Αναφορα για τη 1η εργασια

Εισαγωγη

Στην εργασια αυτη συγκρινουμε τρεις κατηγοριες ταξινομητων:

- **Support Vector Machines (SVMs)**
- **k-Nearest Neighbors (kNN)**
- **Nearest Class Centroid (NCC)**

σε δυο διαφορετικα συνολα δεδομενων:

1. **CIFAR-10** (classification 10 κλασεων)
2. **Breast cancer** (δυαδικη ταξινομηση καλοηθες / κακοηθες)

Δινουμε ιδιαιτερη βαρυτητα στα **SVMs**, τα οποια χρησιμοποιουμε ως βασικη γραμμη συγκρισης και τα οποια σε γενικες γραμμες επιτυγχανουν τις καλυτερες επιδοσεις, ειδικα σε συνδυασμο με καταλληλη επιλογη χαρακτηριστικων.

Κεντρικος στοχος ειναι να δειξουμε πως συμπεριφερονται τα SVMs σε σχεση με απλους ταξινομητες αποστασης (kNN, NCC) καθως και πως επηρεαζει η διασταση και ο τυπος των χαρακτηριστικων την αποδοση των μοντελων.

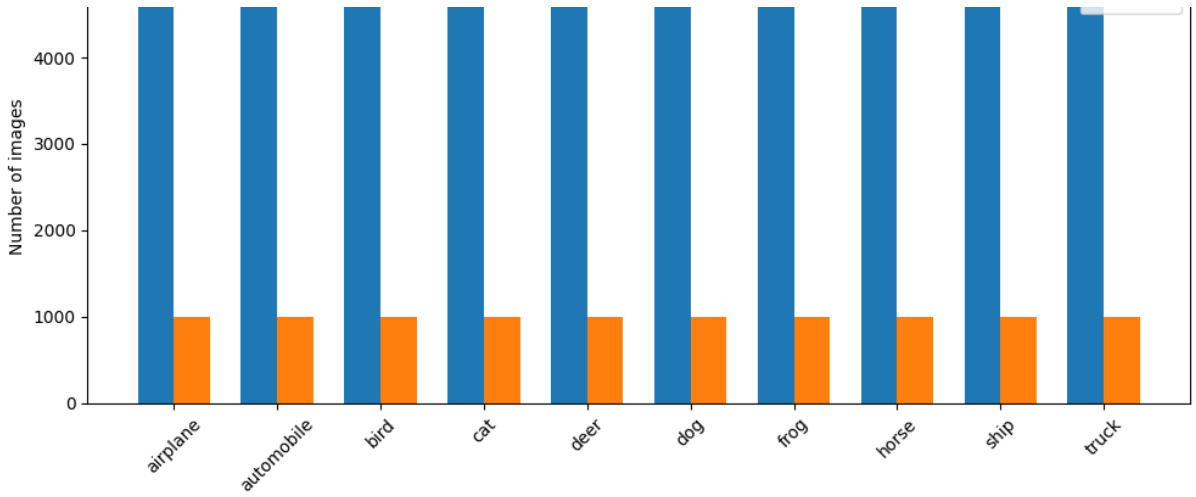
Datasets και Pretraining

CIFAR-10

Για την πειραματικη μελετη χρησιμοποιηθηκε το dataset CIFAR-10, το οποιο περιλαμβανει 10 κλασεις (airplane, automobile, bird, truck, etc) με εικονες διαστασεων 32x32x3 (RGB). Το συνολο δεδομενων αποτελειται απο 50 000 δειγματα για το train set και 10 000 δειγματα για το test set.

Table 1 Τα 5 πρωτα παραδειγματα του CIFAR-10

Table 2 Τα δεδομενα ειναι ισοκατανεμημενα.



Ως βηματα προεπεξεργασιας, αρχικα εγινε φορτωση των δεδομενων. Στην συνεχεια, οι εικονες μετατραπηκαν σε διανυσματα με την εντολη

```
X_train_flat = X_train.reshape(len(X_train), -1)
```

και αντιστοιχα για το test set, ετσι ωστε καθε εικονα να αναπαρισταται ως ενα ενιαιο διανυσμα χαρακτηριστικων. Οποτε από τις διαστασεις (32, 32, 3) παμε στις flattened 3072. Επιπλεον, οι τιμες των pixels κλιμακωθηκαν στην περιοχη [0, 1] με διαιρεση δια 255, ωστε να βελτιωθει η αριθμητικη σταθεροτητα και η αποδοση των ταξινομητων.

Τα δεδομενα χρησιμοποιηθηκαν σε μορφη raw pixels για την εκπαιδευση ταξινομητων SVM, kNN και Nearest Class Centroid (NCC). Επιπλεον, εξηχθησαν HOG χαρακτηριστικα πανω στις ιδιες εικονες και ξαναεκπαιδευτηκαν πανω σε αυτα, επιτυγχανοντας αισθητα καλυτερες επιδοσεις σε σχεση με τα αντιστοιχα μοντελα που εκπαιδευτηκαν απευθειας στα raw pixels.

Table 3 Παραδειγμα HOG μετασχηματισμου.

Το **HOG** (*Histogram of Oriented Gradients*) υπολογιζει για καθε μικρη περιοχη της εικονας (cell) gradients και φτιαχνει ιστογραμμα των κατευθυνσεων τους. Τα histograms αυτα ομαδοποιουνται και κανονικοποιουνται σε blocks, ετσι κωδικοποιουνται οι ακμες και τα σχηματα της εικονας, και ετσι δεν δινεται βαρυτητα στις ακριβεις τιμες των pixels. Με αυτο τον τροπο, τα HOG features ειναι πιο συμπαγη, πιο ανθεκτικα σε αλλαγες φωτισμου και θορυβο και αναδεικνυουν τη δομη του

αντικειμενου, κατι που επιτρεπει στα μοντελα να βρουν ενα καθαροτερο και πιο γραμμικο οριο αποφασης σε σχεση με την εκπαιδευση πανω σε raw pixels.

Breast Cancer

Για το δευτερο μερος της εργασιας χρησιμοποιηθηκε το dataset Breast Cancer, το οποιο αντιστοιχει σε προβλημα δυαδικης ταξινομησης μεταξυ καλοηθων και κακοηθων ογκων (κλασεις 2 για benign και 4 για malignant). Καθε δειγμα από τα 683, που είναι στο συνολο, περιγραφεται απο 9 χαρακτηριστικα, τα οποια ποσοτικοποιουν ιδιοτητες των κυτταρων σε κλιμακα απο 1 εως 10: Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli και Mitoses. Στα χαρακτηριστικα αυτα χρησιμοποιουμε την κλιμακωμενη (scaled) εκδοχη τους στο διαστημα [-1,1], και για τα class labels από 2,4 μετατραπηκαν στις τιμες 0,1, γεγονος που διευκολυνει την εκπαιδευση των μοντελων και την συγκρισιμοτητα μεταξυ των διαφορων διαστασεων του χωρου χαρακτηριστικων.

Για τον διαχωρισμο των δεδομενων σε συνολα εκπαιδευσης και ελεγχου χρησιμοποιηθηκε η συναρτηση train_test_split με ορισμα test_size=0.3, ωστε το 30% των δειγματων να αποτελει το test set και το υπολοιπο 70% το train set.

```
Train Shape:(478, 9)

Test Shape: (205, 9)
```

Επιπλεον, χρησιμοποιησαμε random_state=0 για αναπαραγωγιμοτητα και χρησιμοποιηθηκε η παραμετρος stratify=y, ωστε να διασφαλιστει οτι η αναλογια των δυο κλασεων διατηρειται περιπου ιδια και στα συνολα train και test. Με τον τροπο αυτο αποτρεπεται η δημιουργια μη αντιπροσωπευτικων υποσυνολων και βελτιωνεται η αξιοπιστια της αξιολογησης των ταξινομητων.

Models

SVMs

Τα Support Vector Machines χρησιμοποιηθηκαν ως κυριο μοντελο αναφορας για την συγκριση των υπολοιπων μεθοδων. Η βασικη ιδεα των SVMs ειναι η αναζητηση ενος υπερεπιπεδου που διαχωριζει τις κλασεις με μεγιστο περιθωριο, μεσα απο την λυση ενος προβληματος βελτιστοποιησης που ισορροπει μεταξυ μεγιστοποιησης του περιθωριου (ορος regularization) και μειωσης του σφαλματος ταξινομησης, μεσω της παραμετρου C που ελεγχει το trade off μεταξυ των δυο στοχων. Στην εργασια δοκιμαστηκαν τοσο γραμμικα οσο και μη γραμμικα SVMs, χρησιμοποιωντας τις υλοποιησεις LinearSVC και SVC της βιβλιοθηκης scikit learn.

CIFAR 10

Στο CIFAR 10, λογω του υψηλου υπολογιστικου κοστους σε ολοκληρο το dataset, ο κυριος πειραματισμος με τις υπερπαραμετρους εγινε σε ενα stratified subset (σταθερος αριθμος δειγματων ανα κλαση), οπου δοκιμαστηκαν γραμμικα SVMs (LinearSVC), SVM με πολυωνυμικο kernel βαθμου 3, καθως και RBF SVM με διαφορετικους συνδυασμους των C και gamma το οποιο ελεγχει ποσο "κοντα" πρεπει να ειναι δυο δειγματα για να επηρεαζουν πολυ το ενα το αλλο. Αν το gamma είναι μεγαλο καθε δειγμα εχει μικρη εμβελεια και το SVM μπορει να φτιαξει πιο πολυπλοκα συνορα με κινδυνο, όμως, να παθει overfitting. Ενώ αν είναι μικρο το gamma, καθε δειγμα επηρεαζει μεγαλυτερη περιοχη, τα συνορα γινονται πιο λεια και "απλα".

Τα αποτελεσματα είναι τα εξης:

Table 4 CIFAR 10 – SVM σε raw pixels (stratified subset, 10 000 train / 2 000 test)

Μοντελο	**Χρονος train**	**Test accuracy**	**Macro F1**
LinearSVC_C1	27.46 sec	0.3385	0.3191
SVC_poly_deg3_C1	3 min 35 sec	0.4400	0.4412
SVC_linear_C1	6 min 2 sec	0.2540	0.2167
SVC_rbf_C1_gamma1e-3	6 min 19 sec	0.2430	0.1965
SVC_rbf_C10_gamma1e-3	6 min 18 sec	0.2430	0.1965
SVC_rbf_C100_gamma1e-3	377.75	0.2430	0.1965

Στην συνεχεια τρεξαμε ξανα στο full dataset για το γραμμικο μοντελο λογω του ότι είναι το πιο γρηγορο αλλα και για το πολυωνυμικο το οποιο ειχε τα καλυτερα αποτελεσματα:

Table 5 CIFAR 10 – SVM σε raw pixels (full dataset, 50 000 train / 10 000 test)

Μοντελο	**Χρονος train**	**Test accuracy**	**Macro F1**
LinearSVC_C1	3 min 5 sec	0.3854	0.3707
SVC_poly_deg3_C1	1 hr 31 min	0.5244	0.5254

Στη συνεχεια, υπολογιστηκαν **HOG** χαρακτηριστικα για τις εικονες και εκπαιδευτηκαν εκ νεου SVMs πανω σε αυτο το χωρο χαρακτηριστικων. Στο pipeline χρησιμοποιειται αρχικα το StandardScaler, το οποιο αφαιρει τη μεση τιμη και διαιρει με την τυπικη αποκλιση καθε χαρακτηριστικου, ωστε ολα τα features να βρισκονται σε συγκρισιμη κλιμακα. Αυτο ειναι σημαντικο για τον RBF πυρηνα, επειδη βασιζεται σε ευκλειδειες αποστασεις οπου χωρις καλη κλιμακωση, λιγα features με μεγαλες τιμες θα κυριαρχουσαν στον υπολογισμο της αποστασης και το SVM δεν θα μπορουσε να μαθει ενα σταθερο και καλα γενικευσιμο συνορο αποφασης.

Δοκιμαζουμε LinearSVC και στη συνεχεια RBF SVM με C = 10 και gamma = "scale", στο ιδιο stratified subset. Το “scale” σημαινει ότι αν τα features εχουν μεγαλη διασπορα, το gamma γινεται μικρο, ενώ αν τα features εχουν μικρη διασπορα, το gamma γινεται μεγαλυτερο. Ο συνδυασμος RBF SVM σε HOG χαρακτηριστικα, αποδειχθηκε ο πιο αποδοτικος και για αυτο εκπαιδευτηκε και σε ολοκληρο το CIFAR 10, αποτελωντας το τελικο SVM baseline για το συγκεκριμενο dataset.

Table 6 CIFAR 10 – SVM σε HOG χαρακτηριστικα (stratified subset, 10 000 train / 2 000 test)

Μοντελο	**Χρονος train (sec)**	**Test accuracy**	**Macro F1**
LinearSVC σε HOG	15.54	0.4780	0.4713
RBF SVC σε HOG	25.19	0.5590	0.5585

Τελος, τρεχουμε και στο full dataset το RBF αφου παρατηρουμε ότι τα results είναι promising με δυο διαφορετικες τιμες C.

Table 7 CIFAR 10 – SVM σε HOG χαρακτηριστικα (full dataset, 50 000 train / 10 000 test)

Μοντελο	**Χρονος train (sec)**	**Test accuracy**	**Macro F1**
RBF SVC σε HOG full (C = **10**, gamma=scale)	1087.53	0.6362	0.6360
RBF SVC σε HOG full (C = **150**, gamma=scale)	1203.82	0.6366	0.6364

Συμπερασματικα, σε raw pixels τα SVMs δινουν μετρια αποδοση στο CIFAR 10. Ο πολυωνυμικος kernel σε full dataset βελτιωνει σημαντικα τα metrics, ενώ επιπλεον βελτίωση προκυπτει με την χρηση HOG χαρακτηριστικων. Με RBF SVM σε HOG (ιδιως στο full dataset με C = 150) αποτελει το *καλυτερο* SVM baseline για CIFAR 10 στην πειραματικη διαδικασια.

Breast Cancer

Για αυτό το dataset για την επιλογη υπερπαραμετρων χρησιμοποιηθηκε 5-fold stratified cross validation (StratifiedKFold με n_splits=5, shuffle=True, random_state=0) σε συνδυασμο με Grid Search (GridSearchCV), με κριτηριο scoring="f1" πανω στο training set. Δοκιμαστηκαν τρια βασικα SVM μοντελα ωστε να συγκριθουν γραμμικα και μη γραμμικα συνορα αποφασης στο ιδιο dataset.

- SVC με γραμμικο kernel (kernel="linear"),
- SVC με RBF kernel (kernel="rbf") και
- SVC με πολυωνυμικο kernel (kernel="poly") με βαθμο 3.

Table 8 Αποτελεσματα SVM στο breast cancer (70/30 split, 5 fold CV).

Μοντελο	**Best params**	**Χρονος train (sec, wall)**	**Test accuracy**	**Test F1**
SVM_poly	C=0.1, degree=3, coef0=0.0, gamma='scale'	0.42	0.9756	0.9660
SVM_linear	C=0.1	2.05	0.9610	0.9444
SVM_rbf	C=10, gamma=0.01	1.45	0.9610	0.9444

Ποιοτικα, ολα τα SVM μοντελα αποδιδουν πολυ καλα στο breast cancer dataset. Οι τιμες test accuracy ειναι πανω απο 0.96, ενω το F1 score κινειται περιπου στο ευρος 0.94–0.97. Οπως φαινεται και απο τα classification reports, υπαρχει πολυ καλη ισορροπια μεταξυ precision και recall και για τις δυο κλασεις (benign και malignant), χωρις καποια εμφανη μεροληψια υπερ της μιας ή της αλλης.

Το SVM με polynomial kernel βαθμου 3 (SVM_poly) προκυπτει ως το καλυτερο μοντελο στο notebook και πετυχαινει accuracy γυρω στο 0.976 και F1 περιπου 0.966. Στην confusion matrix βλεπουμε οτι η benign κλαση ταξινομειται σχεδον τελεια (129 σωστα, 4 λαθος), ενω και για την malignant κλαση οι επιδοσεις ειναι αντιστοιχα πολυ καλες (71 σωστα, 1 λαθος). Επιπλεον, ολα αυτα επιτυγχανονται με πολυ μικρο χρονο εκπαιδευσης καθως προκειται για ένα μικρο dataset.

Τα SVMs με γραμμικο kernel (SVM_linear) και με RBF kernel (SVM_rbf) εμφανιζουν σχεδον ταυτιζομενα αποτελεσματα: accuracy γυρω στο 0.961 και F1 περιπου 0.944, με confusion matrix [[129, 4], [4, 68]]. Αυτα τα μοντελα προσφερουν ηδη πολυ υψηλη αποδοση και λειτουργουν ως ισχυρα baselines, δειχνοντας οτι ακομα και σχετικα απλα SVMs, πανω σε καλα χαρακτηριστικα, ειναι ικανα να λυσουν αποτελεσματικα το συγκεκριμενο προβλημα ταξινομησης.

k-Nearest Neighbors (kNN) και Nearest Class Centroid (NCC)

CIFAR 10

Ουσιαστικα, και τα δυο μοντελα (kNN και Nearest Class Centroid) βασιζονται στις αποστασεις μεταξυ δειγματων στον χωρο των χαρακτηριστικων για να προβλεψουν την κλαση. Δεν κανουν πολυπλοκη μαθηση παραμετρων οπως τα SVMs, αλλα χρησιμοποιουν την γεωμετρια των δεδομενων για να παρουν αποφασεις ταξινομησης.

Στο **kNN**, για καθε νεο δειγμα αναζητουνται οι k κοντινοτεροι γειτονες (π.χ. με ευκλειδια ή manhattan αποσταση) και η προβλεψη βασιζεται στην πλειοψηφια των κλασεων τους. Οταν χρησιμοποιουμε weights="distance", οι πιο κοντινοι γειτονες εχουν μεγαλυτερη επιδραση στην αποφαση, κατι που μπορει να βελτιωσει την αποδοση σε περιπτωσεις οπου η τοπικη γειτονια ειναι πυκνη.

Στο Nearest Class Centroid (**NCC**), η λογικη ειναι πιο απλη. Για καθε κλαση υπολογιζεται ενα centroid, δηλαδη το μεσο διανυσμα ολων των δειγματων της. Ενα νεο δειγμα ταξινομειται στην κλαση της οποιας το centroid βρισκεται πιο κοντα. Ετσι, καθε κλαση αναπαρισταται απο ενα μονο σημειο στον χωρο των χαρακτηριστικων, κατι που κανει το μοντελο ιδιαιτερα γρηγορο και ευκολα ερμηνευσιμο, αλλα λιγοτερο αποδοτικο σε σχεση με πιο πολυπλοκα μοντελα.

Όπως και στα SVMs αρχικα δουλευουμε με ένα subset του CIFAR-10, ώστε να βρουμε γρηγορα τις καλυτερες παραμετρους. Επιλεγουμε ενα stratified subset με 10,000 δειγματα για train και 2,000 για test, διατηρωντας την ιδια αναλογια κλασεων οπως στο αρχικο συνολο. Για το *kNN* χρησιμοποιηθηκε Grid Search με ενα πλεγμα τιμων για τον αριθμο γειτονων n_neighbors (1, 3, 5, 7, 9, 11), δυο μετρικες αποστασης (euclidean, manhattan) και δυο σχηματα βαρων (uniform, distance), ενω ο αλγοριθμος αναζητησης γειτονων ειναι στο "auto" ωστε να επιλεγεται αυτοματα η καταλληλοτερη στρατηγικη. Αντιστοιχα, για το *Nearest Class Centroid* οριστηκε grid πανω στη metric (euclidean, manhattan) και στην παραμετρο shrink_threshold, η οποια μπορει να εφαρμοσει shrinkage στα centroids (τιμες [None, 0.05, 0.1, 0.5, 1.0]), με στοχο να βελτιωθει η σταθεροτητα σε θορυβωδη δεδομενα υψηλης διαστασης.

Και στις δυο περιπτωσεις η επιλογη των υπερ-παραμετρων εγινε με GridSearchCV, χρησιμοποιωντας scoring="accuracy", 3-fold cross validation (cv=3) και n_jobs=-1 για παραλληλη εκτελεση, ωστε να βρεθουν οι συνδυασμοι που δινουν την καλυτερη αποδοση στο stratified subset του CIFAR 10.

Table 9 Αποτελεσματα subset του CIFAR 10 για k Nearest Neighbors (kNN) και Nearest Class Centroid (NCC).

Μοντελο	**Καλυτερα params**	**Χρονος εκπαιδευσης (GridSearchCV, sec)**	**Test accuracy**	**Test macro F1**
kNN	algorithm = auto, metric = manhattan, n_neighbors = 9, weights = distance	268.50	0.3225	0.3100

Μοντελο	Καλυτερα params	Χρονος εκπαιδευσης (GridSearchCV, sec)	Test accuracy	Test macro F1
NCC	metric = manhattan, shrink_threshold = None	7.96	0.2615	0.2416

Επειτα, με τις καλυτερες υπερ-παραμετρους που βρισκουμε το τρεχουμε στο full dataset (50 000 train / 10 000 test), οπου παρατηρειται βελτιωση στον kNN.

Table 10 Αποτελεσματα στο full dataset CIFAR 10, για k Nearest Neighbors (kNN) και Nearest Class Centroid (NCC).

Μοντελο	Params (απο subset)	Χρονος fit (sec)	Test accuracy	Test macro F1
kNN	algorithm = auto, metric = manhattan, n_neighbors = 9, weights = distance	0.17	0.3952	0.3895
NCC	metric = manhattan, shrink_threshold = None	8.23	0.2734	0.2528

Στην πειραματικη μας διαδικασια για να εχουμε σωστη συγκριση με τα SVMs εφαρμοστηκε η ιδια μεθοδος με τα HOG features όπως περιγραφηκε στη ενοτητα SVMs και βρηκαμε αυτά τα αποτελεσματα:

Table 11 CIFAR 10 – kNN / NCC σε HOG + StandardScaler (stratified subset, 10 000 train / 2 000 test)

Μοντελο	Best params (απο grid σε raw pixels)	Χρονος fit (sec)	Test accuracy	Test macro F1
kNN σε HOG	algorithm = auto, metric = manhattan, n_neighbors = 9, weights = distance	0.00	0.4780	0.4681
NCC σε HOG	metric = manhattan, shrink_threshold = None	0.08	0.4070	0.4031

Table 12 CIFAR 10 – kNN / NCC σε HOG + StandardScaler (FULL dataset, 50 000 train / 10 000 test)

Μοντελο	Best params (απο grid σε raw pixels)	Χρονος fit (sec)	Test accuracy	Test macro F1
kNN σε HOG full	algorithm = auto, metric = manhattan, n_neighbors = 9, weights = distance	0.01	0.5435	0.5356

Μοντελο	Best params (απο grid σε raw pixels)	Χρονος fit (sec)	Test accuracy	Test macro F1
NCC σε HOG full	metric = manhattan, shrink_threshold = None	0.40	0.4084	0.4034

Breast Cancer

Για τον ταξινομητη **kNN** εφαρμοστηκε διαδικασια grid search προκειμενου να επιλεγουν οι βελτιστες υπερπαραμετροι. Συγκεκριμενα, δοκιμαστηκαν διαφορες τιμες για τον αριθμο γειτονων n_neighbors ([1, 3, 5, 7, 9, 11, 15]), δυο μετρικα αποστασης (euclidean και manhattan), καθως και δυο σχηματα βαρων (uniform και distance). Η εκπαιδευση εγινε με χρηση Grid Search πανω στο train set, το οποιο περιλαμβανει 478 δειγματα, χρησιμοποιωντας 5 fold stratified cross validation ωστε να διατηρειται η αναλογια των κλασεων σε καθε fold. Αφου ολοκληρωθει το grid search, το μοντελο επανεκπαιδευεται (refit) στα συνολικα δεδομενα εκπαίδευσης χρησιμοποιωντας τον συνδυασμο υπερπαραμετρων που εδωσε τη μεγιστη τιμη στην cross-validation accuracy.

Για την αξιολογηση χρησιμοποιειται το test set, το οποιο αποτελειται απο 205 δειγματα. Ως μετρικες επιδοσης υπολογιζονται η accuracy και η F1, σε δυαδικη μορφη, με τη θετικη κλαση να αντιστοιχει στα malignant περιστατικα.

Table 13 Αποτελεσματα kNN

Μοντελο	Best params	Χρονος train (sec)	Test accuracy	Test F1 (binary)
kNN	metric = euclidean, n_neighbors = 7, weights = uniform	2.7610	0.9610	0.9437

Για τον ταξινομητη Nearest Class Centroid (**NCC**) εφαρμοστηκε αντιστοιχη διαδικασια grid search ωστε να επιλεγουν οι βελτιστες υπερπαραμετροι. Συγκεκριμενα, δοκιμαστηκαν δυο μετρικα αποστασης (euclidean και manhattan), καθως και διαφορετικες τιμες για την παραμετρο shrink_threshold ([None, 0.05, 0.1, 0.5, 1.0]), η οποια ελεγχει τον βαθμο συρρικνωσης των κεντροειδων. Ετσι, παλι με Grid Search εχουμε τα εξης αποτελεσματα:

Table 14 Αποτελεσματα NCC.

Μοντελο	Best params	Χρονος train (sec)	Test accuracy	Test F1 (binary)
NCC	metric = euclidean, shrink_threshold = None	0.1149	0.9659	0.9510

Συμπερασματικά, και ο kNN και ο NCC πετυχαινουν πολυ υψηλη επιδοση στο breast cancer dataset. Ο απλος γραμμικος NCC φτανει και ξεπερνα ελαφρως τον kNN, ενω εκπαιδευεται σημαντικα πιο γρηγορα.

Συζητηση

Τα αποτελεσματα της μελετης δειχνουν καθαρα οτι ο ρολος των χαρακτηριστικων ειναι καθοριστικος για την αποδοση των ταξινομητων. Οταν χρησιμοποιουμε raw pixels, η ευκλειδια ή η manhattan αποσταση μεταξυ των εικονων δεν αντανακλα παντα την πραγματικη οπτικη ομοιοτητα, με αποτελεσμα τα μοντελα βασισμενα σε αποστασεις, οπως kNN και NCC, να παρουσιαζουν περιορισμενη επιδοση. Ακομη και τα SVMs, σε αυτη τη μορφη δεδομενων, χρειαζονται ισχυρα kernels για να μπορεσουν να μοντελοποιησουν σωστα τα ορια αποφασης. Αντιθετα, με HOG χαρακτηριστικα στο CIFAR-10, η αποδοση των SVMs βελτιωνεται θεαματικα, δειχνοντας οτι η σωστη αναπαρασταση των δεδομενων ειναι κρισιμη προυποθεση για καλα αποτελεσματα.

Τα SVMs αναδεικνυονται ως πολυ καλη γενικη λυση, ιδιως σε προβληματα υψηλης διαστασης. Σε τετοιες συνθηκες μπορουν να διαχειριστουν καλυτερα το προβλημα σε σχεση με τον kNN/NCC, καθως δεν βασιζονται στον υπολογισμο πληρων αποστασεων απο ολα τα δειγματα του train set για καθε νεο δειγμα. Με την χρηση του kernel trick (π.χ. RBF kernel) μπορουν να προσαρμοσουν πολυπλοκα, μη-γραμμικα ορια αποφασης στον χωρο των χαρακτηριστικων, διατηρωντας παραλληλα καλη ικανοτητα γενικευσης.

Όμως, και τα μοντελα αποστασης, οπως kNN και NCC ειναι ιδιαιτερα χρησιμα ως baselines, ιδιως σε μικρα, καλα κλιμακωμενα datasets. Στο breast cancer dataset, για παραδειγμα, kNN και NCC παρεχουν γρηγορα και ευκολα στην υλοποιηση μοντελα με αποδοση που συχνα πλησιαζει αυτη των SVMs, ιδιαιτερα οταν η σχεση μεταξυ χαρακτηριστικων και κλασης ειναι σχετικα απλη. Ο NCC, συγκεκριμενα, προσφερει ενα εξαιρετικα ελαφρυ υπολογιστικα μοντελο, το οποιο, παρα την απλοτητα του, μπορει σε ορισμενες περιπτωσεις να πλησιασει ή και να ξεπερασει πιο συνθετους ταξινομητες.