 12.h-n λήφθηκαν υπό τις ίδιες συνθήκες, όπως οι Εικόνες 12.a-g.



Εικόνα 12. Ορισμένα χαρακτηριστικά δείγματα των περικομμένων εικόνων που βρίσκονται στη βάση δεδομένων εικόνων προσώπων AR.

5.2.4 Διαχωρισμός δεδομένων σε εκπαίδευση και δοκιμή

Για αυτές τις βάσεις δεδομένων, επιλέγουμε τυχαία τις μισές εικόνες ανά κατηγορία για εκπαίδευση. Συγκεκριμένα, χρησιμοποιούμε έξι (6) εικόνες ανά άτομο από την βάση δεδομένων Yale για εκπαίδευση, επτά (7) εικόνες ανά άτομο από την βάση δεδομένων AR και περίπου τριάντα δύο (32) εικόνες ανά άτομο από την βάση δεδομένων Extended Yale B, ενώ τις υπόλοιπες τις χρησιμοποιούμε για δοκιμή. Καθώς η βάση δεδομένων AR έχει φυσικά χωριστεί σε δύο συνεδρίες, λαμβάνουμε υπόψη και αυτήν την περίπτωση στα πειράματά μας. Απλά χρησιμοποιούμε το "AR_fixed" για να δηλώσουμε τη βάση δεδομένων AR που χωρίζεται με βάση τις δύο σταθερές συνεδρίες, και το "AR_random" για το ένα που χωρίζεται τυχαία.

Στα πειράματά μας, είκοσι (20) σύνολα εκπαίδευσης/δοκιμής παράγονται τυχαία και αναφέρονται οι μέσες ταξινομητικές ακρίβειες πάνω σε αυτές τις διαχωρίσεις.

5.3 Ταξινόμηση

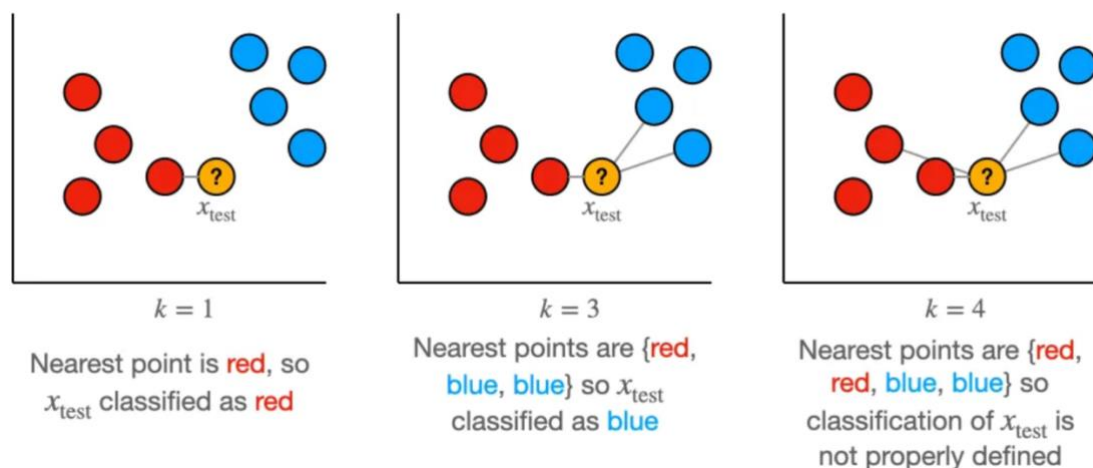
Στον τομέα της αναγνώρισης προσώπων, ο ρόλος ενός ταξινομητή είναι κρίσιμος. Στη βάση του, ένας ταξινομητής εκπαιδεύεται για να διακρίνει διαφορετικά πρότυπα ή κλάσεις σε ένα σύνολο δεδομένων. Για την αναγνώριση προσώπων, αυτές οι κλάσεις

αντιπροσωπεύουν διακριτά άτομα. Μαθαίνοντας τα μοναδικά χαρακτηριστικά και τα χαρακτηριστικά από τα δεδομένα του προσώπου, ένας ταξινομητής μπορεί να αναγνωρίσει ή να επαληθεύσει την ταυτότητα ενός ατόμου, όταν παρουσιάζεται με μια νέα εικόνα προσώπου. Η ακρίβεια αυτής της διαδικασίας αναγνώρισης βασίζεται σε μεγάλο βαθμό στη δυνατότητα του ταξινομητή να γενικεύει από τα δεδομένα εκπαίδευσης σε αδιάκριτα δείγματα. Σε αυτό το κεφάλαιο, εξετάζουμε τους τρεις ταξινομητές που χρησιμοποιούνται στα πειράματά μας: τον ταξινομητή 1-NN, τον ταξινομητή k-NN και τον ταξινομητή SVM, ο καθένας με τα δικά του διακριτά χαρακτηριστικά και μετρικές απόδοσης για την αναγνώριση προσώπων.

5.3.1 Ταξινομητής 1-NN (1-πλησιέστερος γείτονας)

Ο ταξινομητής 1-NN, ή 1-πλησιέστερος γείτονας, είναι μία από τις πιο απλές τεχνικές ταξινόμησης στον τομέα της αναγνώρισης προτύπων. Για την αναγνώριση προσώπων, η αρχή πίσω από τον 1NN είναι απλή: για μια δοθείσα εικόνα προσώπου δοκιμής, ο ταξινομητής βρίσκει το πλησιέστερο ή το πιο παρόμοιο πρόσωπο στο σύνολο εκπαίδευσης και αναθέτει αυτήν την ετικέτα στην εικόνα δοκιμής. Η έννοια της "κοντινότητας" ή "ομοιότητας" καθορίζεται συνήθως από μια μετρική απόστασης, όπως η Ευκλείδεια απόσταση.

Ενώ ο ταξινομητής 1-NN είναι διαισθητικός και εύκολος στην υλοποίηση, μπορεί να είναι υπολογιστικά εντατικός, ειδικά με ένα μεγάλο σύνολο δεδομένων, καθώς απαιτεί σύγκριση με κάθε εκπαιδευτικό δείγμα. Επιπλέον, η απόδοσή του είναι ευαίσθητη στον θόρυβο και τα ακραία δείγματα. Στην Εικόνα 13.α παρουσιάζεται ένα παράδειγμα για τον 1-NN ταξινομητή. Μπορούμε να δούμε πως για το στοιχείο x_{test} που εξετάζουμε, το πιο κοντινό στοιχείο είναι το κόκκινο άρα το x_{test} ταξινομείται στην κόκκινη κλάση.



Εικόνα 13. Παράδειγμα χρησιμοποιώντας τον 1-NN και τον k-NN ταξινομητή για $k=3$ και $k=4$.

5.3.2 Ταξινομητής k-NN (k-πλησιέστεροι γείτονες)

Επεκτείνοντας την έννοια του ταξινομητή 1-NN, ο ταξινομητής k-NN λαμβάνει υπόψη τα 'k' πλησιέστερα δείγματα εκπαίδευσης στην εικόνα προσώπου δοκιμής, για να πραγματοποιήσει την απόφασή του. Η ετικέτα για την εικόνα δοκιμής καθορίζεται συνήθως με πλειοψηφία από αυτούς τους 'k' γείτονες.

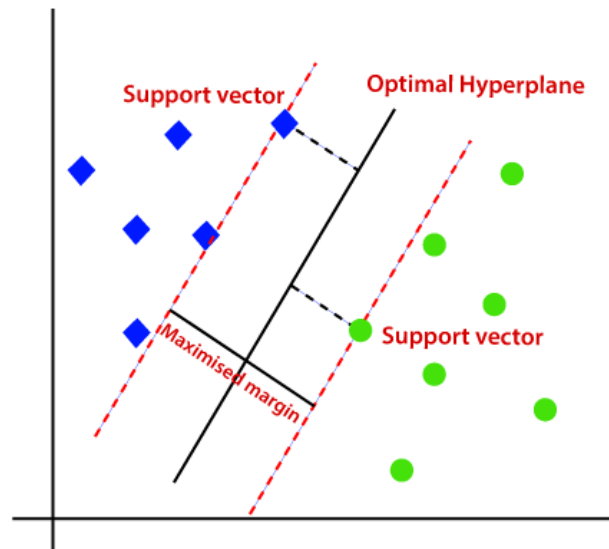
Λαμβάνοντας υπόψη πολλαπλούς γείτονες, ο ταξινομητής k-NN συχνά επιτυγχάνει καλύτερη γενίκευση και ευρωστία σε σύγκριση με το 1-NN, ειδικά σε περιπτώσεις, όπου τα δεδομένα εκπαίδευσης μπορεί να έχουν κάποιο θόρυβο ή ακραίες τιμές.

Η επιλογή του 'k' είναι κρίσιμη. Μια μικρή τιμή μπορεί να καταστήσει τον ταξινομητή ευαίσθητο στο θόρυβο, ενώ μια πολύ μεγάλη τιμή μπορεί να τον καταστήσει υπολογιστικά ακριβό και να μειώσει τη διακριτική του ικανότητα. Στην Εικόνα 13.b και 13.c παρατηρούμε πώς γίνεται η επιλογή της κλάσης για ένα στοιχείο X_{test} με $k=3$ και με $k=4$, αντίστοιχα. Είναι σημαντικό να σημειωθεί ότι στην Εικόνα 13.c, επειδή υπάρχουν δυο κοντινά στοιχεία από κάθε κλάση, ο ταξινομητής μπορεί να δίνει σε κάθε επανεξέταση διαφορετική κλάση ταξινόμησης, κάτι που επιβεβαιώνει ότι η επιλογή του 'k' είναι κρίσιμη για την σωστή ταξινόμηση.

5.3.3 Ταξινομητής SVM (Μηχανή διανυσμάτων στήριξης)

Το Support Vector Machine (SVM) είναι ένας ισχυρός γραμμικός ταξινομητής που λειτουργεί βρίσκοντας το υπερεπίπεδο, το οποίο διαιρεί καλύτερα ένα σύνολο δεδομένων σε κλάσεις. Στο πλαίσιο της αναγνώρισης προσώπων, το SVM στοχεύει να βρει το υπερεπίπεδο που διαχωρίζει καλύτερα τον χώρο χαρακτηριστικών διαφορετικών ατόμων, εξασφαλίζοντας ότι το περιθώριο μεταξύ των κλάσεων είναι μεγιστοποιημένο.

Τα SVMs μπορούν, επίσης, να εξοπλιστούν με συναρτήσεις πυρήνα, επιτρέποντάς τους να αντιμετωπίζουν μη γραμμικά προβλήματα ταξινόμησης. Αυτή η ευελιξία συχνά οδηγεί σε ανώτερες επιδόσεις σε σύνθετα προβλήματα αναγνώρισης. Ωστόσο, η επιλογή του πυρήνα και των παραμέτρων του, παίζουν κρίσιμο ρόλο στην απόδοση του ταξινομητή. Στην παρακάτω Εικόνα 14 παρουσιάζεται ο τρόπος με τον οποίο γίνεται η εύρεση του βέλτιστου υπερεπίπεδου, το οποίο διαχωρίζει καλύτερα τις κλάσεις.



Εικόνα 14. Τρόπος λειτουργίας SVM.

5.4 Μοντέλα επίλυσης

Οι δύο αλγόριθμοι επίλυσης, όπως τους είδαμε στα κεφάλαια 4.1.3 και 4.2.2, για SPP και DSNPE, αντίστοιχα, είναι οι παρακάτω:

Αλγόριθμος 1 Επίλυση SPP μεθόδου.

Είσοδος

- Δεδομένα εκπαίδευσης \mathbf{X} μεγέθους $m \times n$
- Επιθυμητή διάσταση μείωσης d
- Μια σταθερά ανοχής ϵ

Έξοδος

- Πίνακας προβολής μεγέθους $m \times d$

Βήματα αλγορίθμου

1. Αρχικοποίηση

- Πάρε το μέγεθος του πίνακα \mathbf{X} σε $m \times n$
- Αρχικοποίησε τον πίνακα βάρους \mathbf{S} με μηδενικά, μεγέθους $n \times n$

2. Υπολογισμός διανύσματος προβολής

Για κάθε σημείο δεδομένων i στο εύρος n :

- Όρισε \mathbf{x}_i ως τη i^{th} στήλη του \mathbf{X} (τρέχον σημείο δεδομένων)
 - Αρχικοποίησε μια μεταβλητή \mathbf{s}_i ως sparse coefficient vector
 - Όρισε αντικειμενική συνάρτηση για ελαχιστοποίηση της $L1$ νόρμας του \mathbf{s}_i
 - Καθόρισε τους περιορισμούς:
 - α) Το $\|\mathbf{x}_i - \mathbf{X} \times \mathbf{s}_i\|$ να είναι μικρότερο ή ίσο με την σταθερά ϵ
 - β) Το άθροισμα των τιμών του \mathbf{s}_i να είναι ίσο με την μονάδα
 - Λύσε το πρόβλημα ελαχιστοποίησης $L1$ χρησιμοποιώντας τον λύτη *ECOS* από την βιβλιοθήκη *cvxpy*
 - Ενημερώστε τη i^{th} στήλη του \mathbf{S} με την τιμή του \mathbf{s}_i
-

3. Εξαγωγή ιδιοδιανύσματος

- Υπολόγισε το S_β ως $S + S^T - S^T \times S$
- Εξάγαγε ιδιοτιμές και ιδιοδιανύσματα του πίνακα $X \times S_\beta \times X^T$

4. Προβολή υπόχωρου

- Ταξινόμησε τις ιδιοτιμές με φθίνουσα σειρά και αποθήκευσε τους δείκτες
- Επίλεξε τις κορυφαίες d ιδιοτιμές με βάση τους ταξινομημένους δείκτες.
- Όρισε τα επιλεγμένα διανύσματα, επιλέγοντας στήλες που αντιστοιχούν στις κορυφαίες d ιδιοτιμές από τον πίνακα του ιδιοδιανύσματος
- Μετάτρεψε τα επιλεγμένα διανύσματα σε πραγματικές τιμές

5. Επιστροφή

- του πίνακα προβολής των επιλεγμένων διανυσμάτων
-

Αλγόριθμος 2 Επίλυση DSNPE μεθόδου

Είσοδος

- Δεδομένα εκπαίδευσης X μεγέθους $n \times p$
- Ετικέτες y
- Επιθυμητή διάσταση μείωσης d
- Μια σταθερά ανοχής $epsilon$
- Παράμετρος κανονικοποίησης γ

Έξοδος

- Μετασχηματισμένος πίνακας δεδομένων Y και πίνακας προβολής

Βήματα αλγορίθμου

1. Συνάρτηση για τον υπολογισμό s_i

1.1. Είσοδος: Τρέχον σημείο δεδομένων x_i , πίνακας X_i και σταθερά $epsilon$

1.2. Έξοδος: Sparse coefficient vector s_i

- Πάρε τον αριθμό των στηλών n του πίνακα X_i
- Αρχικοποίησε μια μεταβλητή s_i με διάσταση n
- Όρισε αντικειμενική συνάρτηση για ελαχιστοποίηση της $L1$ νόρμας του s_i
- Καθόρισε τον περιορισμό:
 - α) Το $\|x_i - X \times s_i\|$ να είναι μικρότερο ή ίσο με την σταθερά $epsilon$
- Λύσε το πρόβλημα ελαχιστοποίησης $L1$ χρησιμοποιώντας τον λύτη *ECOS* από την βιβλιοθήκη *cvxpy*
- Επέστρεψε την τιμή του s_i

2. Κύρια συνάρτηση DSNPE

2.1. Προεπεξεργασία

- Πάρε το μέγεθος του πίνακα X σε $n \times p$
- Αρχικοποίησε τον πίνακα S με μηδενικά, μεγέθους $n \times n$

Για κάθε σημείο δεδομένων i στο εύρος n :

- Όρισε x_i ως τη i^{th} στήλη του X (τρέχον σημείο δεδομένων)
- Πάρε τον πίνακα X_{i_k} των δειγμάτων που έχουν την ίδια ετικέτα όπως το x_i
- Υπολόγισε s_i χρησιμοποιώντας τη συνάρτηση που ορίστηκε στο Βήμα 1 και ενημερώστε τον πίνακα S

2.2. Υπολογισμός του πίνακα $S_\alpha = I - S - S^T + S^T S$

2.3. Υπολογισμός του πίνακα $X_{sa} = X^T \times S_\alpha \times X$

2.4. Υπολογισμός των μεταξύ-κλάσης και εντός-κλάσης πινάκων διασποράς

- Αρχικοποίησε τον πίνακα \mathbf{X}_{sbt} με μηδενικά, μεγέθους $p \times p$
- Για κάθε κλάση k στις μοναδικές ετικέτες του y :
- Υπολόγισε και πρόσθεσε τη διασπορά των δειγμάτων που ανήκουν στην κλάση k στο \mathbf{X}_{sbt}
- Υπολόγισε τον εντός-κλάσης πίνακα διασποράς \mathbf{X}_{sw}
- Υπολόγισε τον μεταξύ-κλάσης πίνακα διασποράς \mathbf{X}_{sb} χρησιμοποιώντας το \mathbf{X}_{sbt} και το \mathbf{X}_{sw}

2.5. Εξαγωγή ιδιοτιμών και ιδιοδιανυσμάτων $(\mathbf{X} \times (\mathbf{X}_{sa} - \gamma \times \mathbf{X}_{sb}))^T \times \mathbf{X}$

2.6. Προβολή υπόχωρου

- Ταξινόμησε τις ιδιοτιμές με αύξουσα σειρά και επιλέξτε τους κορυφαίους d δείκτες
- Πάρε τον πίνακα προβολής χρησιμοποιώντας τους επιλεγμένους δείκτες από τον πίνακα ιδιοδιανύσματος

2.7. Μετασχηματισμός δεδομένων

- Μετασχημάτισε τα δεδομένα \mathbf{X} χρησιμοποιώντας τον πίνακα προβολής για να πάρεις το Y

2.8. Επιστροφή

- του πίνακα προβολής και των μετασχηματισμένων δεδομένων Y
-

Οι δύο παραπάνω αλγόριθμοι είναι φανερό ότι έχουν αρκετές ομοιότητες μεταξύ τους, με τον DSNPE να είναι αρκετά πιο περίπλοκος.

5.4.1 Επιλογή παραμέτρων

Για την υλοποίηση και των δύο μεθόδων χρησιμοποιήθηκαν οι πιο σταθερές αντικειμενικές συναρτήσεις, δηλαδή η εξίσωση (19) για τον SPP και η εξίσωση (33) για τον DSNPE, αντίστοιχα. Αυτό έγινε γιατί παρατηρήσαμε ότι, όταν χρησιμοποιούμε την σταθερά ανοχής λάθους *epsilon*, έχουμε καλύτερα αποτελέσματα, από όταν παίρνουμε για περιορισμό απλά την $\mathbf{x}_i = \mathbf{X} \times \mathbf{s}_i$. Είναι επίσης σημαντικό να αναφέρουμε ότι η σταθερά αυτή είχε, καθ' όλη τη διάρκεια των πειραμάτων μας, την ίδια τιμή *epsilon* = 0.05 για τον SPP και *epsilon* = 0.01 για τον DSNPE. Για το PCA η μόνη παράμετρος του μοντέλου είναι η διάσταση του υποχώρου, αντίστοιχα ορίζεται και το LDA. Για το LPP και το LLE, οι παράμετροι του κάθε μοντέλου περιλαμβάνουν την επιλογή των γειτόνων n και την διάσταση του υποχώρου.

Για τους ταξινομητές k-NN και SVM η επιλογή των παραμέτρων έγινε με την διαδικασία της διασταυρούμενης επικύρωσης (cross-validation). Πιο συγκεκριμένα, για τον k-NN η επιλογή γειτόνων έγινε από μια λίστα με εύρος από 1-20 γείτονες, για τα βάρη η αναζήτηση έγινε μεταξύ του 'uniform' και του 'distance', και για τον αλγόριθμο είχαμε είτε το 'auto', είτε το 'ball_tree', είτε το 'kd_tree' ή το 'brute'. Για το SVM οι παράμετροι είχαν ως εξής, ο πυρήνας (kernel) επιλέχθηκε από τα 'linear', 'rbf', 'poly' και 'sigmoid'. Η παράμετρος

κανονικοποίησης C δοκιμάστηκε για 0.1, 1, 10 και 100, εξίσου ίδιες τιμές είχε και το γ , με επιπρόσθετα τα *'scale'* και *'auto'*.

5.4.2 Λεπτομέρειες υλοποίησης – Λογισμικό και βιβλιοθήκες

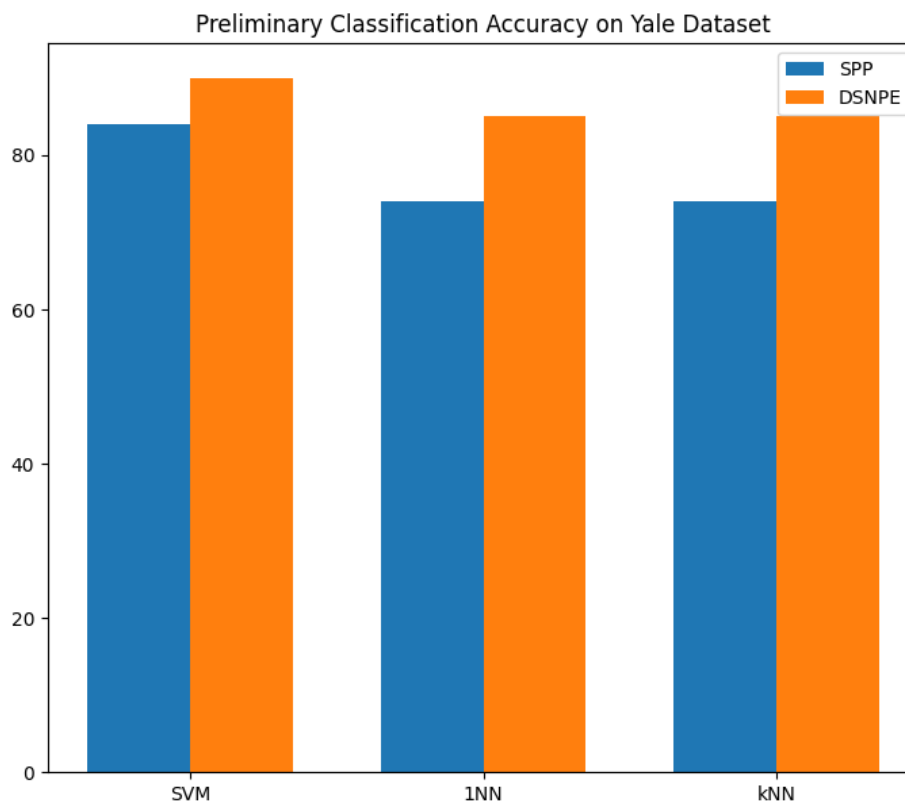
Κατά τη διάρκεια της διεξαγωγής των πειραμάτων, χρησιμοποιήσαμε τη γλώσσα προγραμματισμού Python, γνωστή για την αξιοπιστία της στην ανάλυση δεδομένων και τις εργασίες μηχανικής μάθησης. Για να διευκολύνουμε την ανάπτυξη και την εκτέλεση των αλγορίθμων μείωσης διάστασης, χρησιμοποιήθηκε μια σειρά εξειδικευμένων βιβλιοθηκών. Η βιβλιοθήκη *'scikit-learn'*, ένας ακρογωνιαίος λίθος στον τομέα της μηχανικής μάθησης, προσέφερε μια σειρά εργαλείων για την προεπεξεργασία δεδομένων, την εκπαίδευση μοντέλων και την αξιολόγηση. Για αριθμητικές λειτουργίες και διαχείριση δεδομένων, η *'numpy'* ήταν απαραίτητη, προσφέροντας αποτελεσματική διαχείριση μεγάλων παρατάξεων δεδομένων (data arrays) και πινάκων. Η βιβλιοθήκη *'os'* διευκόλυνε τη διαχείριση αρχείων και καταλόγων, εξασφαλίζοντας ομαλή εισαγωγή και εξαγωγή δεδομένων. Οι εργασίες επεξεργασίας και διαχείρισης οπτικών δεδομένων πραγματοποιήθηκαν με τη χρήση του *'cv2'*, ενός μοντέλου OpenCV προσαρμοσμένου για εφαρμογές υπολογιστικής όρασης. Για τη οπτικοποίηση των αποτελεσμάτων και των κατανομών δεδομένων, χρησιμοποιήθηκε το *'matplotlib'*, επιτρέποντας τη δημιουργία διασθητικών γραφημάτων και διαγραμμάτων. Τέλος, για τα προβλήματα βελτιστοποίησης, χρησιμοποιήθηκε το *'cvxpy'*, προσφέροντας μια πλατφόρμα για τον καθορισμό και την επίλυση προβλημάτων κυρτής βελτιστοποίησης χωρίς δυσκολία. Συνολικά, αυτές οι βιβλιοθήκες αποτέλεσαν τον κορμό της υποδομής λογισμικού, επιτρέποντας έναν γραμμικό και αποτελεσματικό πειραματικό ρυθμό εργασίας.

5.4.3 Εκπαίδευση μοντέλων

Όπως έχει ήδη συζητηθεί στην διπλωματική αυτή, και οι δύο μέθοδοι του Sparsity Preserving Projections (SPP) και του Discriminant Sparse Neighborhood Preserving Embedding (DSNPE) περιλαμβάνουν ένα μοναδικό αρχικό βήμα προβολής των συνόλων δεδομένων σε έναν υπόχωρο PCA. Αυτή η προεπεξεργασία PCA δεν είναι μια ενιαία λύση για όλες τις περιπτώσεις, αλλά προσαρμόζεται ανάλογα με το συγκεκριμένο σύνολο δεδομένων που χρησιμοποιείται. Για το σύνολο δεδομένων Yale, εκτελείται πλήρης PCA ορίζοντας την παράμετρο *n_components* σε None, διατηρώντας έτσι όλα τα διαθέσιμα χαρακτηριστικά. Αντίθετα, για τα σύνολα δεδομένων Extended Yale B και AR, χρησιμοποιείται μια αναλογία PCA του 0,98, η οποία στοχεύει στη διατήρηση του 98% της διακύμανσης στο σύνολο δεδομένων. Αυτή η μειωμένη διαστατικότητα διευκολύνει την υπολογιστική αποδοτικότητα και βοηθά στην εξάλειψη τυχόν θορύβου στα σύνολα δεδομένων. Ο χώρος χαρακτηριστικών του συνόλου δεδομένων Yale επιτρέπει τη χρήση μιας πλήρους PCA, χωρίς να θέτει σε κίνδυνο την υπολογιστική ή την απόδοση της ταξινόμησης, ενώ τα σύνολα δεδομένων Extended Yale B και AR επωφελούνται από την ελαφριά μείωση της διαστατικότητας. Αυτή η προεπεξεργασία PCA δεν υπηρετεί μόνο την υπολογιστική ευκολία, αλλά είναι και ένα κρίσιμο βασικό βήμα για την εφαρμογή του SPP

και του DSNPE, επηρεάζοντας, τόσο το υπολογιστικό κόστος, όσο και τη συνολική αποτελεσματικότητα των ταξινομητών.

Μετά την προεπεξεργασία PCA, προχωρούμε στους κύριους αλγορίθμους των ανακατασκευασμένων μεθόδων μας, SPP και DSNPE. Οι διαδικασίες εκπαίδευσης και οι ρυθμίσεις υπερπαραμέτρων για τους ταξινομητές SVM, 1-NN, και k-NN, είναι όπως έχουν λεπτομερώς περιγραφεί σε παραπάνω.



Εικόνα 15. Απόδοση SPP και DSNPE στη βάση δεδομένων Yale στους τρεις ταξινομητές SVM, 1-NN και k-NN.

Όπως φαίνεται στην Εικόνα 15, και τα δύο SPP και DSNPE δείχνουν ανταγωνιστική απόδοση στο σύνολο δεδομένων Yale, χρησιμοποιώντας τους ταξινομητές SVM, 1-NN, και k-NN. Για παράδειγμα, το SPP επιτυγχάνει μέση ταξινομητική ακρίβεια της τάξεως του 86% χρησιμοποιώντας τον SVM, ενώ το DSNPE αναφέρει μέση ακρίβεια της τάξεως του 90%.

Αυτά τα αρχικά ευρήματα διασφαλίζουν, όχι μόνο τη μοναδική προεπεξεργασία και τα βήματα εκπαίδευσης, αλλά προετοιμάζουν και το έδαφος για μια πιο ολοκληρωμένη ανάλυση, η οποία θα παρουσιαστεί στο επόμενο κεφάλαιο 'Αποτελέσματα'.

5.5 Αποτελέσματα

Σε αυτό το κρίσιμο κεφάλαιο, βυθιζόμαστε στην εμπειρική αξιολόγηση της μελέτης μας, επικεντρώνοντας το φως στην αποτελεσματικότητα και τις λεπτομέρειες κάθε τεχνικής

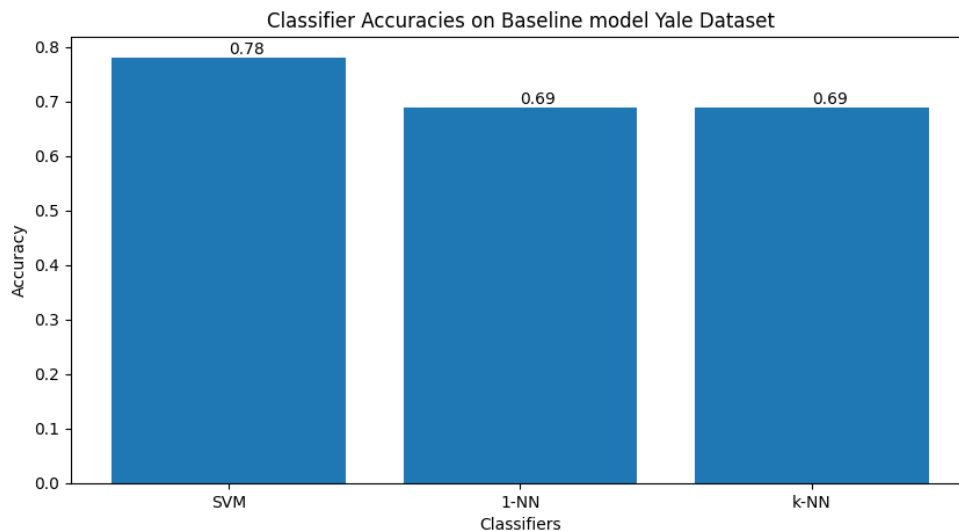
μείωσης της διαστατικότητας που λαμβάνεται υπόψη. Εναντία στο υπόβαθρο πολλαπλών συνόλων δεδομένων και ταξινομητών, αυτά τα αποτελέσματα προσφέρουν μια πολύπλευρη οπτική γωνία στο τοπίο των επιδόσεων των σύγχρονων μεθόδων μείωσης της διαστατικότητας. Από τις παραδοσιακές προσεγγίσεις, όπως η ανάλυση κύριων συνιστωσών (PCA) και η γραμμική διακριτική ανάλυση (LDA), έως τις κεντρικές μεθόδους της μελέτης μας, τις προβολές διατήρησης αραιότητας (SPP) και τη διακριτική αραιή ενσωμάτωση που διατηρεί τη γειτονιά (DSNPE), τα επόμενα υποτμήματα παρέχουν μια εξαντλητική και λεπτομερή παρουσίαση. Επιπλέον, σε μια προσπάθεια να παράσχουμε μια ολιστική κατανόηση, συμπεριλαμβάνουμε αποτελέσματα βάσης (baseline) ως αναφορικό σημείο σύγκρισης. Αυτή η ολοκληρωμένη αξιολόγηση, όχι μόνο επικυρώνει τις θεωρητικές βάσεις που συζητήθηκαν νωρίτερα, αλλά επιπλέον φωτίζει την εφαρμογή και τους περιορισμούς στον πραγματικό κόσμο. Το τμήμα κορυφώνεται σε μια συγκριτική ανάλυση που συνθέτει αυτές τις πληροφορίες, ακολουθούμενη από μια συζήτηση που ερμηνεύει τα δεδομένα στο ευρύτερο πλαίσιο της έρευνας για τη μείωση της διαστατικότητας.

5.5.1 Βάση (baseline)

Σε κάθε εμπειρική αξιολόγηση, η θέσπιση μιας βάσης είναι αναπόφευκτο βήμα. Θέτει το πλαίσιο για μια ρεαλιστική σύγκριση, παρέχοντας ένα μετρικό απόδοσης 'χωρίς πολλά πολλά' που παρουσιάζει τι μπορεί να επιτευχθεί χωρίς τη βοήθεια εξειδικευμένων τεχνικών μείωσης της διαστατικότητας. Για την παρούσα μελέτη, η βάση περιλαμβάνει την εκτέλεση των ταξινομητών μας - SVM, 1-NN, και k-NN - απευθείας στα αρχικά σύνολα δεδομένων: Yale, Extended Yale B και AR. Αυτό λειτουργεί ως η κατάσταση ελέγχου μας, μια απλή αλλά αποκαλυπτική άσκηση που θα παρέχει το πλαίσιο για τα αυξητικά κέρδη ή τις απώλειες που προκύπτουν από κάθε επόμενη μέθοδο. Ακολουθούν τα αποτελέσματα των ταξινομητών μας που λειτουργούν σε αυτόν τον αμετάβλητο χώρο χαρακτηριστικών, τα οποία θα χρησιμεύσουν αργότερα ως σημείο σύγκρισης για τις προηγμένες μεθόδους που εξετάζονται.

Yale

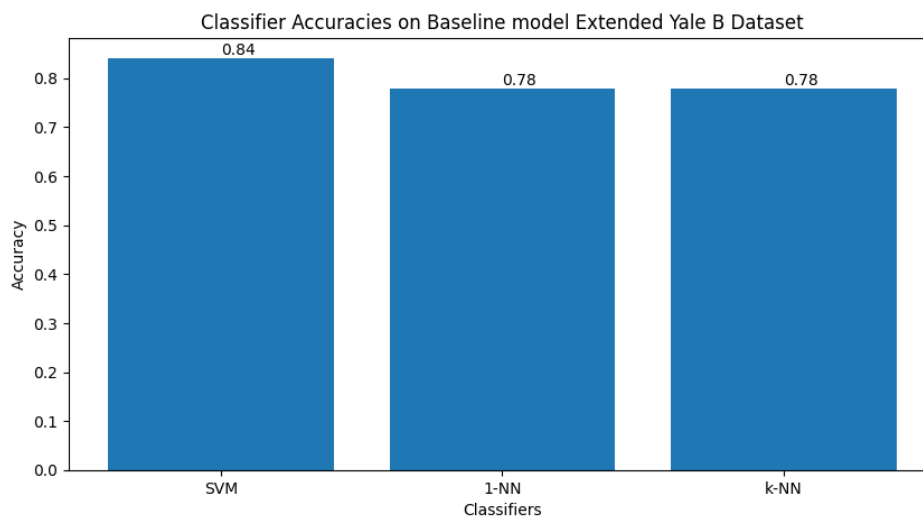
Στην Εικόνα 16 παρουσιάζονται τα αποτελέσματα της βάσης στους τρεις ταξινομητές για την βάση δεδομένων του Yale. Στο SVM παρατηρούμε απόδοση της τάξης του 78%, ενώ για το 1-NN και το k-NN έχουμε κοινή απόδοση στο 69%. Αυτά τα αποτελέσματα αποτελούν μια αρκετά καλή βάση.



Εικόνα 16. Απόδοση baseline στη βάση δεδομένων Yale στους τρεις ταξινομητές SVM, 1-NN και k-NN.

Extended Yale B

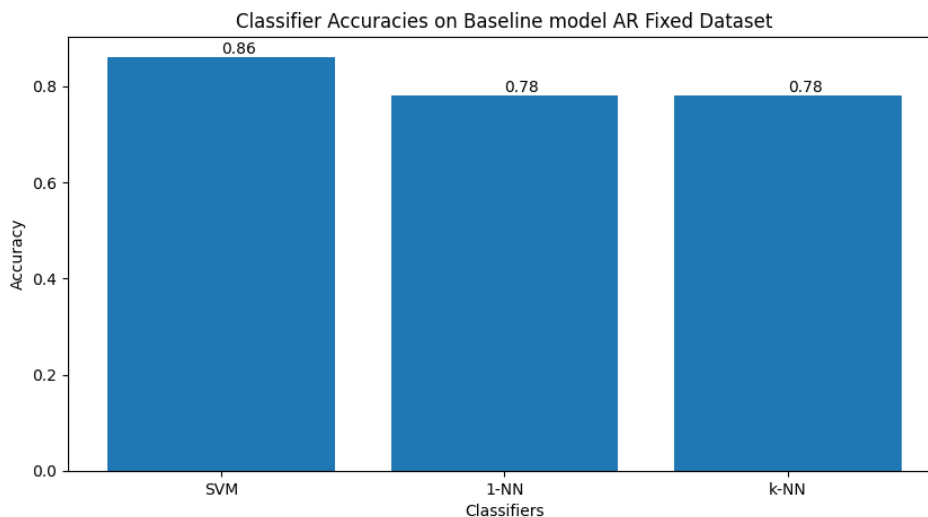
Επόμενη βάση είναι το Extended Yale B, όπου εδώ παρατηρούμε στην Εικόνα 17 μεγαλύτερη απόδοση για την βάση μας, με το SVM να επιτυγχάνει ποσοστό αναγνώρισης 84% και τα 1-NN και k-NN ποσοστό 78%.



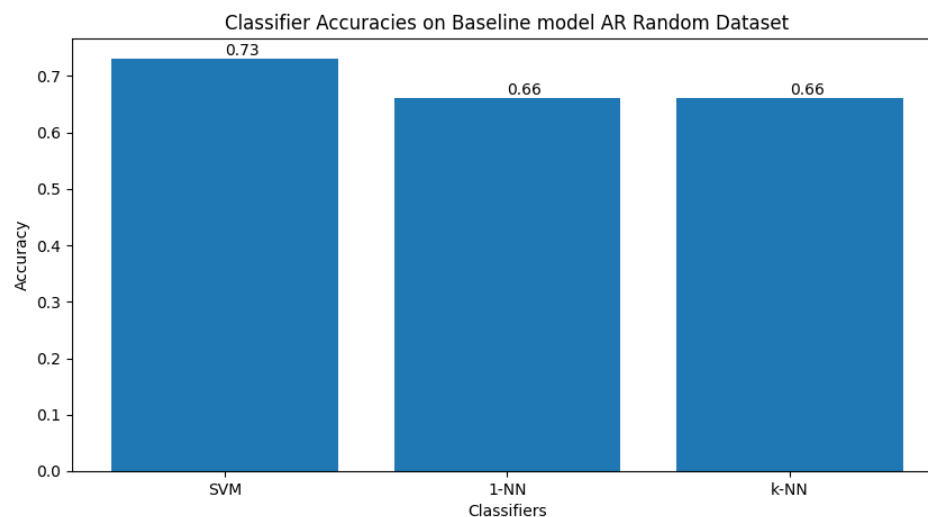
Εικόνα 17. Απόδοση baseline στη βάση δεδομένων Extended Yale B στους τρεις ταξινομητές SVM, 1-NN και k-NN.

AR fixed - AR random

Για την βάση δεδομένων του AR, ο τρόπος διαχωρισμού των δεδομένων σε fixed και random αποτελεί κρίσιμο στοιχείο, όπως μπορούμε να παρατηρήσουμε στις αντίστοιχες αποδόσεις στην Εικόνα 18 για AR fixed και Εικόνα 19 για AR random. Με το SVM να έχει απόδοση 86% για τον fixed διαχωρισμό και 73% για τον random, και τα 1-NN/k-NN 78% για fixed και 66% για random, αντίστοιχα.



Εικόνα 18. Απόδοση baseline στη βάση δεδομένων AR fixed στους τρεις ταξινομητές SVM, 1-NN και k-NN.



Εικόνα 19. Απόδοση baseline στη βάση δεδομένων AR random στους τρεις ταξινομητές SVM, 1-NN και k-NN.

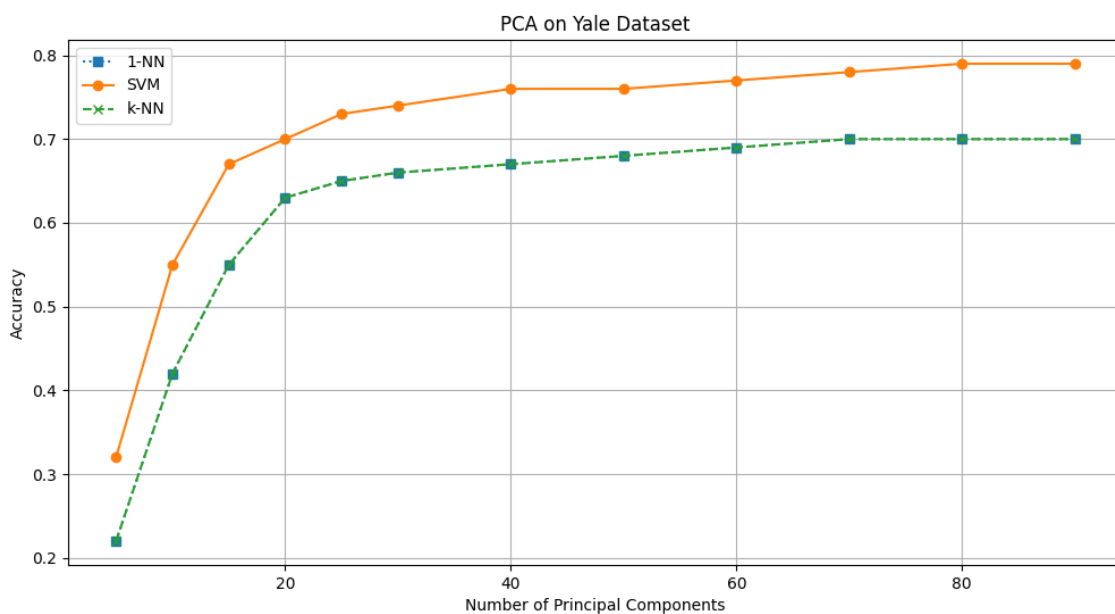
5.5.2 Ανάλυση κύριων συνιστωσών (PCA)

Έχοντας αναλύσει τις μετρικές επίδοσης της βάσης, στρέφουμε τώρα την προσοχή μας σε μία από τις πιο εδραιωμένες τεχνικές στον τομέα της μείωσης της διαστατικότητας, την ανάλυση κύριων συνιστωσών (PCA). Λειτουργώντας ως ένα θεμελιώδες στοιχείο, τόσο στη μηχανική μάθηση, όσο και στη στατιστική, το PCA επιδιώκει την προβολή των δεδομένων σε έναν υποχώρο μειωμένης διαστατικότητας, διατηρώντας όσο το δυνατόν περισσότερη διακύμανση. Αυτή η ενότητα θα εξετάσει τον αντίκτυπο της εφαρμογής του PCA ως βήμα προεπεξεργασίας, προτού εκτελέσουμε τους ταξινομητές μας στα σύνολα δεδομένων. Τα αποτελέσματα αυτής της ενότητας θα λειτουργήσουν, όχι μόνο ως ένα σημαντικό σημείο σύγκρισης για τις κύριες μεθόδους μας, τα SPP και τα DSNPE, αλλά θα προσφέρουν επίσης

επιστημονικά στοιχεία για την αποτελεσματικότητα των παραδοσιακών τεχνικών μείωσης της διαστατικότητας.

Yale

Στην Εικόνα 20 παραθέτουμε ένα γράφημα για την εξέλιξη της απόδοσης, σε σύγκριση με την αντίστοιχη διάσταση μείωσης, χρησιμοποιώντας το PCA στη βάση δεδομένων του Yale. Από το γράφημα μπορούμε να δούμε ότι η μέγιστη απόδοση επιτυγχάνεται για το SVM στη διάσταση 80 με ποσοστό 79%, ενώ για τα 1-NN/k-NN επιτυγχάνεται στην διάσταση 70, αλλά με μικρότερο ποσοστό 70%.

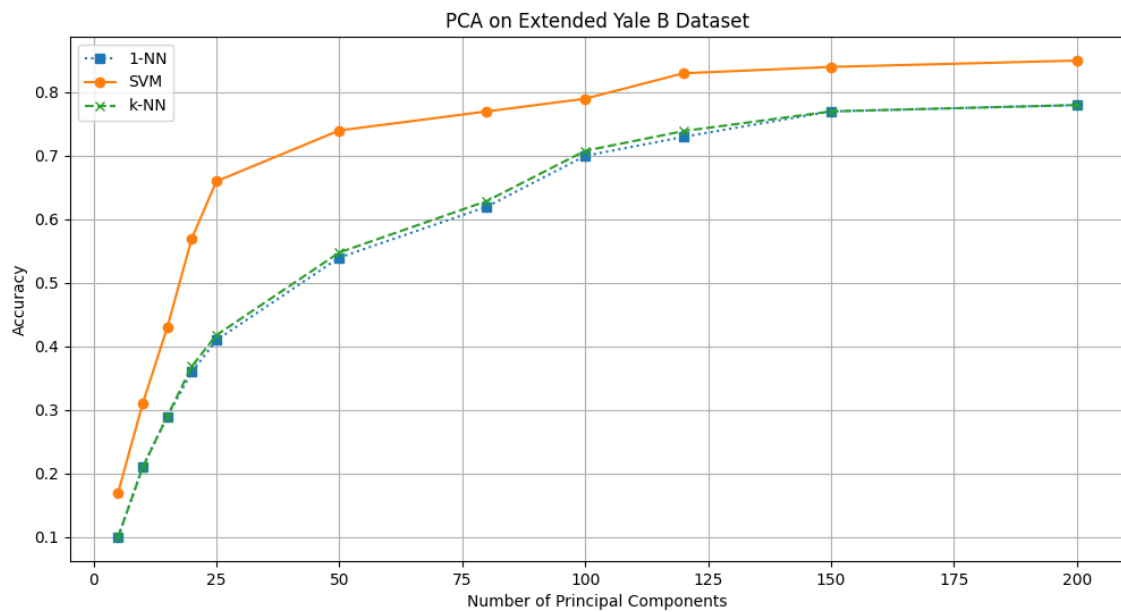


Εικόνα 20. Γράφημα απόδοσης – διάστασης, στο Yale χρησιμοποιώντας την μέθοδο PCA.

Είναι σημαντικό να παρατηρήσουμε ότι οι ταξινομητές 1-NN (μπλε χρώμα) και k-NN (πράσινο χρώμα) συγκλίνουν στην ίδια πορεία, κάτι που υποδηλώνει ότι ο καλύτερος αριθμός γειτόνων για το k-NN μετά την διασταυρωμένη επικύρωση είναι ο ένας γείτονας. Αυτό είναι κάτι που παρατηρείται και στις υπόλοιπες μεθόδους, καθ' όλη τη διάρκεια των πειραμάτων μας.

Extended Yale B

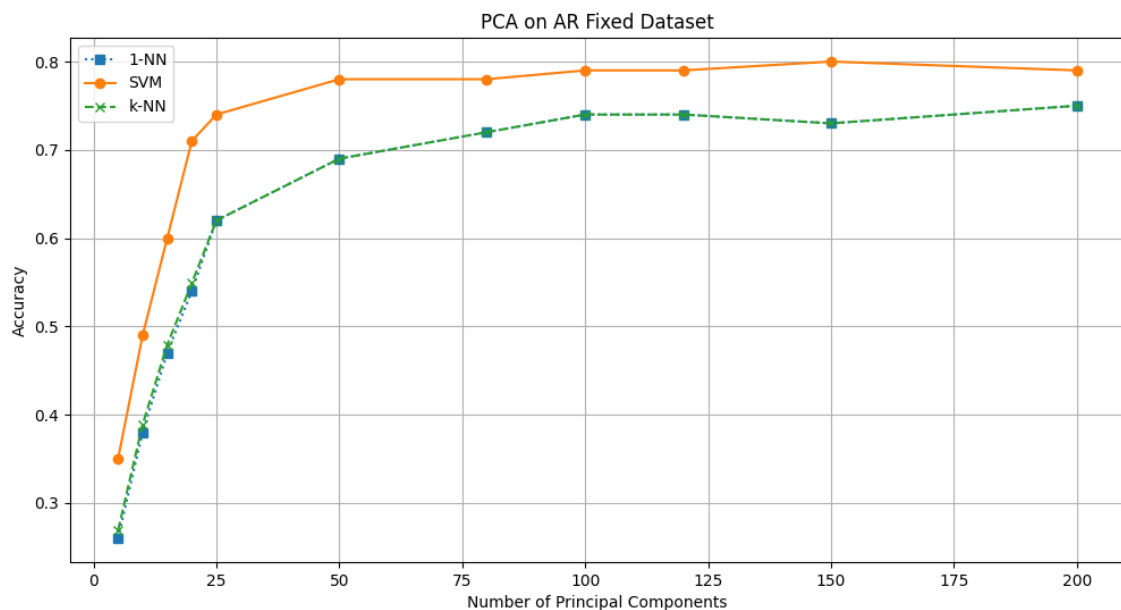
Στην παρακάτω Εικόνα 21 παρατηρούμε τις καμπύλες απόδοσης του PCA στο Extended Yale B. Η μέγιστη απόδοση επιτυγχάνεται στην διάσταση 200 και για τους τρεις ταξινομητές, με τον SVM να έχει ποσοστό 85% και τα 1-NN/k-NN 79%.



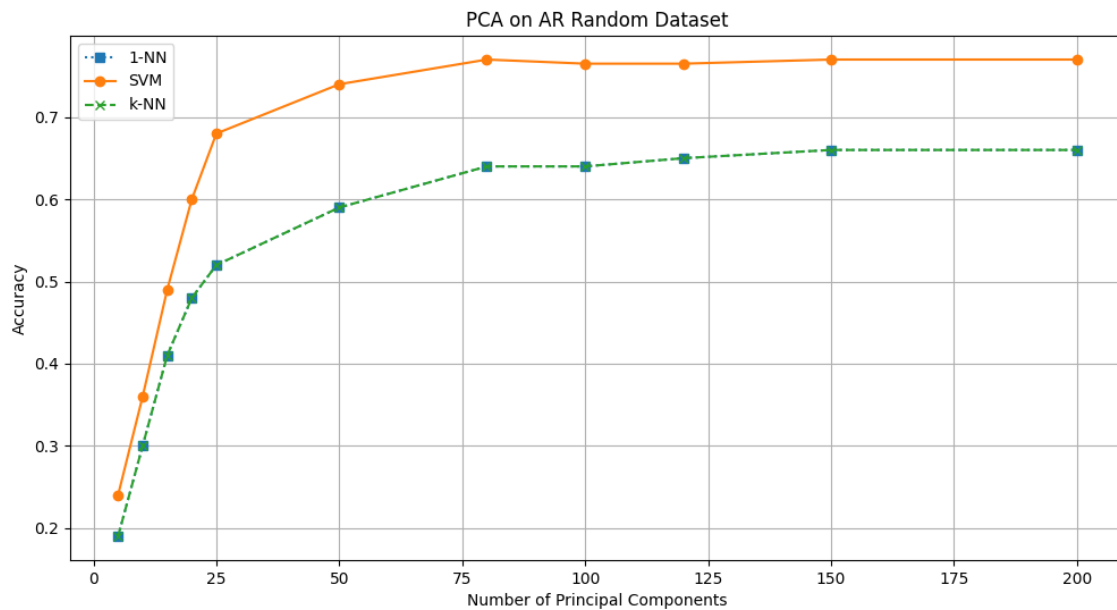
Εικόνα 21. Γράφημα απόδοσης – διάστασης, στο Extended Yale B χρησιμοποιώντας την μέθοδο PCA.

AR fixed - AR random

Στις εικόνες 22 και 23 δίνονται τα γραφήματα για AR fixed και AR random με χρήση PCA. Ο SVM έχει μέγιστη απόδοση 80% στην διάσταση 150 για τον fixed διαχωρισμό και στην ίδια διάσταση 77% στο random. Αντίστοιχα, στα 1-NN/k-NN, επιτυγχάνει μέγιστη απόδοση 75% στην διάσταση 200 στον fixed και 66% στην διάσταση 150 στο random.



Εικόνα 22. Γράφημα απόδοσης – διάστασης, στο AR fixed χρησιμοποιώντας την μέθοδο PCA.



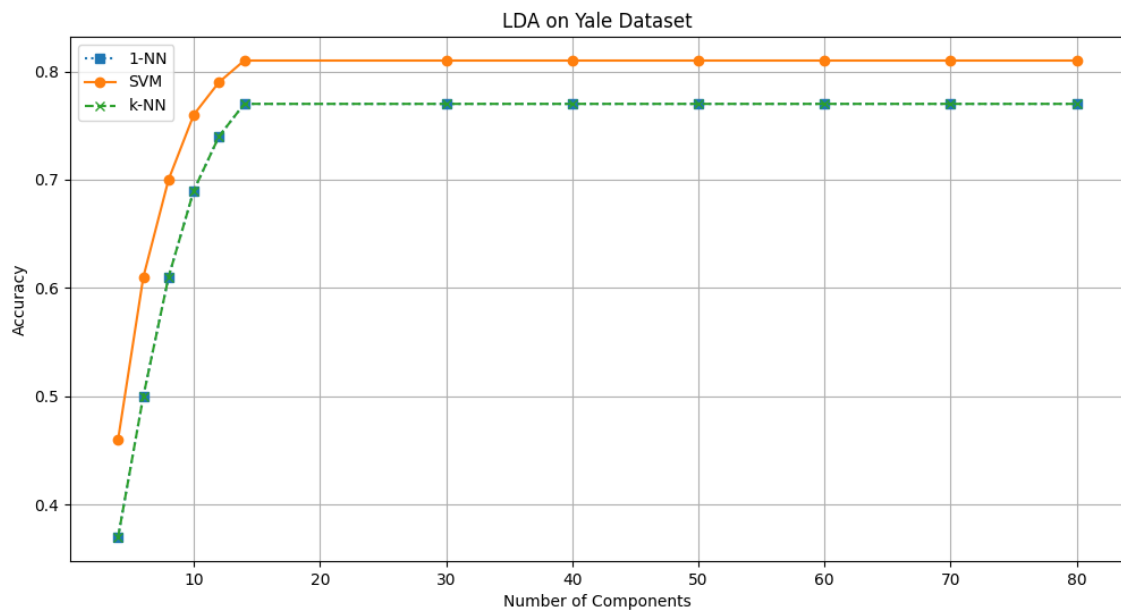
Εικόνα 23. Γράφημα απόδοσης – διάστασης, στο AR random χρησιμοποιώντας την μέθοδο PCA.

5.5.3 Γραμμική διακριτική ανάλυση (LDA)

Μεταβαίνοντας από το PCA, εξετάζουμε στη συνέχεια την γραμμική διακριτική ανάλυση (LDA), μια άλλη κλασική μέθοδος με διαφορετικό όμως επίκεντρο στη μεγιστοποίηση της διαχωριστικότητας των κλάσεων. Σε αντίθεση με το PCA, το οποίο είναι μη επιβλεπόμενο, το LDA λαμβάνει υπόψη τις ετικέτες των κλάσεων για να βελτιστοποιήσει τη διακύμανση μεταξύ των κλάσεων, ενώ ελαχιστοποιεί τη διακύμανση εντός των κλάσεων. Αυτή η ενότητα θα αναλύσει τα αποτελέσματα που προκύπτουν από την εφαρμογή της LDA στα σύνολα δεδομένων μας, Yale, Extended Yale B και AR.

Yale

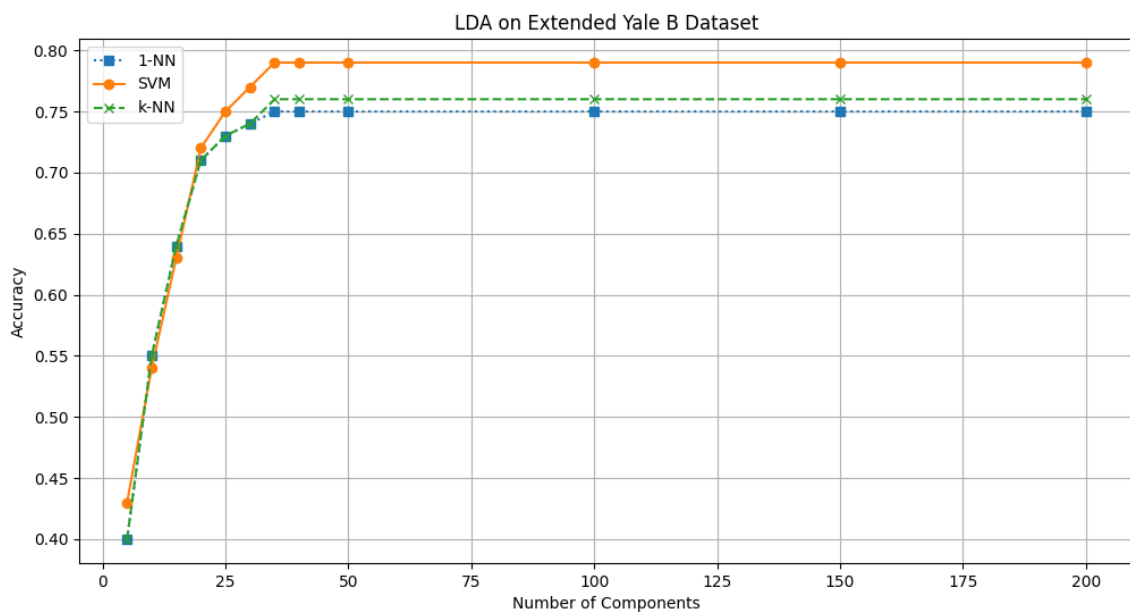
Τα αποτελέσματά μας για το LDA στο Yale παρουσιάζονται στην Εικόνα 24. Για το SVM έχουμε μέγιστη απόδοση 81% στην διάσταση 14, ενώ για τα 1-NN/k-NN έχουμε απόδοση 77% για την ίδια διάσταση.



Εικόνα 24. Γράφημα απόδοσης – διάστασης, στο Yale χρησιμοποιώντας την μέθοδο LDA.

Extended Yale B

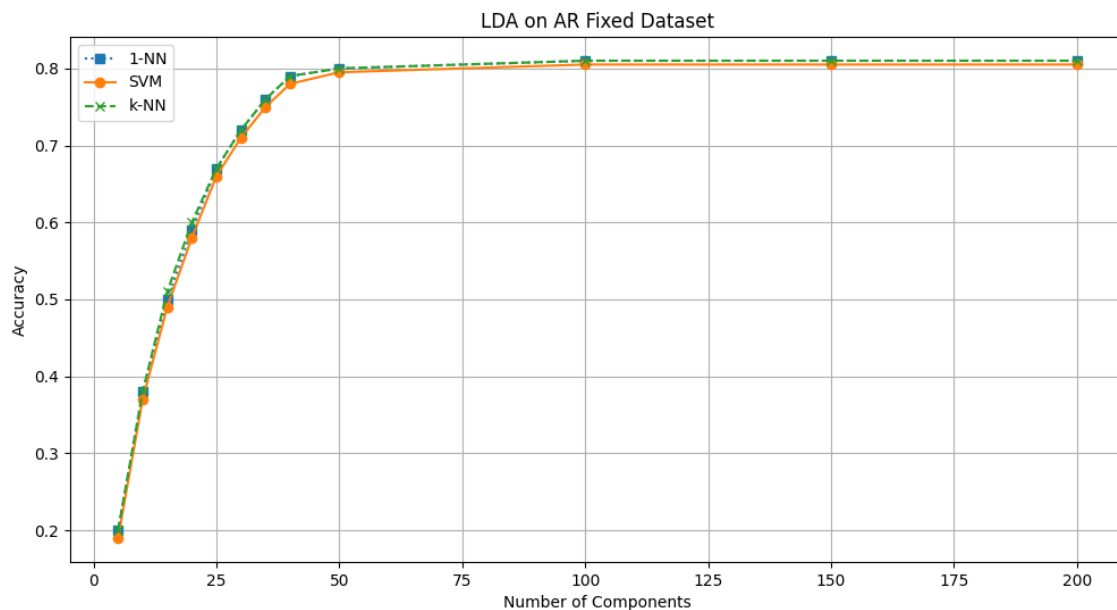
Στην Εικόνα 25 παρουσιάζονται τα αποτελέσματα για τον Extended Yale B με την χρήση του LDA. Μέγιστη απόδοση της τάξης του 79% για το SVM στην διάσταση 35, για το 1-NN έχουμε απόδοση 75% ενώ για το k-NN 76%, για την ίδια διάσταση, 35.



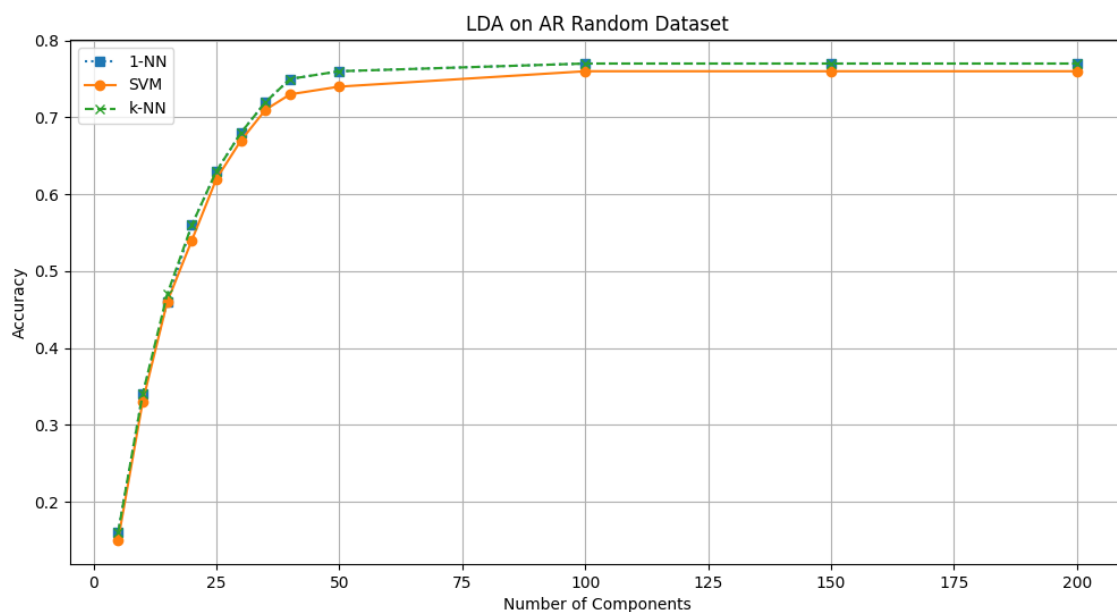
Εικόνα 25. Γράφημα απόδοσης – διάστασης, στο Extended Yale B χρησιμοποιώντας την μέθοδο LDA.

AR fixed - AR random

Για τον AR fixed και τον AR random στο LDA είναι οι μοναδικές φορές που παρατηρούμε μεγαλύτερη απόδοση στους ταξινομητές 1-NN και k-NN, με αποδόσεις 81% (fixed) και 77% (random) για διάσταση 100, σε σύγκριση με τον SVM που επιτυγχάνει 80% (fixed) και 76% (random) αντίστοιχα στην ίδια διάσταση, όπως φαίνεται στις εικόνες 26 και 27.



Εικόνα 26. Γράφημα απόδοσης – διάστασης, στο AR fixed χρησιμοποιώντας την μέθοδο LDA.



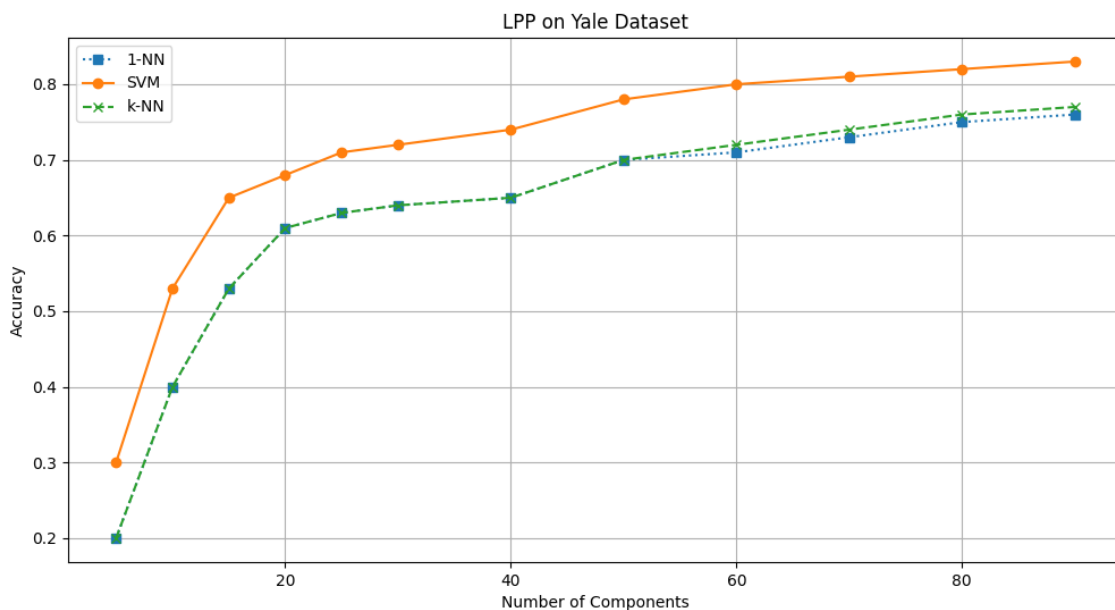
Εικόνα 27. Γράφημα απόδοσης – διάστασης, στο AR random χρησιμοποιώντας την μέθοδο PCA.

5.5.4 Προβολές διατήρησης τοπικότητας (LPP)

Κινούμενοι πέρα από τις παραδοσιακές γραμμικές μεθόδους, παρουσιάζουμε τις προβολές διατήρησης τοπικότητας (LPP). Εδραιωμένες στην πολλαπλή (manifold) μάθηση, οι LPP στοχεύουν στη διατήρηση της τοπικής δομής των δεδομένων κατά την προβολή τους σε έναν χώρο μειωμένης διαστατικότητας. Αυτό καθιστά τις LPP ιδιαίτερα χρήσιμες για σύνολα δεδομένων, όπου κυριαρχούν οι μη γραμμικές σχέσεις. Σε αυτήν την ενότητα, θα συζητήσουμε την απόδοση των LPP. Η αποτελεσματικότητα των LPP, τόσο στην ακρίβεια της ταξινόμησης, όσο και στην υπολογιστική αποδοτικότητα θα αξιολογηθεί, παρέχοντας μια ολοκληρωμένη κατανόηση των δυνατοτήτων τους.

Yale

Στην Εικόνα 28 παρατηρούμε την απόδοση των τριών ταξινομητών μας, με τη χρήση των LPP. Στην διάσταση 85 επιτυγχάνεται η μέγιστη απόδοση αναγνώρισης προσώπων με ποσοστά, 85% για SVM, 76% για 1-NN και τέλος, 77% για k-NN.

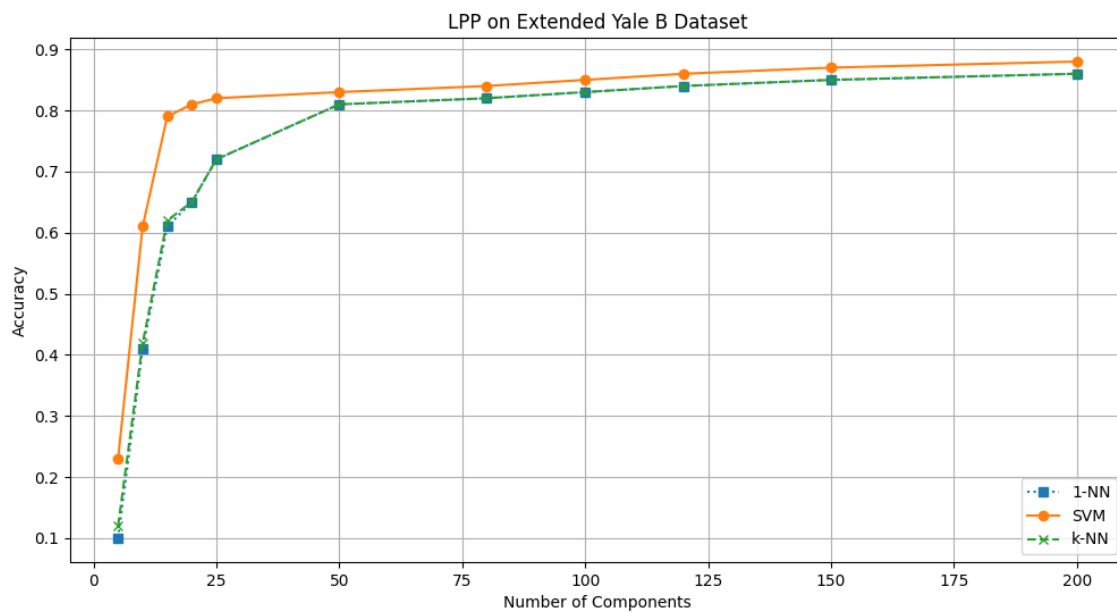


Εικόνα 28. Γράφημα απόδοσης – διάστασης, στο Yale χρησιμοποιώντας την μέθοδο LPP.

Extended Yale B

Αρκετά μεγαλύτερη είναι η απόδοση και των τριών ταξινομητών στον Extended Yale B, όπως μπορούμε να παρατηρήσουμε στην Εικόνα 29, όπου η απόδοση φτάνει το 88% για τον SVM και 86% 1-NN/k-NN για διάσταση μείωσης 200. Ακόμα, είναι ενδιαφέρον να

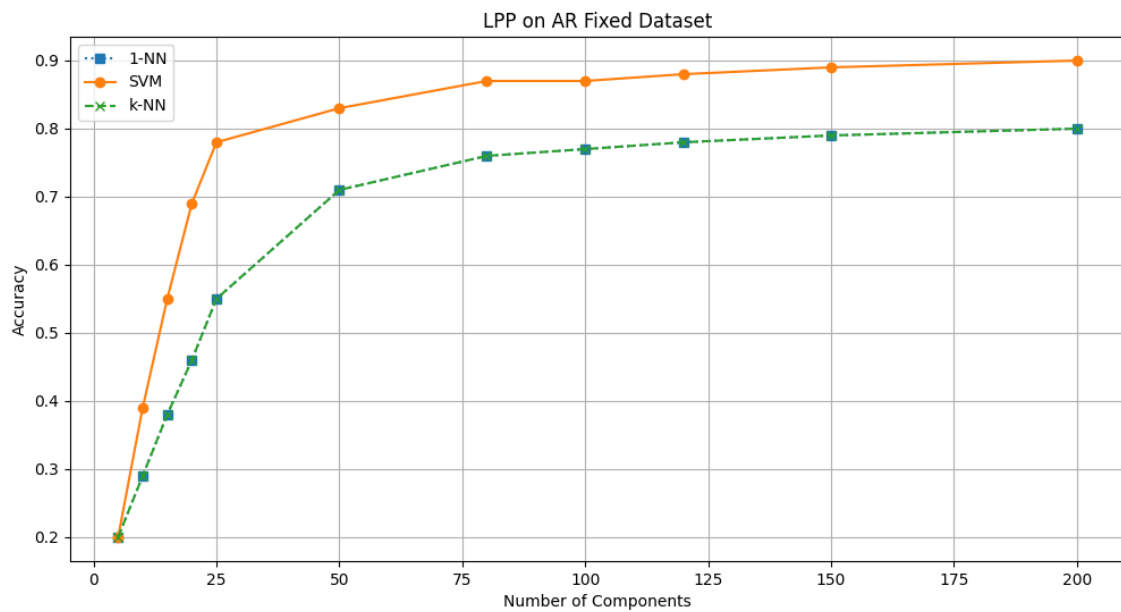
αναφέρουμε ότι και για μικρότερη διάσταση, 25, έχουμε εξίσου καλή απόδοση της τάξης του 81% και 71%, αντίστοιχα.



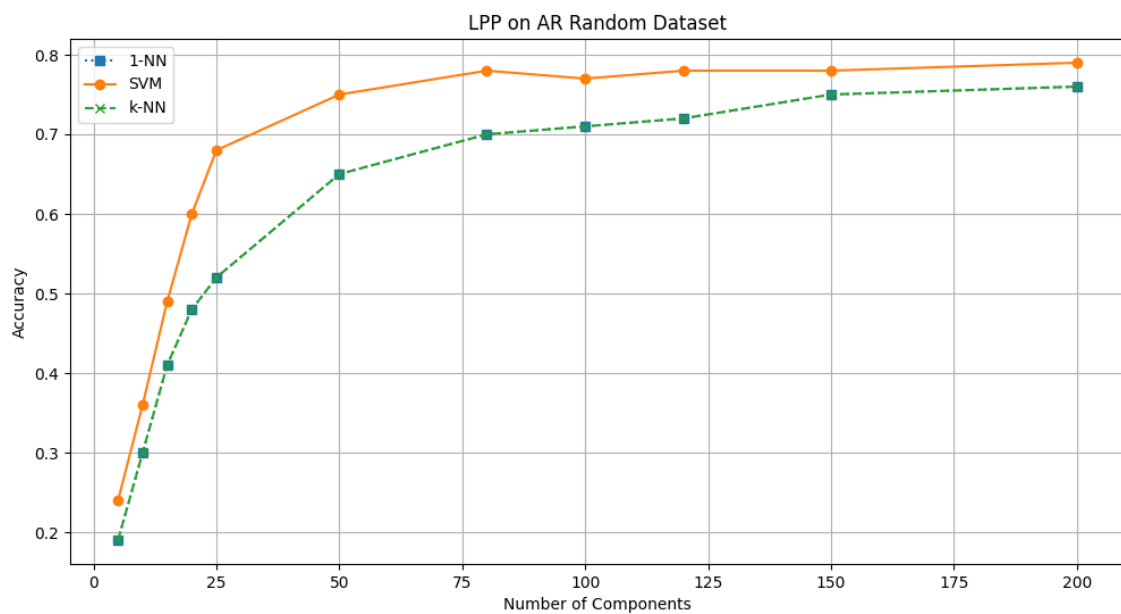
Εικόνα 29. Γράφημα απόδοσης – διάστασης, στο Extended Yale B χρησιμοποιώντας την μέθοδο LPP.

AR fixed - AR random

Εξίσου υψηλή απόδοση έχουμε και στον fixed διαχωρισμό της AR βάσης, χρησιμοποιώντας τις LPP. Όπως βλέπουμε στην παρακάτω Εικόνα 30, έχουμε επιτυχία αναγνώρισης 90% και 80% για τους ταξινομητές SVM και 1-NN/k-NN στη διάσταση 200. Από την άλλη πλευρά, στον διαχωρισμό random επιτυγχάνουμε ποσοστό 79% στον SVM και 76% στα 1-NN/k-NN στην διάσταση 200, Εικόνα 31.



Εικόνα 30. Γράφημα απόδοσης – διάστασης, στο AR fixed χρησιμοποιώντας την μέθοδο LPP.



Εικόνα 31. Γράφημα απόδοσης – διάστασης, στο AR random χρησιμοποιώντας την μέθοδο LPP.

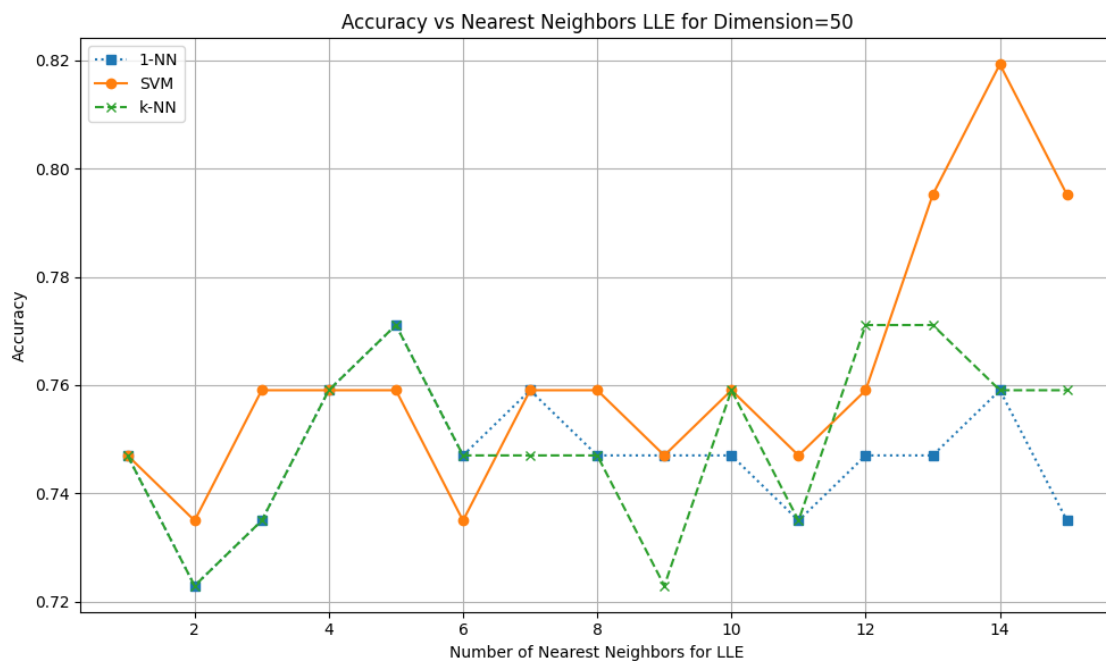
5.5.5 Τοπικά γραμμική ενσωμάτωση (LLE)

Καθώς συνεχίζουμε να εξερευνούμε τεχνικές πολλαπλής μάθησης, φτάνουμε στην τοπικά γραμμική ενσωμάτωση (LLE). Γνωστή για την ικανότητά της να ξεδιπλώνει δεδομένα υψηλής διαστατικότητας που βρίσκονται σε μια πολλαπλότητα χαμηλότερης

διαστατικότητα, η LLE προσφέρει μια μοναδική οπτική γωνία στη μείωση της διαστατικότητας. Αυτή η ενότητα θα επικεντρωθεί στην εφαρμογή της LLE στα σύνολα δεδομένων της μελέτης μας και θα παράσχει μια λεπτομερή ανάλυση του πώς αυτή επηρεάζει την ακρίβεια της ταξινόμησης.

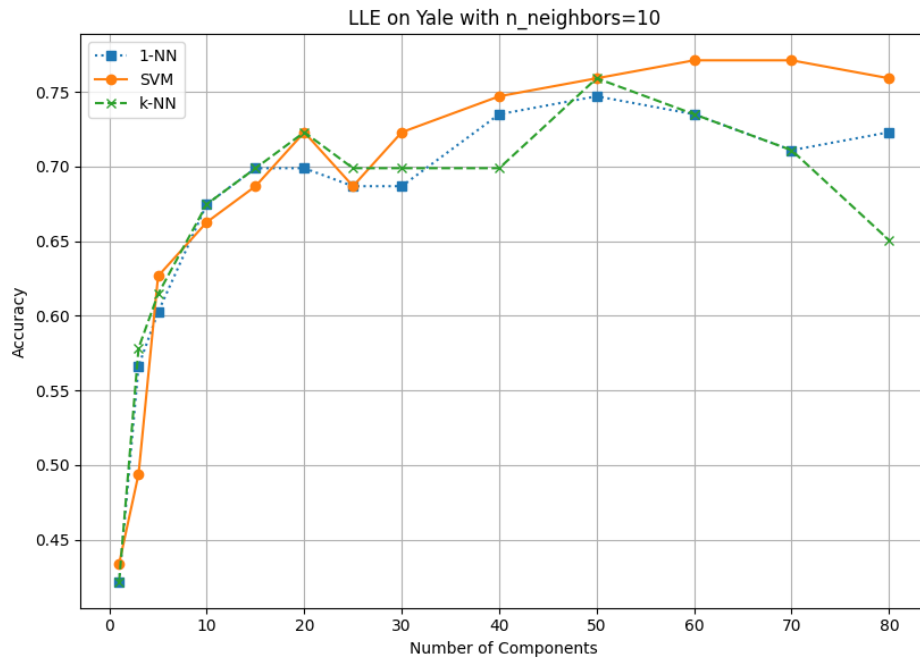
Yale

Το LLE έχει την ιδιοτροπία ότι απαιτεί δύο παραμέτρους για να λειτουργήσει σωστά, η μια είναι αυτή που παίρνουν όλες οι μέθοδοι που είναι η διάσταση μείωσης και η δεύτερη είναι η επιλογή των γειτόνων. Στην Εικόνα 32 παρουσιάζουμε την αλλαγή της απόδοσης σε διαφορετικές τιμές γειτόνων για μια σταθερή τιμή διάστασης 50. Όπως παρατηρούμε, η απόδοση στον SVM μεγιστοποιείται με ποσοστό 82% για 14 γείτονες, ενώ τα 1-NN/k-NN παρέχουν -σε γενικές γραμμές- σταθερές τιμές, κοντά στο 76%.



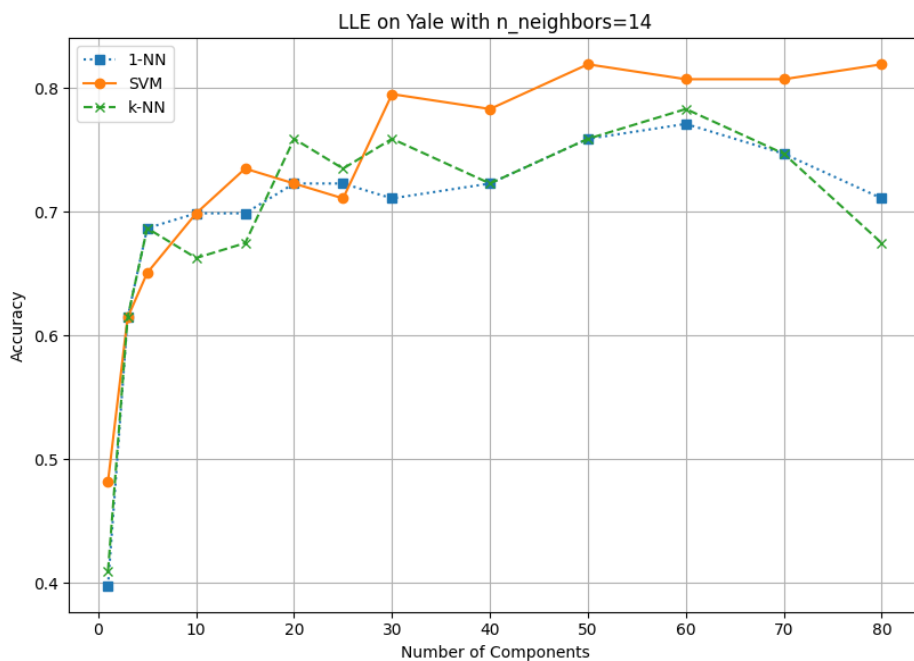
Εικόνα 32. Γράφημα απόδοσης – γείτονα, στο Yale χρησιμοποιώντας την μέθοδο LLE για διάσταση 50.

Παρακάτω παρουσιάζουμε τα αποτελέσματα σύγκρισης της απόδοσης της διάστασης μείωσης για δυο διαφορετικές τιμές γειτόνων, 10 και 14 γείτονες. Αρχικά, στην Εικόνα 33 παρατηρούμε την εξέλιξη της απόδοσης με την παράμετρο των 10 γειτόνων. Επιτυγχάνεται αναγνώριση 78% για τον SVM, σε διάσταση 60, 74% για τον 1-NN και 76% για τον k-NN, στην διάσταση 50.



Εικόνα 33. Γράφημα απόδοσης – διάστασης, στο Yale χρησιμοποιώντας την μέθοδο LLE για 10 γείτονες.

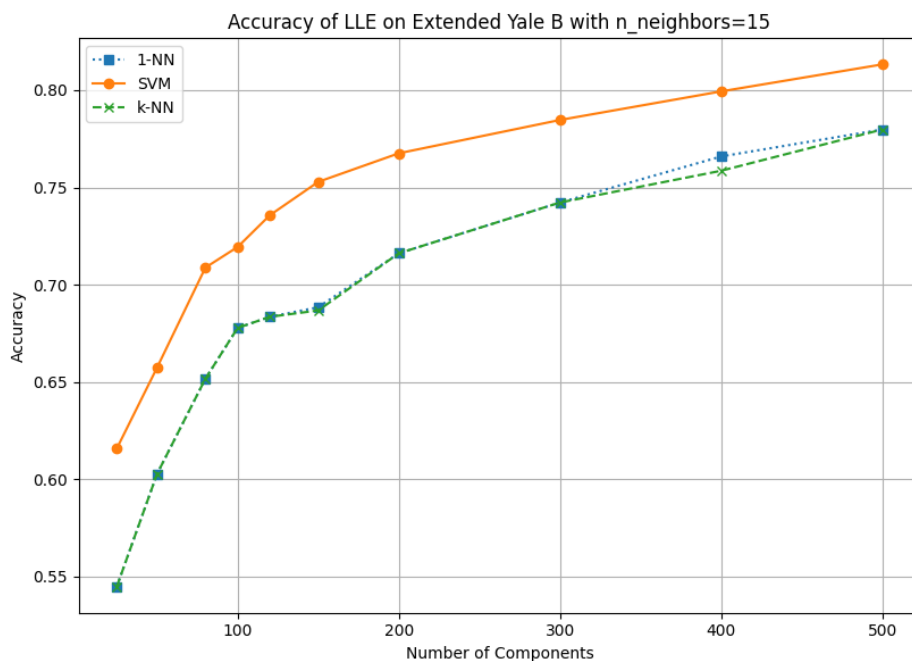
Έπειτα εξετάζουμε την απόδοση για 14 γείτονες, στην Εικόνα 34, όπου, όπως παρατηρούμε, η απόδοση ανεβαίνει αισθητά για τον SVM με ποσοστό 82% στη διάσταση 50, ενώ τα 1-NN και k-NN έχουν μια πιο μικρή αύξηση σε 77% και 78%, αντίστοιχα, για τη διάσταση 60.



Εικόνα 34. Γράφημα απόδοσης – διάστασης, στο Yale χρησιμοποιώντας την μέθοδο LLE για 14 γείτονες.

Extended Yale B

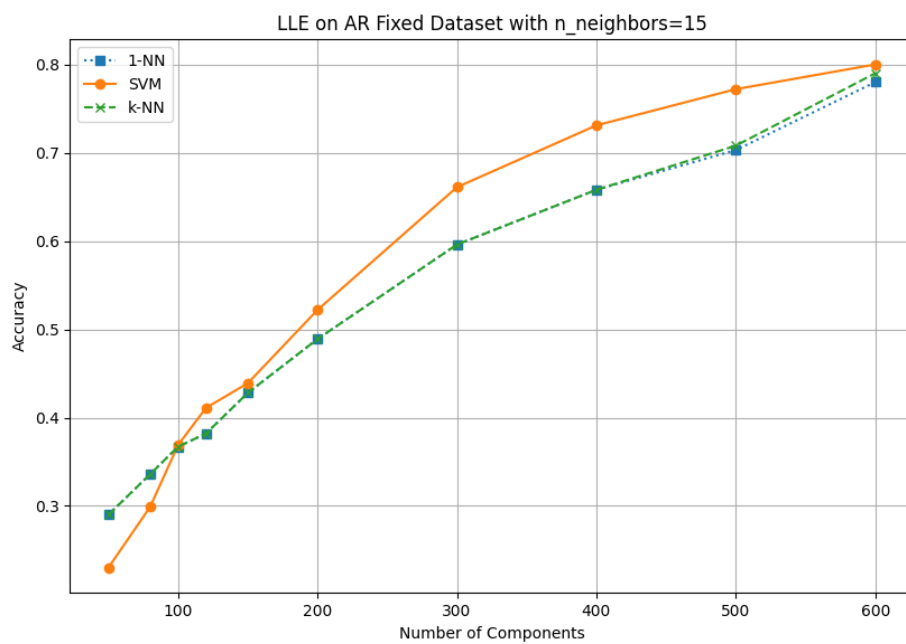
Για τον Extended Yale B, πραγματοποιήσαμε διασταυρωμένη επικύρωση για την εύρεση της καλύτερης τιμής γειτόνων και παρατηρήσαμε ότι η απόδοση μεγιστοποιείται για 15 γείτονες. Το SVM φτάνει απόδοση 82% για διάσταση 500, ενώ τα 1-NN/k-NN 78%, αντίστοιχα, Εικόνα 35.



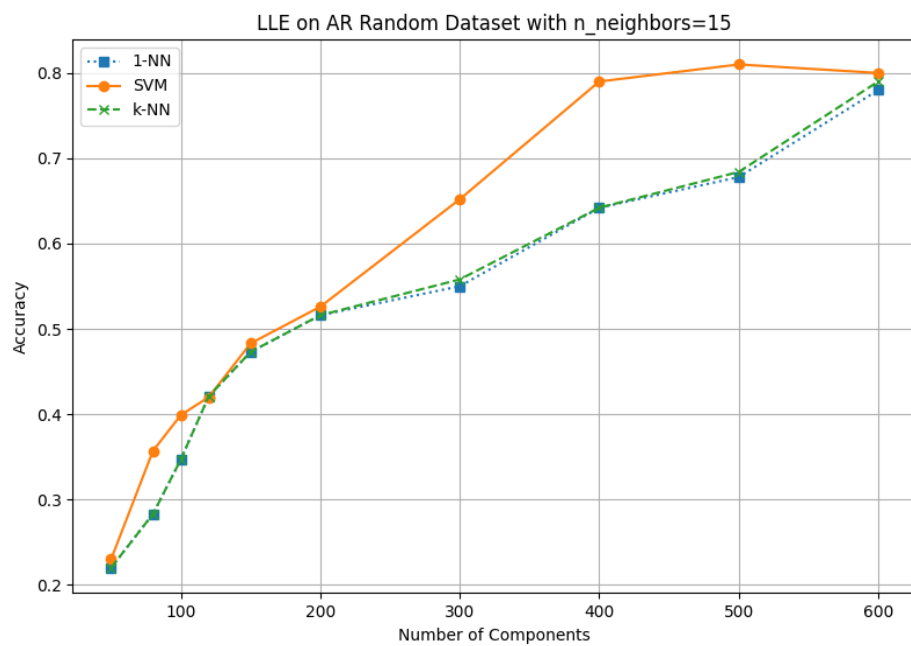
Εικόνα 35. Γράφημα απόδοσης – διάστασης, στο Extended Yale B χρησιμοποιώντας την μέθοδο LLE για 15 γείτονες.

AR fixed - AR random

Μεγάλα ποσοστά απόδοσης παρατηρούμε και στους δύο διαχωρισμούς της βάσης δεδομένων AR. Με τον fixed διαχωρισμό να επιτυγχάνει ποσοστά 80%, 78% και 79%, για τους τρεις ταξινομητές SVM, 1-NN και k-NN στην διάσταση 600, με 15 γείτονες, Εικόνα 36. Παρόμοια αποτελέσματα είχε και ο random διαχωρισμός, με ποσοστά 81% για διάσταση 500 στον SVM, 78% και 79% για διάσταση 600 στα 1-NN/k-NN, με επιλογή 15 γειτόνων, Εικόνα 37.



Εικόνα 36. Γράφημα απόδοσης – διάστασης, στο AR fixed χρησιμοποιώντας την μέθοδο LLE για 15 γείτονες.



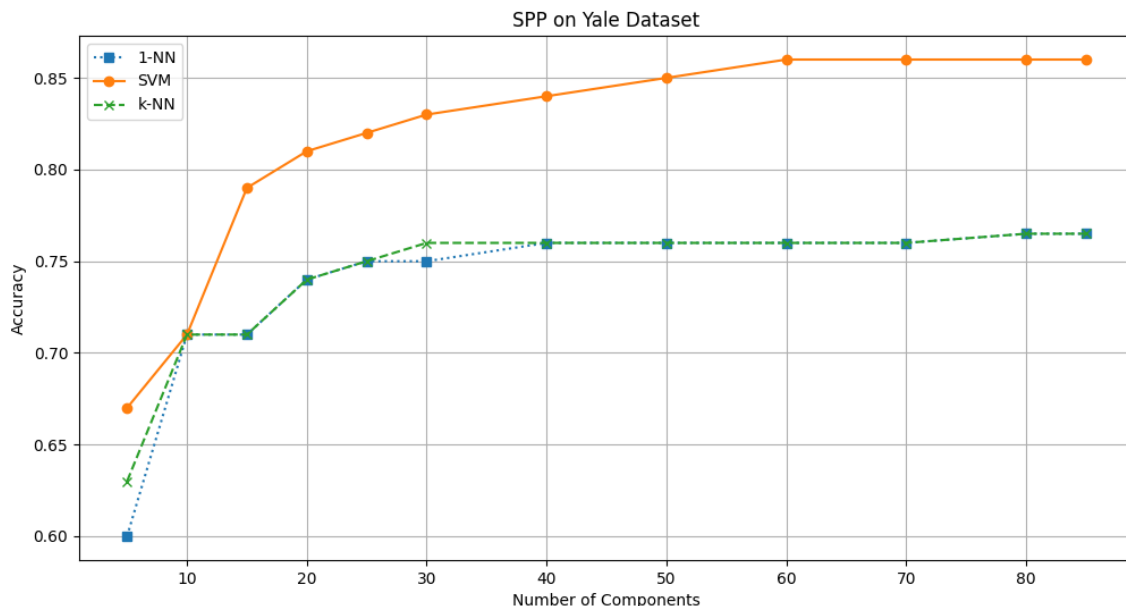
Εικόνα 37. Γράφημα απόδοσης – διάστασης, στο AR random χρησιμοποιώντας την μέθοδο LLE για 15 γείτονες.

5.5.6 Προβολές διατήρησης αραιότητας (SPP)

Μετά την εξερεύνηση μιας σειράς καθιερωμένων μεθόδων, προχωρούμε τώρα στην πρώτη από τις κύριες τεχνικές αυτής της εργασίας, τις προβολές διατήρησης αραιότητας (SPP). Σχεδιασμένο για να διατηρεί την αραιή αναπαράσταση των δεδομένων κατά τη μείωση της διαστατικότητάς τους, το SPP προσφέρει μια λεπτομερή προσέγγιση στην επιλογή χαρακτηριστικών. Αυτή η ενότητα παρέχει μια λεπτομερή ανάλυση της απόδοσης του SPP στα σύνολα δεδομένων Yale, Extended Yale B και AR.

Yale

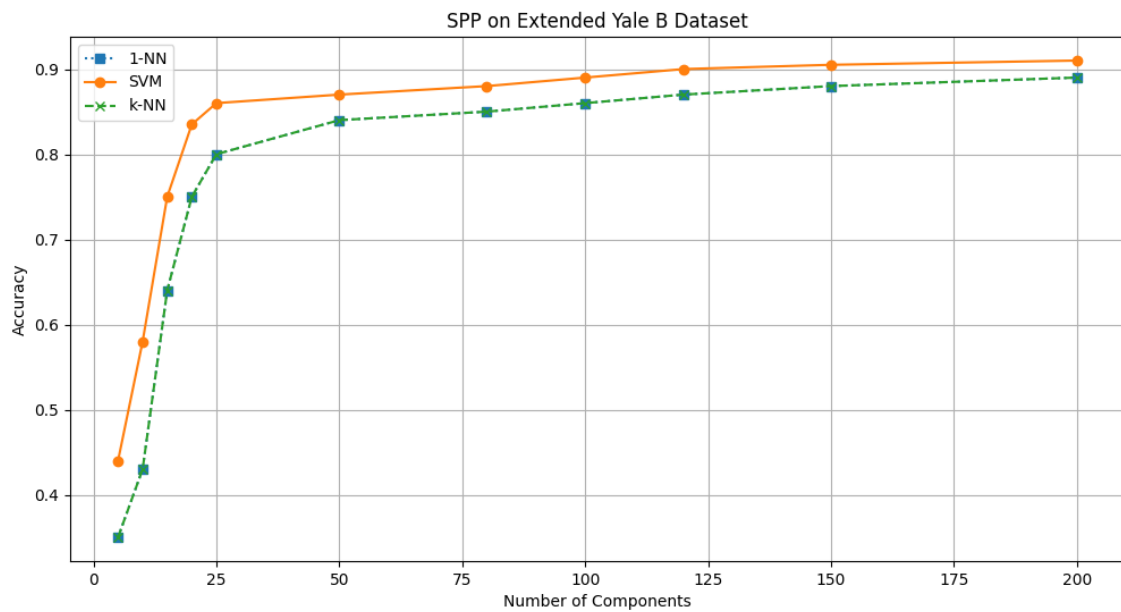
Ξεκινώντας με την πρώτη βάση δεδομένων, Εικόνα 38, παρατηρούμε μια σημαντική άνοδο στην απόδοση του ταξινομητή SVM, με ποσοστό επιτυχία 86% για την διάσταση 60. Τα 1-NN και k-NN προσφέρουν απόδοση της τάξης του 77% για διάσταση 80.



Εικόνα 38. Γράφημα απόδοσης – διάστασης, στο Yale χρησιμοποιώντας την μέθοδο SPP.

Extended Yale B

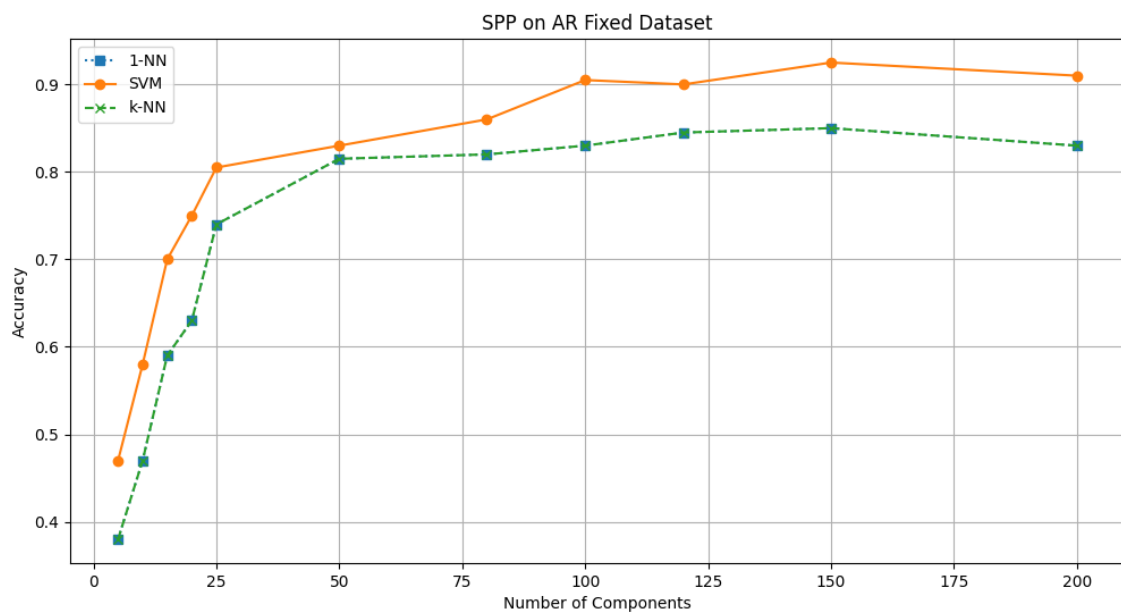
Συνεχίζοντας στον Extended Yale B, Εικόνα 39, παρατηρούμε ότι η ανοδική πορεία της απόδοσης συνεχίζεται. Με τον SVM να πιάνει ποσοστό 91% και τα 1-NN/k-NN ποσοστό 89%, για την διάσταση 200.



Εικόνα 39. Γράφημα απόδοσης – διάστασης, στο Extended Yale B χρησιμοποιώντας την μέθοδο SPP.

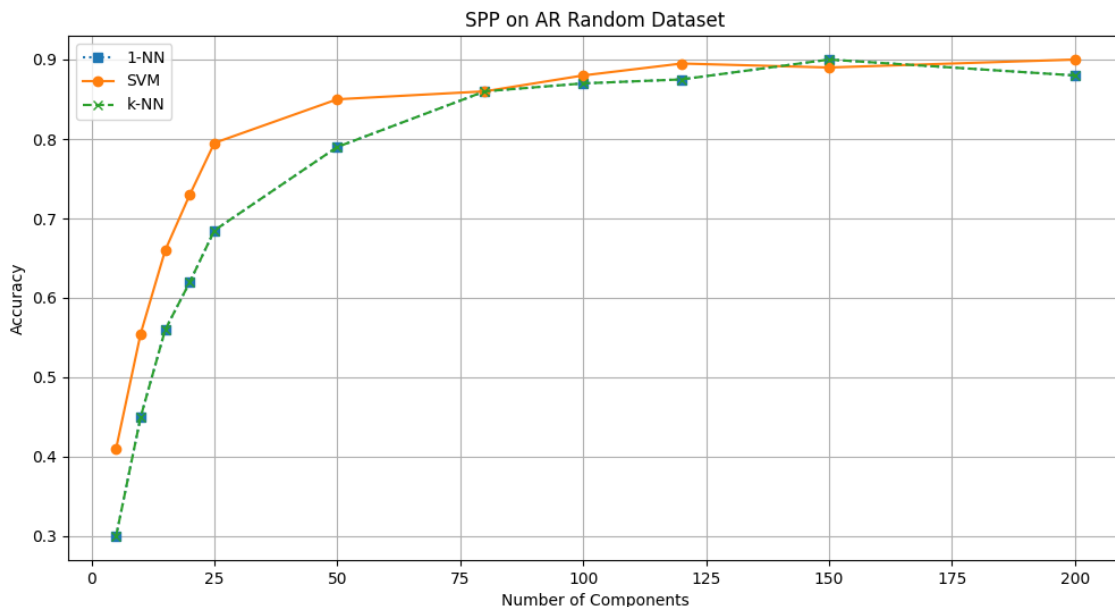
AR fixed - AR random

Στον AR fixed διαχωρισμό, της Εικόνας 40, με την χρήση του SPP επιτυγχάνουμε ποσοστό αναγνώρισης προσώπου 93% για τον SVM ταξινομητή, για την διάσταση 150 και 85% στην ίδια διάσταση για τους ταξινομητές 1-NN και k-NN.



Εικόνα 40. Γράφημα απόδοσης – διάστασης, στο AR fixed χρησιμοποιώντας την μέθοδο SPP.

Εξίσου μεγάλο ποσοστό αναγνώρισης συναντάμε και στον διαχωρισμό random, Εικόνα 41, με το ποσοστό να αγγίζει το 90% και στους τρεις ταξινομητές, για διαστάσεις 200 στον SVM και 150 στα 1-NN/k-NN, αντίστοιχα.



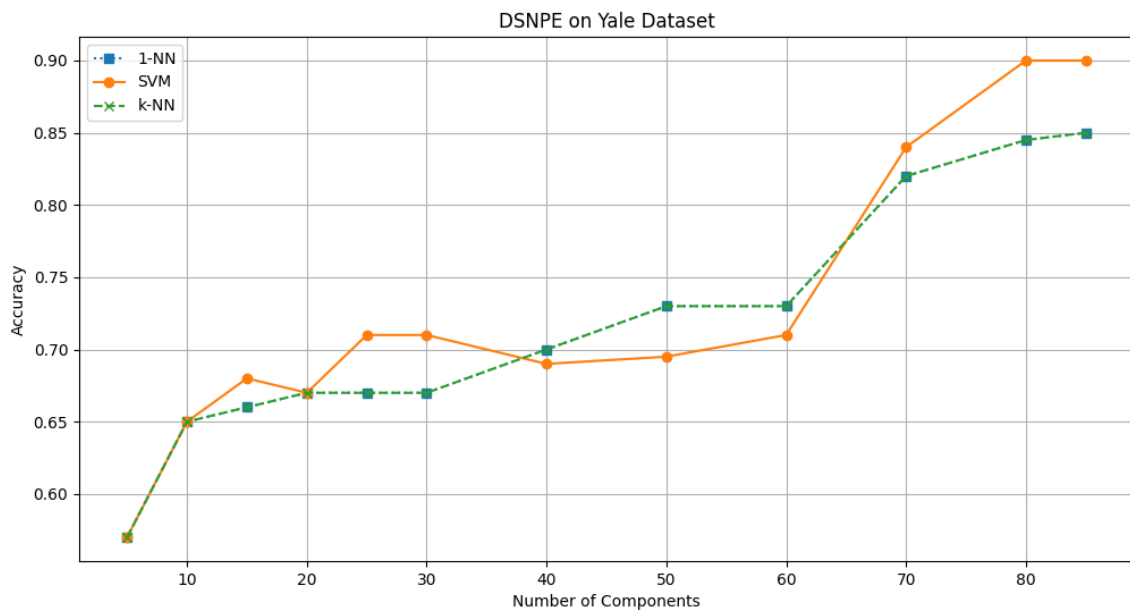
Εικόνα 41. Γράφημα απόδοσης – διάστασης, στο AR random χρησιμοποιώντας την μέθοδο SPP.

5.5.7 Διακριτική αραιή ενσωμάτωση που διατηρεί τη γειτονιά (DSNPE)

Τέλος, φτάνουμε στη δεύτερη κύρια τεχνική που ερευνάται σε αυτή τη διπλωματική εργασία, την διακριτική αραιή ενσωμάτωση που διατηρεί τη γειτονιά (DSNPE). Αυτή η μέθοδος συνδυάζει τα πλεονεκτήματα των προβολών διατήρησης αραιότητας, προσφέροντας μια εξελιγμένη εναλλακτική λύση για τη μείωση της διαστατικότητας. Σε αυτήν την ενότητα, θα εμβαθύνουμε στις λεπτομέρειες της απόδοσης του DSNPE στα επιλεγμένα σύνολα δεδομένων μας.

Yale

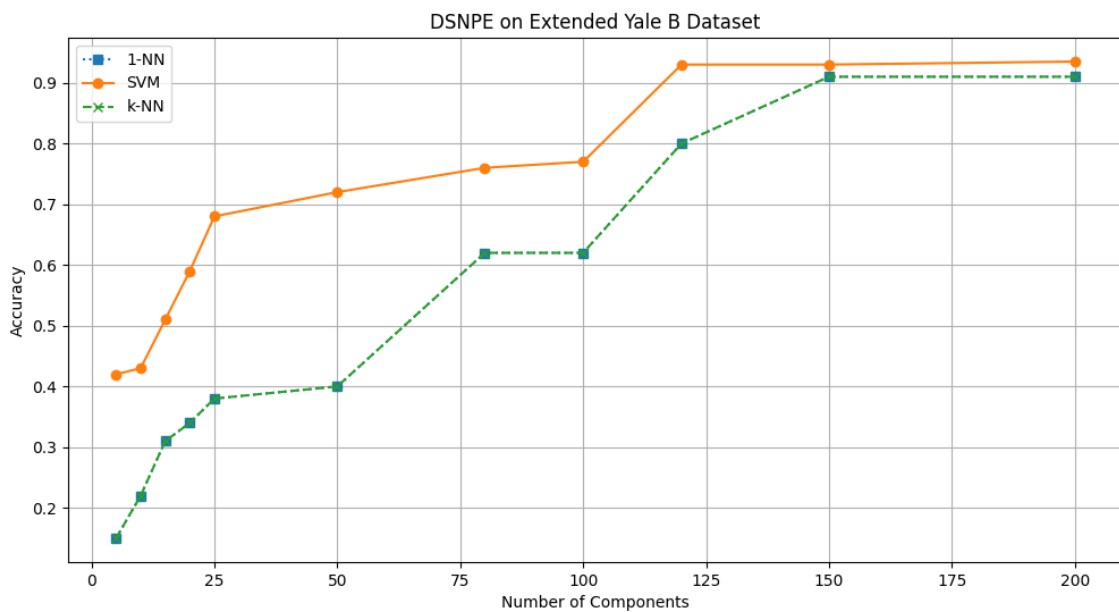
Η αναγνώριση στο Yale με την χρήση του DSNPE έχει ποσοστά αναγνώρισης 90% για τον SVM στην διάσταση 80. Αντίστοιχα, πολύ καλό ποσοστό έχουμε και με τους 1-NN και k-NN, όπου επιτυγχάνουμε 85% αναγνώρισης για διάσταση μείωσης 85. Στην παρακάτω Εικόνα 42 παραθέτουμε το γράφημα με τα αποτελέσματά μας.



Εικόνα 42. Γράφημα απόδοσης – διάστασης, στο Yale χρησιμοποιώντας την μέθοδο DSNPE.

Extended Yale B

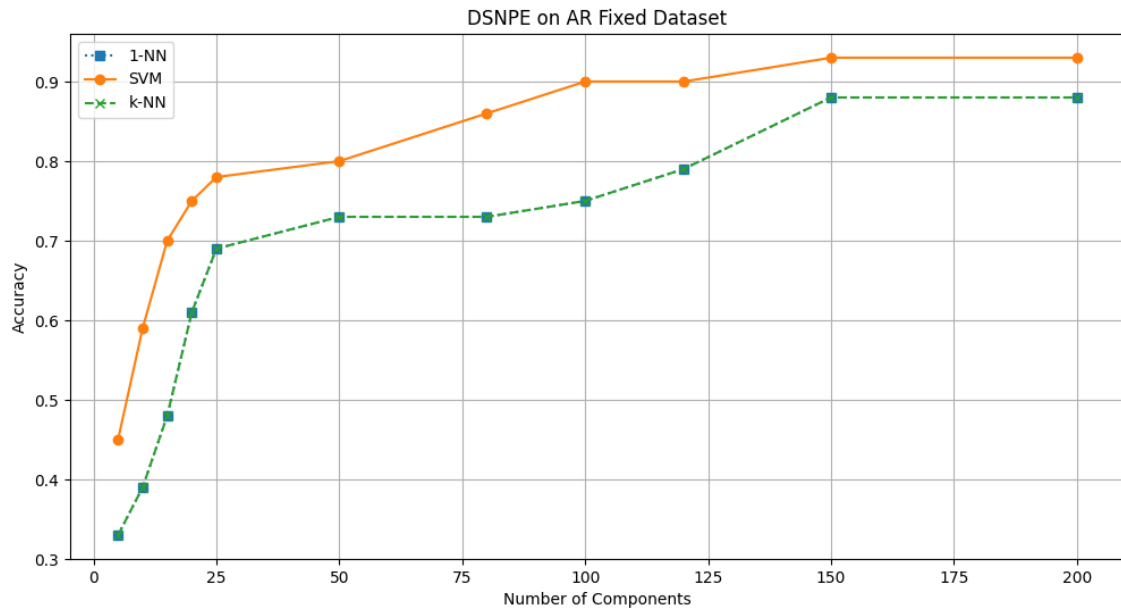
Για τον Extended Yale B τα αποτελέσματα μάς εντυπωσίασαν καθώς ο DSNPE κατάφερε να αγγίξει ποσοστά μεγαλύτερα του 90% και για τους τρεις ταξινομητές, Εικόνα 43. Συγκεκριμένα, στην διάσταση 120 ο SVM επιτυγχάνει 94%, ενώ τα 1-NN και k-NN στην διάσταση 150 αγγίζουν 91% αναγνώρισης.



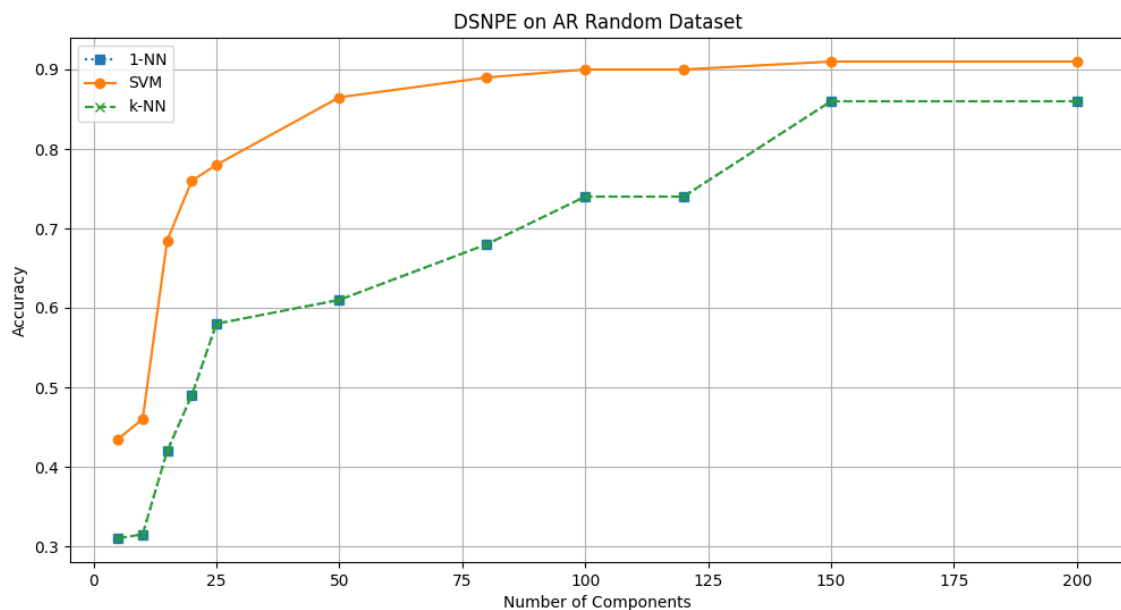
Εικόνα 43. Γράφημα απόδοσης – διάστασης, στο Extended Yale B χρησιμοποιώντας την μέθοδο DSNPE.

AR fixed - AR random

Για τους AR fixed (Εικόνα 44) και AR random (Εικόνα 45) έχουμε πάλι υψηλό ποσοστό αναγνώρισης. Ο SVM μας δίνει ποσοστά 93% και 91% στην διάσταση 150, σε fixed και random. Από την άλλη πλευρά, για 1-NN και k-NN έχουμε ποσοστά 88% και 86% στην ίδια διάσταση 150, για τους δύο διαχωρισμούς μας fixed και random, αντίστοιχα.



Εικόνα 44. Γράφημα απόδοσης – διάστασης, στο AR fixed χρησιμοποιώντας την μέθοδο DSNPE.

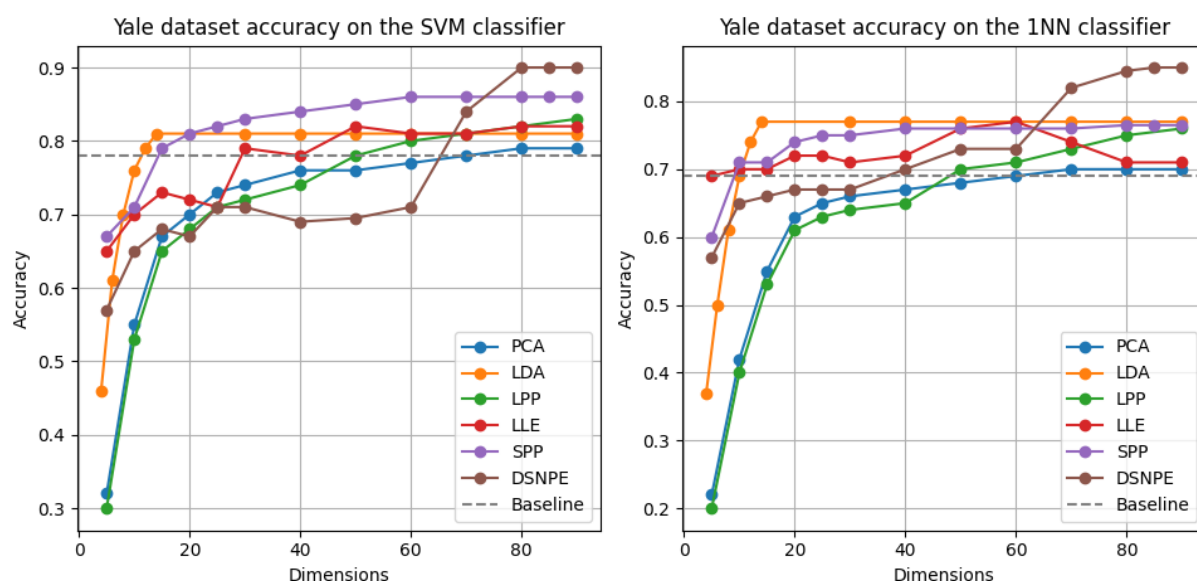


Εικόνα 45. Γράφημα απόδοσης – διάστασης, στο AR random χρησιμοποιώντας την μέθοδο DSNPE.

5.5.8 Σύγκριση των μεθόδων

Έχοντας αναλύσει λεπτομερώς τις ατομικές επιδόσεις των μεθόδων βάσης, PCA, LDA, LPP, LLE, SPP και DSNPE, φτάνουμε τώρα στο τελικό και ίσως πιο διαφωτιστικό στάδιο της πειραματικής μας ανάλυσης, τη σύγκριση. Αυτή η ενότητα συνθέτει τα αποτελέσματα των προηγούμενων υποενοτήτων και παρέχει μια συνολική άποψη για το πώς αποδίδει κάθε τεχνική μείωσης διαστατικότητας, όσον αφορά την ακρίβεια ταξινόμησης, την υπολογιστική αποδοτικότητα και την ανθεκτικότητα. Μέσω μιας ποικιλίας συγκριτικών μετρικών και οπτικών βοηθημάτων, θα αποσαφηνίσουμε την πολύπλοκη αλληλεπίδραση των παραγόντων που συμβάλλουν στην αποτελεσματικότητα κάθε μεθόδου. Αυτή η περιεκτική ανάλυση στοχεύει στην παροχή μιας πολυδιάστατης κατανόησης των μεθόδων που είναι πιο κατάλληλες για συγκεκριμένους τύπους συνόλων δεδομένων και υπό ποιες συνθήκες.

Yale



Εικόνα 46. Γραφήματα απόδοσης – διάστασης, σύγκρισης όλων των μεθόδων στη βάση δεδομένων Yale για τους ταξινομητές SVM και 1-NN.

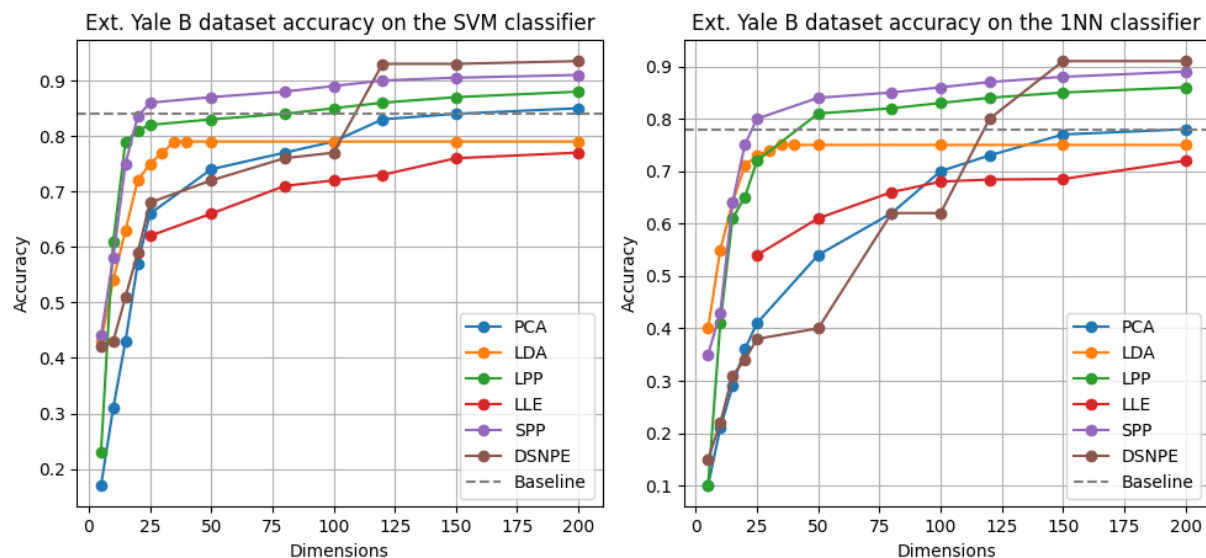
Ξεκινάμε τη σύγκριση των μεθόδων με την βάση δεδομένων Yale. Ο Πίνακας 1 δείχνει τα μέγιστα μέσα ποσοστά αναγνώρισης σε 20 εκτελέσεις κάθε μεθόδου, υπό τους ταξινομητές SVM, 1-NN και k-NN και τις αντίστοιχες διαστάσεις, όπου τα καλύτερα αποτελέσματα είναι τονισμένα με έντονα γράμματα. Οι καμπύλες ποσοστού αναγνώρισης των διαφορετικών αλγορίθμων φαίνονται στην Εικόνα 46, για τους ταξινομητές SVM και 1-NN. Λόγω του περιορισμένου χώρου, παραλείψαμε τον ταξινομητή k-NN, ο οποίος, όπως θα δούμε και στον Πίνακα 1, έχει σχεδόν τα ίδια αποτελέσματα με τον ταξινομητή 1-NN. Για τη μέθοδο βάσης, απλώς πραγματοποιήσαμε αναγνώριση προσώπου στον αρχικό χώρο εικόνας 1024 διαστάσεων. Σημειώστε ότι το ανώτατο όριο της διαστατικότητας της LDA είναι $c-1$, όπου c είναι ο αριθμός των ατόμων [2]. Όπως μπορεί να διαπιστωθεί, οι αλγόριθμοί μας SPP και

DSNPE υπερτερούν όλων των άλλων μεθόδων, εκτός του ταξινομητή k-NN, όπου ο LLE υπερτερεί του SPP κατά 1%, ενώ η μέθοδος PCA ήταν η χειρότερη σε όλες τις περιπτώσεις. Είναι πολύ ενδιαφέρον ότι η μέθοδος PCA και η βασική μέθοδος έχουν σχεδόν την ίδια απόδοση, η οποία συμφωνεί με τα αποτελέσματα σε πολλές δημοσιεύσεις, όπως το [40].

Πίνακας 1. Μέγιστα ποσοστά αναγνώρισης στο Yale για κάθε ταξινομητή και οι αντίστοιχες διαστάσεις.

Ταξινομητής	Μέθοδος	Baseline	PCA	LDA	LPP	LLE	SPP	DSNPE
SVM	Ποσοστό % αναγνώρισης	78	79	81	83	82	86	90
	Διάσταση	1024	80	14	89	50	60	80
1-NN	Ποσοστό % αναγνώρισης	69	70	77	76	77	77	85
	Διάσταση	1024	76	14	87	60	80	82
k-NN	Ποσοστό % αναγνώρισης	69	70	77	77	78	77	85
	Διάσταση	1024	76	14	87	60	80	82

Extended Yale B



Εικόνα 47. Γραφήματα απόδοσης – διάστασης, σύγκρισης όλων των μεθόδων στη βάση δεδομένων Extended Yale B για τους ταξινομητές SVM και 1-NN.

Ο πειραματικός σχεδιασμός είναι ίδιος με αυτόν του Yale. Το μέγιστο μέσο ποσοστό αναγνώρισης και η αντίστοιχη διαστατικότητα σε 20 εκτελέσεις των δοκιμών κάθε μεθόδου φαίνονται στον Πίνακα 2. Τα καλύτερα αποτελέσματα είναι τονισμένα με έντονα γράμματα. Επιπλέον, σχεδιάζουμε τις καμπύλες ποσοστού αναγνώρισης των αλγορίθμων στην Εικόνα 47. Όπως μπορεί να διαπιστωθεί, οι αλγόριθμοί μας SPP και DSNPE έδωσαν τα

καλύτερα αποτελέσματα, σε όλες τις περιπτώσεις. Επιπλέον, η μέθοδος LDA έχει τη χειρότερη απόδοση σε όλες τις περιπτώσεις, που την τοποθετεί κάτω από τη μέθοδο της βάσης, σε θέμα αναγνώρισης προσώπων, στην συγκεκριμένη βάση δεδομένων.

Πίνακας 2. Μέγιστα ποσοστά αναγνώρισης στο Extended Yale B για κάθε ταξινομητή και οι αντίστοιχες διαστάσεις.

Ταξινομητής	Μέθοδος	Baseline	PCA	LDA	LPP	LLE	SPP	DSNPE
SVM	Ποσοστό % αναγνώρισης	84	85	79	88	82	91	94
	Διάσταση	1024	200	38	175	500	200	120
1-NN	Ποσοστό % αναγνώρισης	78	79	75	86	78	89	91
	Διάσταση	1024	200	36	175	500	200	150
k-NN	Ποσοστό % αναγνώρισης	78	79	76	86	78	89	91
	Διάσταση	1024	200	35	175	500	200	150

AR fixed - AR random

Εδώ πάλι ακολουθούμε την ίδια πορεία, όπως και με τις άλλες βάσεις δεδομένων. Το μέγιστο ποσοστό αναγνώρισης κάθε μεθόδου και η αντίστοιχη μειωμένη διάσταση παρατίθενται στους Πίνακες 3 και 4, για διαχωρισμό δεδομένων fixed και random, αντίστοιχα. Οι καμπύλες ποσοστού αναγνώρισης αλγορίθμων, σε σχέση με τις διάφορες διαστάσεις, για τους δυο ταξινομητές SVM και 1-NN φαίνονται στην Εικόνα 48 για fixed διαχωρισμό και στην Εικόνα 49 για random διαχωρισμό. Από τα πειραματικά αποτελέσματα, μπορούμε να δούμε ότι τα SPP και DSNPE επιτυγχάνουν τα υψηλότερα ποσοστά αναγνώρισης. Είναι πολύ ενδιαφέρον ότι είναι η μοναδική φορά που το SPP, για διαχωρισμό random των δεδομένων, υπερτερεί και του DSNPE, καθώς σε όλες τις άλλες βάσεις, ο DSNPE έρχεται πάντα πρώτος, με τον SPP να ακολουθεί με πολύ μικρή διαφορά.

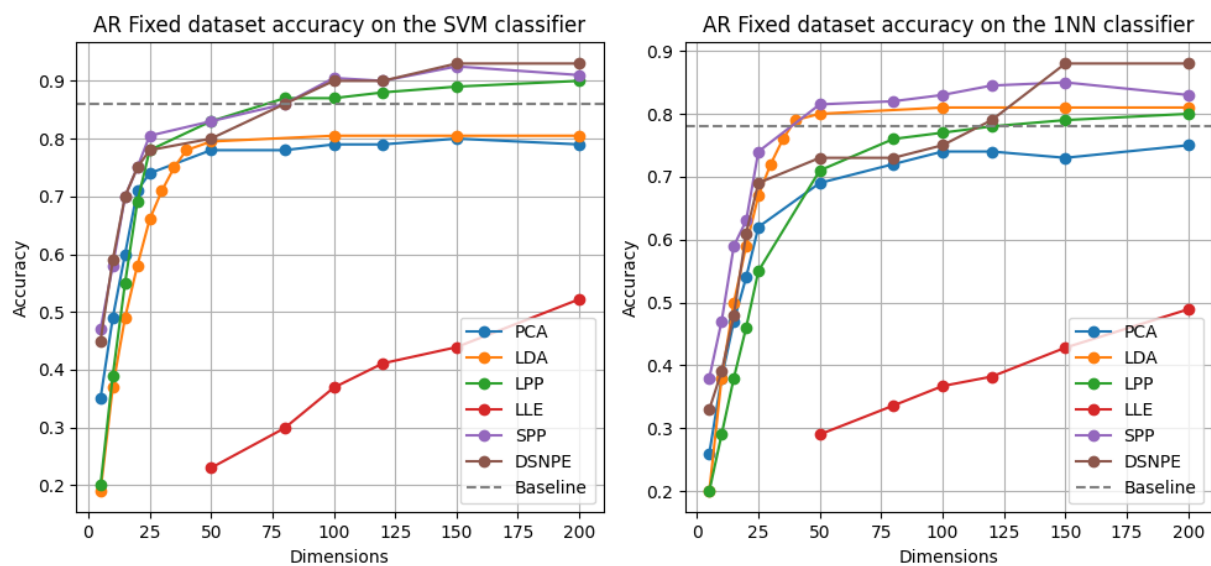
Πίνακας 3. Μέγιστα ποσοστά αναγνώρισης στο AR fixed για κάθε ταξινομητή και οι αντίστοιχες διαστάσεις.

Ταξινομητής	Μέθοδος	Baseline	PCA	LDA	LPP	LLE	SPP	DSNPE
SVM	Ποσοστό % αναγνώρισης	86	80	80	90	80	93	93
	Διάσταση	3168	150	100	200	600	150	150
1-NN	Ποσοστό % αναγνώρισης	78	75	81	80	78	85	88
	Διάσταση	3168	200	100	200	600	150	150

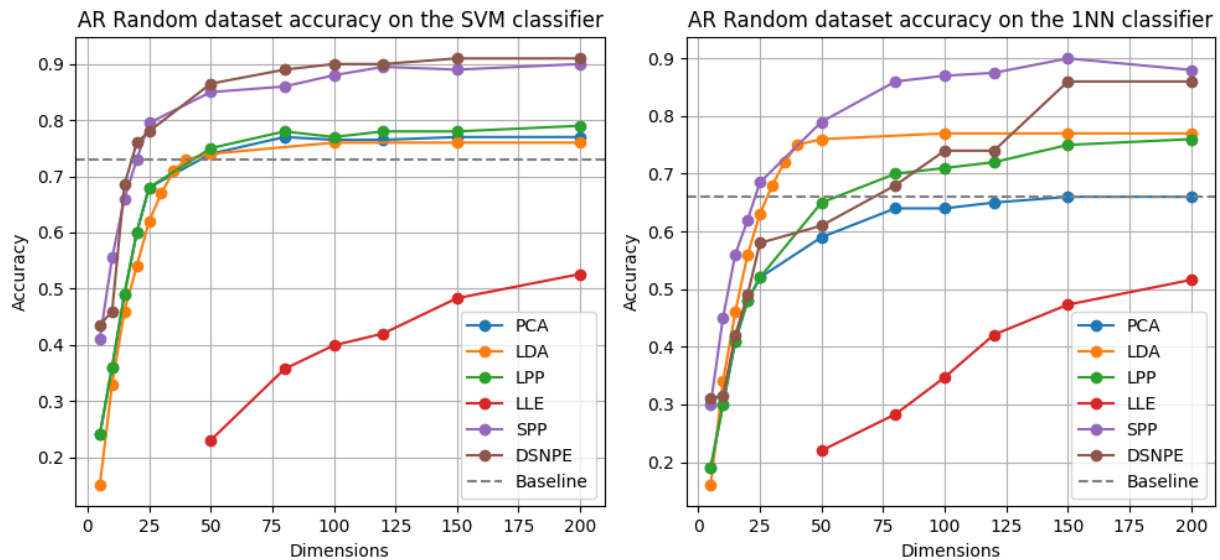
k-NN	Ποσοστό % αναγνώρισης	78	75	81	80	79	85	88
	Διάσταση	3168	200	100	200	600	150	150

Πίνακας 4. Μέγιστα ποσοστά αναγνώρισης στο AR random για κάθε ταξινομητή και οι αντίστοιχες διαστάσεις.

Ταξινομητής	Μέθοδος	Baseline	PCA	LDA	LPP	LLE	SPP	DSNPE
SVM	Ποσοστό % αναγνώρισης	73	77	76	79	81	90	91
	Διάσταση	3168	79	100	200	500	200	175
1-NN	Ποσοστό % αναγνώρισης	66	66	77	76	78	90	86
	Διάσταση	3168	150	100	200	600	150	175
k-NN	Ποσοστό % αναγνώρισης	66	66	77	76	79	90	86
	Διάσταση	3168	150	100	200	600	150	175



Εικόνα 48. Γραφήματα απόδοσης – διάστασης, σύγκρισης όλων των μεθόδων στη βάση δεδομένων AR fixed για τους ταξινομητές SVM και 1-NN.



Εικόνα 49. Γραφήματα απόδοσης – διάστασης, σύγκρισης όλων των μεθόδων στη βάση δεδομένων AR random για τους ταξινομητές SVM και 1-NN.

5.5.9 Συνολικές παρατηρήσεις και συζήτηση των παραπάνω πειραματικών αποτελεσμάτων

Έχοντας εξετάσει συστηματικά την απόδοση διάφορων τεχνικών μείωσης διαστατικότητας, σε πολλαπλά σύνολα δεδομένων και ταξινομητές, φτάνουμε τώρα σε μια στιγμή σύνθεσης. Αυτή η τελική ενότητα θα συγκεντρώσει τις παρατηρήσεις μας, προσφέροντας μια συνεκτική ερμηνεία των αποτελεσμάτων. Συγκρίνοντας τις διάφορες μεθόδους κατά κεφαλήν, στοχεύουμε να παράσχουμε μια λεπτομερή κατανόηση των σχετικών πλεονεκτημάτων και περιορισμών τους, ενημερώνοντας έτσι τη μελλοντική έρευνα και τις πρακτικές εφαρμογές στον τομέα της μηχανικής μάθησης και της επιστήμης των δεδομένων.

1. Το PCA είναι απλό στην εκτέλεση, αλλά γενικά έχει πολύ χειρότερη απόδοση συγκριτικά με τις υπόλοιπες μεθόδους μας. Και στους τρεις ταξινομητές, τα ποσοστά αναγνώρισής του είναι απλά κοντά στη βασική γραμμή, αν όχι και χειρότερα, σε κάποιες περιπτώσεις, σε όλες τις χρησιμοποιημένες βάσεις δεδομένων. Αυτό συμφωνεί με τα αποτελέσματα σε πολλές δημοσιεύσεις, όπως στο [40].
2. Τα LPP και LLE υπερτερούν πάντα του PCA, όταν η διάσταση υποχώρου υπερβαίνει ένα συγκεκριμένο όριο. Αυτό δείχνει ότι, διατηρώντας την τοπική δομή των δεδομένων, το ποσοστό αναγνώρισης μπορεί να βελτιωθεί. Δηλαδή, όταν λαμβάνεται υπόψη η αναζήτηση του κοντινότερου γείτονα, η τοπική δομή φαίνεται να είναι πιο σημαντική από την παγκόσμια δομή. Ωστόσο, τα LPP και LLE είναι λιγότερο διαχειρίσιμα από το PCA, λόγω της δυσκολίας επιλογής παραμέτρων που εμπλέκονται.
3. Στις δοκιμασμένες βάσεις δεδομένων, ο SPP και ο DSNPE υπερτερούν των PCA, LDA, LPP και LLE με τους ταξινομητές SVM, 1-NN και k-NN, ακόμη και αν δεν χρειάζεται να

προσαρμοστεί καμία παράμετρος. Αυτό υποδηλώνει ότι, οι προβολές που βρίσκονται από το SPP και το DSNPE, μπορούν να διατηρήσουν περισσότερες πληροφορίες διάκρισης από αυτές των συγκρινόμενων μεθόδων.

4. Όπως παρατηρούμε στους πίνακες συγκρίσεων, οι ταξινομητές επηρεάζουν, επίσης, σημαντικά την απόδοση αναγνώρισης. Ωστόσο, τα προτεινόμενα SPP και DSNPE μπορούν γενικά να επιτύχουν καλύτερη απόδοση από το PCA, LDA, LPP και LLE, βασισμένα στους περισσότερους ταξινομητές (δηλαδή, 1NN, k-NN και SVM). Η μόνη εξαίρεση είναι ότι για τον ταξινομητή k-NN, και για την βάση δεδομένων Yale το LLE υπερτερεί του SPP κατά 1%. Επιπλέον, συνδυάζοντας τον αλγόριθμο DSNPE και τον ταξινομητή SVM, μπορούμε να επιτύχουμε την καλύτερη απόδοση ανάμεσα σε όλους τους συνδυασμούς των συγκρινόμενων εξαγωγέων χαρακτηριστικών και ταξινομητών.
5. Ο DSNPE, ενώ είναι ο καλύτερος από άποψη μέγιστης απόδοσης, μερικές φορές απαιτεί αρκετά υψηλή διάσταση για να επιτύχει το καλύτερο ποσοστό, με αποτέλεσμα σε χαμηλές διαστάσεις να υστερεί από άλλους αλγορίθμους. Αντίθετα, το SPP αν και δεν έχει το ίδιο υψηλό ποσοστό απόδοσης, όπως ο DSNPE, επιτυγχάνει πιο γρήγορα μέγιστη απόδοση ακόμα και σε μικρές διαστάσεις.

6. Κεφάλαιο

Συμπεράσματα και μελλοντική εργασία

6.1 Γενικά συμπεράσματα

Στην παρούσα διπλωματική εργασία ανακατασκευάστηκαν με επιτυχία οι δύο αλγόριθμοι για μείωση διαστάσεων, βασισμένοι στην αραιή αναπαράσταση (sparse representation) και στην πολλαπλή μάθηση (manifold learning), ο SPP και ο DSNPE. Και οι δυο μέθοδοι δίνουν προτεραιότητα στη διατήρηση των αραιών βαρών ανακατασκευής, με τον DSNPE να συνδυάζει μαζί με το κριτήριο αραιότητας και το κριτήριο μέγιστου περιθωρίου (MMC). Αρχικά, με τον SPP παρατηρήσαμε ότι υπερτερεί του PCA, LDA, LPP και LLE σε όλα τα σύνολα δεδομένων που χρησιμοποιήθηκαν και είναι πολύ απλό να εκτελεστεί όπως το PCA, αποφεύγοντας τη δυσκολία επιλογής παραμέτρων, όπως σε άλλους αλγορίθμους. Επειδή παραμένει ασαφές πώς να ορίσουμε θεωρητικά την "τοπικότητα" για πολλούς αλγορίθμους που βασίζονται στην τοπικότητα όπως το LPP, το SPP μπορεί να θεωρηθεί ως μια εναλλακτική λύση για αυτούς. Από την άλλη πλευρά, ο DSNPE έχει τόσο το πλεονέκτημα της ανθεκτικότητας της αραιής αναπαράστασης, όσο και το πλεονέκτημα της διακριτότητας του MMC, προσφέροντας έτσι μια καλή λύση για αναγνώριση μοτίβων. Η προτεινόμενη μέθοδος DSNPE εφαρμόστηκε και αυτή σε όλα τα σύνολα δεδομένων και παρατηρήσαμε εξίσου καλά αποτελέσματα, όπως στον SPP, όταν συγκρίθηκε με άλλους γνωστούς αλγορίθμους.

6.2 Προβλήματα

Ωστόσο, κάθε προσέγγιση έχει τα δικά της πλεονεκτήματα και αδυναμίες. Παρόλο που ο SPP είναι ένας πολύ δυνατός αλγόριθμος για μείωση διάστασης, εμφανίζει μια ευαισθησία σε μεγάλες διακυμάνσεις στη στάση του προσώπου, μια ευαισθησία που παρατηρείται και σε άλλους γνωστούς αλγορίθμους, όπως το PCA και το LPP. Ως αποτέλεσμα, η εφαρμογή του περιορίστηκε κυρίως σε εικόνες προσώπου από μπροστά, με διάφορες παραλλαγές στο φωτισμό και στην έκφραση. Παρόμοια, ενώ το DSNPE έχει επαινεθεί για την αποτελεσματικότητά του, εφαρμόστηκε άμεσα σε πρωτότυπες εικόνες προσώπου, αντί για εικόνες που έχουν υποστεί επεξεργασία σε τοπικά χαρακτηριστικά τους, όπως με φίλτρα Gabor ή με τοπικά δυαδικά μοτίβα (LBP). Αυτό οδήγησε σε μια παρατηρήσιμη διαφορά στην ακρίβεια αναγνώρισης προσώπου, όταν συγκρίθηκε με πρακτικές εφαρμογές.

6.3 Μελλοντικές επεκτάσεις

Για να αντιμετωπίσουμε τις ευαισθησίες του SPP, κυρίως στις παραλλαγές της στάσης του προσώπου, θα προσπαθήσουμε να αξιοποιήσουμε μια στρατηγική βασισμένη σε υπο-μοτίβο (sub-pattern). Επιπλέον, η πιθανή ένταξη εποπτευόμενης (supervised) πληροφορίας στον αλγόριθμο θα μπορούσε να αυξήσει σημαντικά τις δυνατότητές του. Για τον DSNPE, οι μελλοντικές κατευθύνσεις είναι προσανατολισμένες προς την εφαρμογή του σε προηγμένα οπτικά χαρακτηριστικά, βαίνοντας πέρα από τους περιορισμούς των αρχικών τιμών έντασης των εικονοστοιχείων (pixel). Τα σχέδια περιλαμβάνουν, επίσης, την αξιοποίηση στρατηγικών, όπως είναι η τοπικοποίηση των εικόνων εκπαίδευσης (localizing the training images) και η σύνθεση εικονικών δειγμάτων (synthesizing virtual samples). Στόχος των παραπάνω είναι η σημαντική ενίσχυση της ακρίβειας αναγνώρισης μοτίβων σε ποικίλες βιομετρικές διαδικασίες, όπως είναι η αναγνώριση προσώπου, της ίριδας και των αποτυπωμάτων της παλάμης.

Παράρτημα Α΄

Ακρωνύμια και συντομογραφίες

PCA	Principal Components Analysis
LDA	Linear Discriminant Analysis
LPP	Locality Preserving Projection
LLE	Locally Linear Embedding
SPP	Sparsity Preserving Projections
DSNPE	Discriminant Sparse Neighborhood Preserving Embedding
k-NN	k-Nearest Neighbors
SVM	Support Vector Machine
SR	Sparse Representation
LASSO	Least Absolute Shrinkage and Selection Operator
NP	Non-deterministic Polynomial time
MSR	Modified Sparse Representation
MMC	Maximum Margin Criterion
LBP	Local Binary Patterns
NumPy	Numerical Python
OS	Operating System
OpenCV	Open Source Computer Vision Library
Matplotlib	MATLAB Plotting Library for Python
CVXPY	Convex Programming in Python

7. Βιβλιογραφία

- [1] A. Jain, R. Duin, J. Mao, Statistical pattern recognition: a review, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (1) (2000) 4–37.
- [2] P. Belhumeur, J. Heapanha, D. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (7) (1997) 711–720.
- [3] D. Xu, S. Yan, D. Tao, S. Lin, H. Zhang, Marginal Fisher analysis and its variants for human gait recognition and content-based image retrieval, *IEEE Trans. Image Process.* 16 (11) (2007) 2811–2821.
- [4] H. Li, T. Jiang, K. Zhang, Efficient and robust feature extraction by maximum margin criterion, *IEEE Trans. Neural Networks* 17 (1) (2006) 157–165.
- [5] J. Liu, S. Chen, X. Tan, D. Zhang, Comments on efficient and robust feature extraction by maximum margin criterion, *IEEE Trans. Neural Networks* 18 (6) (2007) 1862–1864.
- [6] M. Turk, A. Pentland, Eigenfaces for recognition, *J. Cognitive Neurosci.* 3 (1) (1991) 71–86.
- [7] X. He, P. Niyogi, Locality preserving projections, in: *Proceedings of Conference on Advances in Neural Information Processing Systems (NIPS)*, 2003.
- [8] D. Zhang, Z. Zhou, S. Chen, Semi-supervised dimensionality reduction, in: *SIAM Conference on Data Mining (ICDM)*, 2007.
- [9] D. Cai, X. He, J. Han, Semi-supervised discriminant analysis, in: *Proceedings of International Conference on Computer Vision (ICCV)*, 2007.
- [10] Y. Song, F. Nie, C. Zhang, S. Xiang, A unified framework for semi-supervised dimensionality reduction, *Pattern Recognition* 41 (9) (2008) 2789–2799.
- [11] B. Scholkopf, A. Smola, K. Muller, Kernel principal component analysis, in: B. Scholkopf, C. Burges, A. Smola (Eds.), *Advances in Kernel Methods—Support Vector Learning*, MIT Press, Cambridge, MA, 1999, pp. 327–352.
- [12] J. Tenenbaum, Mapping a manifold of perceptual observations, in: *Advances in Neural Information Processing Systems (NIPS)*, 1998.
- [13] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science* 290 (5500) (2000) 2323–2326.
- [14] M. Belkin, P. Niyogi, Laplacian Eigenmaps for dimensionality reduction and data representation, *Neural Comput.* 15 (6) (2003) 1373–1396.
- [15] L. Maaten, E. Postma, H. Herik, Dimensionality reduction: a comparative review, available at: http://ticc.uvt.nl/~lvdrmaaten/Laurens_van_der_Maaten/Publications.html.
- [16] Y. Bengio, J. Paiement, P. Vincent, O. Delalleau, N. Roux, M. Ouimet, Out-of-sample extensions for LLE, ISOMAP, MDS, Eigenmaps, and spectral clustering, in: *Advances in Neural Information Processing Systems (NIPS)*, 2004.
- [17] D. Cai, X. He, J. Han, Spectral regression for dimensionality reduction, Technical Report UIUCDCS-R-2007-2856, Computer Science Department, UIUC, May 2007. [18] X. He, D. Cai, S. Yan, H. Zhang, Neighborhood preserving embedding, in: *Proceedings in International Conference on Computer Vision (ICCV)*, 2005. [19] Y. Fu, T. Huang, Locally linear embedded eigenspace analysis, IFP-TR, University of Illinois at Urbana-Champaign, January 2005.

- [20] D. Cai, X. He, J. Han, Isometric projection, in: Proceedings of AAAI Conference on Artificial Intelligence, 2007.
- [21] J. Wright, A. Yang, S. Sastry, Y. Ma, Robust face recognition via sparse representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (2) (2009) 210–227.
- [22] K. Huang, S. Aviyente, Sparse representation for signal classification, in: *Advances in Neural Information Processing Systems (NIPS)*, 2006.
- [23] M. Davenport, M. Duarte, M. Wakin, D. Takhar, K. Kelly, R. Baraniuk, The smashed filter for compressive classification and target recognition, in: *Proceedings of IS&T/SPIE Symposium on Electronic Imaging: Computational Imaging*, January 2007.
- [24] M. Davenport, M. Wakin, R. Baraniuk, Detection and estimation with compressive measurements, Technical Report, January 24, 2007.
- [25] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: a general framework for dimensionality reduction, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (1) (2007) 40–51.
- [26] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using Laplacianfaces, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (3) (2005).
- [27] J. Ham, D. Lee, S. Mika, B. Scholkopf, A kernel view of the dimensionality reduction of manifolds, in: *Proceedings of International Conference on Machine Learning*, 2004, pp. 47–54.
- [28] J. Murray, K. Kreutz-Delgado, Visual recognition and inference using dynamic overcomplete sparse learning, *Neural Comput.* 19 (2007) 2301–2352.
- [29] R. Duda, P. Hart, D. Stork, *Pattern Classification*, second ed., Wiley, New York, 2001.
- [30] M. Marcellin, M. Gormish, A. Bilgin, M. Boliek, An overview of JPEG-2000, in: *Proceedings of the Data Compression Conference*, 2000.
- [31] M. Elad, M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, *IEEE Trans. Image Process.* 15 (12) (2006) 3736–3745.
- [32] J. Yang, J. Wright, Y. Ma, T. Huang, Image super-resolution as sparse representation of raw image patches, in: *Computer Vision and Pattern Recognition (CVPR)*, 2008.
- [33] R. Tibshirani, Regression shrinkage and selection via the LASSO, *J. R. Stat. Soc. B* 58 (1) (1996) 267–288.
- [34] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, *J. R. Stat. Soc. Ser. B* 67 (2) (2005) 301–320.
- [35] R. Baraniuk, A lecture on compressive sensing, *IEEE Signal Process. Mag.* 24 (4) (2007) 118–121.
- [36] D. Donoho, Compressed sensing, *IEEE Trans. Inf. Theory* 52 (4) (2006) 1289–1306.
- [37] A. Yang, J. Wright, Y. Ma, S. Sastry, Feature selection in face recognition: a sparse representation perspective, UC Berkeley Technical Report UCB/EECS-2007-99, 2007.
- [38] S. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal Process.* 41 (12) (1993) 3397–3415.
- [39] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, *IEEE Trans. Signal Process.* 56 (6) (2008) 2346–2356.
- [40] M. Wu, Kai. Yu, S. Yu, B. Scholkopf, Local learning projections, in: *International Conference on Machine Learning (ICML)*, 2007.
- [41] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, *SIAM Rev.* 43 (1) (2001) 129–159.

- [42] D. Donoho, Y. Tsaig, Fast solution of l_1 -norm minimization problems when the solution may be sparse, Technical Report, Institute for Computational and Mathematics and Engineering, Stanford University, USA, 2006.
- [43] M.E. Tipping, Sparse Bayesian learning and the relevance vector machine, *J. Mach. Learn. Res.* 1 (2001) 211–244.
- [44] D. Cai, X. He, J. Han, Spectral regression: a unified approach for sparse subspace learning, in: *Proceedings of International Conference on Data Mining (ICDM)*, 2007.
- [45] H. Zhou, T. Hastie, R. Tibshirani, Sparse principle component analysis, Technical Report, Statistics Department, Stanford University, USA, 2004.
- [46] R. Zass, A. Shashua, Non-negative sparse PCA, in: *Advances in Neural Information Processing systems (NIPS)*, 2007.
- [47] P. Hoyer, Non-negative matrix factorization with sparseness constraints, *J. Mach. Learn. Res.* 5 (2004) 1457–1469.
- [48] S. Chen, Y. Zhu, Subpattern-based principal component analysis, *Pattern Recognition* 37 (1) (2004) 1081–1083.
- [49] A.M. Martinez, A.C. Kak, PCA versus LDA, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (2) (2001) 228–233.
- [50] K. Lee, J. Ho, D. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (5) (2005) 684–698.
- [51] K. Zhang, J.T. Kwok, Density-weighted Nystrom method for computing large kernel eigen-systems, *Neural Comput.* 21 (1) (2009) 121–146.