

## 2<sup>η</sup> εργασία του μαθήματος «Νευρωνικά Δίκτυα»

### Χρυσολόγου Γεώργιος (ΑΕΜ: 10782)

Για την εκπόνηση της παρούσας εργασίας, κλήθηκα να υλοποιήσω ένα Support Vector Machine το οποίο θα επιλύει το πρόβλημα κατηγοριοποίησης πολλαπλών κλάσεων. Επέλεξα να κατασκευάσω το πρόγραμμα σε γλώσσα python, χωρίς την χρήση των σχετικών frameworks (Scikit-learn, libsvm, κλπ), το οποίο εκπαιδεύεται στο data set Cifar-10.

**Περιγραφή αλγορίθμου:** Ο αλγόριθμος kernel SVM χρησιμοποιείται για την ταξινόμηση δεδομένων τα οποία δεν είναι γραμμικά διαχωρίσιμα, όπως συμβαίνει στο data set Cifar-10. Συγκεκριμένα, το support vector machine βρίσκει μία υπερεπιφάνεια που διαχωρίζει τα δείγματα δύο διαφορετικών κλάσεων. Με χρήση του kernel trick, είναι δυνατόν να βρεθεί μία υψηλότερη διάσταση, στην οποία τα δεδομένα προβάλλονται και είναι γραμμικά διαχωρίσιμα.

Η υπερεπιφάνεια αυτή καθορίζεται από ορισμένα δείγματα τα οποία ονομάζονται support vectors. Τα support vectors ορίζονται ως τα δείγματα που βρίσκονται πλησιέστερα στην διαχωριστική επιφάνεια. Η θέση και ο προσανατολισμός της επηρεάζονται αποκλειστικά από αυτά τα δείγματα, και όχι από τα υπόλοιπα.

Για την πραγματοποίηση του kernel trick, χρησιμοποιείται μία kernel συνάρτηση, η οποία μετασχηματίζει τα δείγματα σε έναν χώρο υψηλότερης διάστασης. Στην εργασία αυτή, χρησιμοποιήθηκαν ο γραμμικός kernel, ο πολυωνυμικός kernel διαφορετικών βαθμών και ο RBF kernel.

Σύμφωνα με την μέθοδο του SVM, επιλύεται ένα τετραγωνικό πρόβλημα βελτιστοποίησης με στόχο την μεγιστοποίηση του περιθωρίου μεταξύ των δειγμάτων των δύο διαφορετικών κλάσεων και, παράλληλα, την ελαχιστοποίηση των σφαλμάτων κατηγοριοποίησης.

**Περιγραφή κώδικα:** Σύμφωνα με τον κώδικα, αρχικά, αποθηκεύονται τα 50000 δεδομένα εκπαίδευσης σε έναν πίνακα 50000 x 3072 με όνομα data\_tr, όπου κάθε γραμμή αποτελεί ένα δείγμα εκπαίδευσης. Αντίστοιχα, αποθηκεύονται τα 10000 δείγματα ελέγχου σε έναν πίνακα 10000 x 3072 με όνομα data\_test, όπου κάθε γραμμή αποτελεί ένα δείγμα ελέγχου.

Τα δεδομένα αυτά κανονικοποιούνται μέσω διαίρεσης όλων των στοιχείων των δύο παραπάνω πινάκων δια 255. Επιπλέον, εφαρμόζεται η μέθοδος PCA, όπου εξασφαλίζεται η συγκράτηση του 90% της πληροφορίας, για την μείωση των διαστάσεων των πινάκων, με σκοπό την απλοποίηση του προβλήματος, την επιτάχυνση των υπολογισμών και τον περιορισμό του συνολικού χρόνου εκπαίδευσης.

Για την πραγματοποίηση κατηγοριοποίησης πολλών κλάσεων, και συγκεκριμένα 10 για το data set Cifar-10, επιλέχθηκε η χρήση της στρατηγικής one vs all. Για την υλοποίηση της, είναι απαραίτητη η κατασκευή 10 διαφορετικών SVMs, καθένα από τα οποία αντιστοιχεί σε μία κλάση και διαχωρίζει τα δείγματα αυτής από τα δείγματα όλων των υπολοίπων κλάσεων. Τα δείγματα που ανήκουν στην κλάση στην οποία αντιστοιχεί το εκάστοτε SVM κατηγοριοποιούνται στην κλάση +1, ενώ τα υπόλοιπα στην κλάση -1. Για την αντιμετώπιση του προβλήματος αδυναμίας κατασκευής του kernel matrix  $H$  με διαστάσεις 50000x50000 (τα training δείγματα είναι συνολικά 50000), επιλέχθηκε η εκπαίδευση με χρήση batches training δειγμάτων. Αναλυτικότερα, ορίζονται batches δειγμάτων ίδιου μεγέθους, των οποίων, όπως παρατηρήθηκε, κατά μέσο όρο το 10% των συνολικών δειγμάτων ανήκουν στην κλάση +1 ενώ το 90% στην κλάση -1. Το 1<sup>ο</sup> batch αποτελείται από τα πρώτα batch size, σε αριθμό, δείγματα (γραμμές) του πίνακα data\_tr, το 2<sup>ο</sup> batch από τα επόμενα batch size, σε αριθμό, δείγματα και ούτω καθεξής. Στην συνέχεια, δημιουργείται για κάθε batch ένα SVM, το οποίο εκπαιδεύεται στα training δείγματα αυτού του batch. Πρώτο βήμα της εκπαίδευσης αποτελεί ο υπολογισμός του kernel matrix  $H$  διαστάσεων (batch size)x(batch size), μέσω της σχέσης  $H[i, j] = Y[i] * Y[j] * Kernel(X[i], X[j])$ , όπου για  $X[i]$  και  $X[j]$  τα training

δείγματα  $i$  και  $j$ ,  $Y[i]$  και  $Y[j]$  τα αντίστοιχα labels (+1 ή -1), για κάθε στοιχείο  $H[i, j]$  του πίνακα.

Ακολουθεί η επίλυση του τετραγωνικού προβλήματος ελαχιστοποίησης με αντικειμενική συνάρτηση την εξής:

$\frac{1}{2} * a^T * H * a - c^T a * a$ , όπου  $a$  οι πολλαπλασιαστές Lagrange οι οποίοι υπολογίζονται μέσω της επίλυσης του τετραγωνικού προβλήματος,  $H$  ο kernel matrix και  $c$  ένα διάνυσμα διάστασης batch size με όλα τα στοιχεία του μονάδες, υπό τους εξής περιορισμούς:

$a \geq 0$  και  $a \leq C$ , όπου  $C$  παράμετρος που επιλέγεται από τον χρήστη και καθορίζει το βαθμό σημαντικότητας μεταξύ της εξασφάλισης μεγάλου περιθωρίου των δειγμάτων των δύο κλάσεων από την υπερεπιφάνεια και των σφαλμάτων κατηγοριοποίησης.

Ως τελευταίο στάδιο της εκπαίδευσης, υπολογίζεται το bias με τον εξής τρόπο:

Για κάθε support vector, υπολογίζεται ένα bias  $b_{sv} = Y[sv] - \sum_{i=1}^{batch\_size} a_i * Y[i] * Kernel(X[i], X[sv])$ , όπου  $X[sv]$  το support vectors και  $Y[sv]$  η ετικέτα του (+1 ή -1),  $X[i]$  training δείγμα του batch και  $Y[i]$  η ετικέτα του (+1 ή -1),  $a_i$  ο  $i$ -οστός πολλαπλασιαστής Lagrange.

Κατόπιν, υπολογίζεται το bias ως ο μέσος όρος όλων των biases  $b_{sv}$ . Αναλυτικά:

$$b = \frac{1}{S} * \sum_{sv=1}^S b_{sv}, \text{ όπου } S \text{ το πλήθος των support vectors του batch.}$$

Μετά την εκπαίδευση του SVM του batch, πραγματοποιείται ο υπολογισμός των προβλέψεων του SVM τόσο για όλα τα test δείγματα, όσο και για τα ίδια τα training δείγματα του batch στο οποίο εκπαιδεύτηκε. Η διαδικασία υπολογισμού των προβλέψεων ξεκινάει με υπολογισμό ενός πίνακα  $W$  διαστάσεων (πλήθος δειγμάτων στα οποία γίνεται πρόβλεψη της κλάσης τους +1 ή -1) x (batch\_size), κάθε στοιχείο  $W[i, j]$  του οποίου υπολογίζεται μέσω της σχέσης  $W[i, j] = Y_{tr}[j] * Kernel(X_{pred}[i] * X_{tr}[j])$ , όπου  $X_{tr}[j]$  το  $j$ -οστο δείγμα εκπαίδευσης που ανήκει στο batch και  $Y_{tr}[j]$  η ετικέτα του (+1 ή -1),  $X_{pred}[i]$  το  $i$ -οστο δείγμα στο οποίο γίνεται πρόβλεψη κλάσης.

Κατόπιν, για το  $i$ -οστο δείγμα στο οποίο γίνεται πρόβλεψη κλάσης, υπολογίζεται η τιμή της πρόβλεψης μέσω της σχέσης  $Prediction[i] = W[i, :] * a + b$ , όπου  $a$  το διάνυσμα των πολλαπλασιαστών Lagrange και  $b$  το bias. Η τιμή αυτή είναι αλγεβρική (μπορεί να έχει θετικό ή και αρνητικό πρόσημο).

Αντί, λοιπόν, να υλοποιείται ένα SVM για κάθε κλάση του data set, υλοποιούνται 50000 / batch size, σε αριθμό, SVMs, καθένα από τα οποία εκπαιδεύεται σε ένα batch training δειγμάτων. Ο συνολικός αριθμός των SVMs που κατασκευάζονται είναι  $10 * 50000 / \text{batch size}$  και τον ρόλο του SVM της κάθε κλάσης αναλαμβάνουν τα 50000 / batch size, τα οποία εκπαιδεύονται για να διαχωρίζουν τα δείγματα της κλάσης αυτής από τα δείγματα όλων των υπολοίπων κλάσεων.

Όσον αφορά τις τελικές προβλέψεις για τα training δείγματα, κάθε SVM πραγματοποιεί προβλέψεις για το batch στο οποίο εκπαιδεύτηκε. Για κάθε training δείγμα ενός batch θα έχουμε, συμπερασματικά, 10 διαφορετικές προβλέψεις, μία από κάθε SVM που αντιστοιχεί σε μία κλάση και εκπαιδεύεται στο συγκεκριμένο batch. Ως τελική πρόβλεψη, επιλέγεται η πρόβλεψη με την μεγαλύτερη αλγεβρική τιμή και το training δείγμα κατηγοριοποιείται στην κλάση της οποίας το σύνολο των SVMs πραγματοποίησε την πρόβλεψη αυτή.

Όσον αφορά τις τελικές προβλέψεις για τα test δείγματα, καθένα από τα  $10 * 50000 / \text{batch size}$ , σε αριθμό, SVMs πραγματοποιεί προβλέψεις για όλα τα δείγματα αυτά. Για τα 50000 / batch size, σε αριθμό, SVMs που αντιστοιχούν σε μία κλάση, ως πρόβλεψη για κάθε test δείγμα ορίζεται ο μέσος όρος των προβλέψεων όλων των SVMs αυτών. Μόλις ολοκληρωθεί αυτή η διαδικασία για τα SVMs όλων των κλάσεων, έχουμε  $10 * (\text{πλήθος test δειγμάτων})$ , σε αριθμό, προβλέψεις. Σε κάθε test δείγμα αντιστοιχούν 10 προβλέψεις, καθεμία από τις οποίες είναι ο μέσος όρος των προβλέψεων όλων των SVMs κάθε κλάσης.

Όπως και στην περίπτωση των training δειγμάτων, ως τελική πρόβλεψη για κάθε test δείγμα επιλέγεται η πρόβλεψη με την μεγαλύτερη αλγεβρική τιμή και το test δείγμα κατηγοριοποιείται στην κλάση της οποίας το σύνολο των SVMs πραγματοποίησε την πρόβλεψη αυτή.

## Παραδείγματα ορθής και εσφαλμένης κατηγοριοποίησης

Ορθή κατηγοριοποίηση:

1)

```
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 0: -1.7188346317234795
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 1: -1.4244704523975928
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 2: -2.726542493935235
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 3: -2.2644360149411296
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 4: -1.0985485951433354
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 5: -2.874186102022167
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 6: -3.619504047825401
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 7: -2.086448279639563
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 8: 0.4525720865643016
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 9: -0.9139615831081509
Τελική πρόβλεψη για το δείγμα: Κλάση 8
Πραγματική κλάση του δείγματος: Κλάση 8
```

Παρατηρείται ότι όλες οι τιμές των προβλέψεων είναι αρνητικές, εκτός από την τιμή της πρόβλεψης που αντιστοιχεί στην κλάση 8. Αυτό υποδεικνύει ότι μόνο το σύνολο των SVMs της κλάσης 4 προέβλεψε ότι το δείγμα ανήκει στην κλάση αυτή.

2)

```
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 0: -1.4215648572105315
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 1: -1.287294842527171
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 2: -2.174169943992559
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 3: -1.7341243345712223
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 4: -0.8996678767550754
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 5: -1.1291112073449414
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 6: -0.6080773548390908
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 7: -0.9641571667976959
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 8: -1.407012007464571
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 9: -0.610314877482252
Τελική πρόβλεψη για το δείγμα: Κλάση 6
Πραγματική κλάση του δείγματος: Κλάση 6
```

Παρατηρείται ότι καμία τιμή πρόβλεψης δεν είναι θετική. Συμπεραίνεται, λοιπόν, ότι κανένα σύνολο SVMs κάποιας κλάσης δεν προέβλεψε ότι το δείγμα ανήκει στην κλάση αυτή. Ως τελική πρόβλεψη επιλέχθηκε η κλάση στην οποία αντιστοιχεί μεγαλύτερη, αλγεβρικά, τιμή πρόβλεψης, η οποία αποδείχθηκε και σωστή.

3)

```
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 0:-0.608747153044706
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 1:-1.179071488468013
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 2:-1.7608443392070032
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 3:-0.38366375279172527
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 4:-0.15179119946793995
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 5:-1.0329409043810454
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 6:-0.8969153693205609
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 7:-0.9650620328850981
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 8:-0.6672718059882098
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 9:-0.867791853108852
Τελική πρόβλεψη για το δείγμα: Κλάση 4
Πραγματική κλάση του δείγματος: Κλάση 4
```

Το ίδιο φαινόμενο παρατηρείται και σε αυτό το παράδειγμα. Όπως και παραπάνω, καμία τιμή πρόβλεψης δεν είναι θετική, παρόλο που το δίκτυο κατηγοριοποιεί το δείγμα σωστά.

Εσφαλμένη κατηγοριοποίηση:

1)

```
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 0:-0.87232976913363
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 1:-0.19221428376704364
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 2:-0.5773641224834517
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 3:-0.2362841986798933
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 4:0.13846378939353196
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 5:-0.8010692736072784
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 6:-0.5623595701138833
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 7:-0.5237762938579443
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 8:-0.7405067663067284
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 9:-0.428235542244696
Τελική πρόβλεψη για το δείγμα: Κλάση 4
Πραγματική κλάση του δείγματος: Κλάση 3
```

Παρατηρείται ότι όλες οι τιμές των προβλέψεων είναι αρνητικές, εκτός από την τιμή της πρόβλεψης που αντιστοιχεί στην κλάση 4. Αυτό υποδεικνύει ότι μόνο το σύνολο των SVMs της κλάσης 4 προέβλεψε ότι το δείγμα ανήκει στην κλάση αυτή. Επιπλέον, όλες οι τιμές βρίσκονται κοντά στο 0, γεγονός που δείχνει ότι το δίκτυο “δυσκολεύεται” να δώσει πρόβλεψη για μία συγκεκριμένη κλάση αποκλείοντας τις υπόλοιπες.

2)

```
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 0:-4.384384867356135
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 1:-4.675877363337898
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 2:-6.564366027825276
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 3:-6.710313666847426
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 4:-0.025070562474652572
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 5:-4.148118432248548
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 6:-5.3133716069265455
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 7:-2.9630138275803626
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 8:-5.010444195158241
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 9:0.026345082337972142
Τελική πρόβλεψη για το δείγμα: Κλάση 9
Πραγματική κλάση του δείγματος: Κλάση 8
```

Παρατηρείται ότι οι τιμές των προβλέψεων για την πλειονότητα των κλάσεων είναι αρνητικές και αρκετά μικρότερες από το μηδέν. Αυτό υποδεικνύει ότι το δίκτυο προβλέπει ότι η πιθανότητα το δείγμα να ανήκει σε κάποια από αυτές τις κλάσεις είναι μικρή. Αντίθετα, οι τιμές των προβλέψεων για τις κλάσεις 4 και 9 είναι πολύ κοντά στο 0, με την δεύτερη να είναι ελαφρώς θετική. Ως συμπέρασμα, το δίκτυο προβλέπει ότι το δείγμα ανήκει σε μία από αυτές τις δύο κλάσεις, “χωρίς να δίνει την τελική του απόφαση με μεγάλη βεβαιότητα”.

3)

```
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 0: -2.7068097531513367
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 1: -1.1473066705678734
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 2: -2.037438885031576
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 3: -2.76785305912706
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 4: -0.38593941599095816
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 5: -1.9262396371744719
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 6: -3.1107696736276416
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 7: -2.051387991920307
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 8: -1.7532533991824326
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 9: -0.5844839900704669
Τελική πρόβλεψη για το δείγμα: Κλάση 4
Πραγματική κλάση του δείγματος: Κλάση 8
```

Σε αυτό το παράδειγμα, όπως παρατηρήθηκε και σε ορισμένα παραδείγματα ορθής κατηγοριοποίησης, καμία τιμή πρόβλεψης δεν είναι θετική. Διαπιστώνεται, άρα, ότι κανένα σύνολο SVMs κάποιας κλάσης δεν προέβλεψε ότι το δείγμα ανήκει στην κλάση αυτή. Ως τελική πρόβλεψη επιλέχθηκε η κλάση της οποίας η τιμές πρόβλεψης ήταν αλγεβρικά η μεγαλύτερη.

## Πειράματα:

Πραγματοποιήθηκαν διάφορα πειράματα με διαφορετικές τιμές των παραμέτρων σε μικρότερα τμήματα του data set για να μειωθεί ο συνολικός χρόνος εκπαίδευσης και να δημιουργηθεί μία γενική εικόνα για την συμπεριφορά του αλγορίθμου ανάλογα τις τιμές αυτές. Με βάση τις παρατηρήσεις στα πειράματα αυτά, επιλέχθηκαν ορισμένα για διεξαγωγή σε όλο το data set.

Παρακάτω δίνονται τα αποτελέσματα πειραμάτων με διαφορετικούς πυρήνες (Kernel συναρτήσεις), διαφορετική τιμή της παραμέτρου C και διαφορετικό batch size για εκπαίδευση σε όλο το data set.

Πείραμα 1: Γραμμικός πυρήνας, C = 1, Batch size = 500

Ποσοστό επιτυχίας εκπαίδευσης (training accuracy)	Ποσοστό επιτυχίας test (test accuracy)	Συνολικός χρόνος εκπαίδευσης (σε λεπτά)
48.49%	38.99%	240.93

Τα αποτελέσματα αυτά είναι λογικά, δεδομένου ότι το Cifar-10 είναι ένα πολύπλοκο data set, με αποτέλεσμα ο γραμμικός πυρήνας να αδυνατεί να πετύχει ιδιαίτερα υψηλά ποσοστό επιτυχίας τόσο στο training όσο και στο test.

Πείραμα 2: Πολυωνυμικός πυρήνας 1<sup>ου</sup> βαθμού, C = 2, Batch size = 1000

Ποσοστό επιτυχίας εκπαίδευσης (training accuracy)	Ποσοστό επιτυχίας test (test accuracy)	Συνολικός χρόνος εκπαίδευσης (σε λεπτά)
38.07%	35.2%	293.69

Παρά την αύξηση του βαθμού του πολυωνυμικού πυρήνα και την αύξηση του batch size (ώστε κάθε SVM να εκπαιδεύεται σε περισσότερα training δείγματα), η απόδοση μειώθηκε σε σχέση με το προηγούμενο πείραμα, ενδεχομένως λόγω της αύξησης του C.

Πείραμα 3: Πολυωνυμικός πυρήνας 2<sup>ου</sup> βαθμού, C = 0.1, Batch size = 500

Ποσοστό επιτυχίας εκπαίδευσης (training accuracy)	Ποσοστό επιτυχίας test (test accuracy)	Συνολικός χρόνος εκπαίδευσης (σε λεπτά)
98.67%	42.47%	254.88

Παρατηρείται ότι η χρήση πολυωνυμικού Kernel 2<sup>ου</sup> βαθμού οδήγησε στην ραγδαία αύξηση του training accuracy, ενώ παράλληλα αυξήθηκε το test accuracy. Ως συμπέρασμα, ο πολυωνυμικός Kernel 2<sup>ου</sup> βαθμού δείχνει να είναι καταλληλότερος για το πρόβλημα σε σχέση με τους προηγούμενους δύο.

Πείραμα 4: Πολυωνυμικός πυρήνας 2<sup>ου</sup> βαθμού, C = 0.1, Batch size = 1000

Ποσοστό επιτυχίας εκπαίδευσης (training accuracy)	Ποσοστό επιτυχίας test (test accuracy)	Συνολικός χρόνος εκπαίδευσης (σε λεπτά)
99.15%	46.61%	318.57

Το πείραμα αυτό είναι παρόμοιο με το προηγούμενο, με μόνη διαφορά την αύξηση του batch size. Είναι εμφανές από τα αποτελέσματα ότι αυτή η τροποποίηση οδήγησε σε αρκετά υψηλότερο test accuracy, ενώ το training accuracy παρέμεινε στα ίδια υψηλά επίπεδα τιμής.

Πείραμα 5: Πολυωνυμικός πυρήνας 3<sup>ου</sup> βαθμού, C = 0.1, Batch size = 500

Ποσοστό επιτυχίας εκπαίδευσης (training accuracy)	Ποσοστό επιτυχίας test (test accuracy)	Συνολικός χρόνος εκπαίδευσης (σε λεπτά)
96.77%	38.4%	264.64

Παρατηρούμε ότι η αύξηση του βαθμού του πολυωνυμικού Kernel σε 3 δεν οδήγησε σε περαιτέρω βελτίωση της απόδοσης του δικτύου.

Πείραμα 6: Πολυωνυμικός πυρήνας 3<sup>ου</sup> βαθμού, C = 0.1, Batch size = 1000

Ποσοστό επιτυχίας εκπαίδευσης (training accuracy)	Ποσοστό επιτυχίας test (test accuracy)	Συνολικός χρόνος εκπαίδευσης (σε λεπτά)
100%	41.37%	418.16

Τα συμπεράσματα αυτού του πειράματος συμφωνούν με τα αντίστοιχα του προηγούμενου. Για τις ίδιες τιμές C και batch size, ο πολυωνυμικός πυρήνας 3<sup>ου</sup> βαθμού δεν είναι αποδοτικότερος από πολυωνυμικό πυρήνα 2<sup>ου</sup> βαθμού. Συμπεραίνεται, λοιπόν, ότι ο τελευταίος αποτελεί, πιθανότατα, τον πλέον κατάλληλο πολυωνυμικό Kernel για τις ανάγκες αυτού του data set.

Πείραμα 7: RBF πυρήνας, C = 1, Batch size = 1000

Ποσοστό επιτυχίας εκπαίδευσης (training accuracy)	Ποσοστό επιτυχίας test (test accuracy)	Συνολικός χρόνος εκπαίδευσης (σε λεπτά)
95.72%	44.53%	590.72

Ο RBF πυρήνας αποδείχθηκε ιδιαίτερα αποδοτικός, πλησιάζοντας την τιμή του test accuracy του πολυωνυμικού Kernel 2<sup>ου</sup> βαθμού, για ίδιο batch size άλλα διαφορετική τιμή C. Ωστόσο, ο χρόνος εκπαίδευσης του πρώτου ήταν ιδιαίτερα αυξημένος σε σχέση με τον αντίστοιχο του δευτέρου.



Πείραμα 8: Πολυωνυμικός πυρήνας 2<sup>ου</sup> βαθμού, C = 0.1, Batch size = 2500

Ποσοστό επιτυχίας εκπαίδευσης (training accuracy)	Ποσοστό επιτυχίας test (test accuracy)	Συνολικός χρόνος εκπαίδευσης (σε λεπτά)
100%	48.81%	1093.18

Δεδομένου ότι το καλύτερο test accuracy σημειώθηκε για πολυωνυμικό Kernel 2<sup>ου</sup> βαθμού, πραγματοποιήθηκε το ίδιο πείραμα με batch size = 2500. Το ποσοστό επιτυχίας στο test και στο training αυξήθηκαν με αυτήν την αλλαγή. Παρ' όλα αυτά, ο χρόνος εκπαίδευσης των 1093 λεπτών (18.21 ώρες) αποτελεί αποτρεπτικό παράγοντα όσον αφορά την διεξαγωγή και άλλων πειραμάτων με τόσο μεγάλο batch size.

**Τροποποιημένος κώδικας:** Παρά την επίτευξη ικανοποιητικών αποτελεσμάτων, ο αλγόριθμος που ακολουθεί ο κώδικας που περιγράφηκε παραπάνω παρουσιάζει μία κύρια αδυναμία. Κατά την φάση της εκπαίδευσης, κάθε batch training δειγμάτων περιλαμβάνει δείγματα της κλάσης +1 σε ποσοστό περίπου 10%. Αυτό το ποσοστό είναι ιδιαίτερα χαμηλό, με αποτέλεσμα το SVM που εκπαιδεύεται σε αυτό το batch να τείνει να δίνει πολλές περισσότερες αρνητικές προβλέψεις κατά την φάση του υπολογισμού των τιμών των προβλέψεων. Αυτό γίνεται εμφανές και από τα παραδείγματα κατηγοριοποίησης τα οποία παρατέθηκαν προηγουμένως. Σε πολλά δείγματα, μάλιστα, οι 10 προβλέψεις που υπολογίστηκαν, μία για κάθε κλάση, ήταν όλες αρνητικές. Ως ερμηνεία, κανένα σύνολο SVMs που αντιστοιχούν σε μία κλάση δεν προέβλεπε ότι το δείγμα ανήκει στην κλάση αυτή. Η τελική πρόβλεψη, που ήταν η μέγιστη από όλες τις προβλέψεις, ήταν και αυτή αρνητική.

Για να αντιμετωπιστεί αυτό το πρόβλημα, υλοποιήθηκε ένας νέος τροποποιημένος κώδικας. Σύμφωνα με αυτόν, πριν την κατασκευή των SVMs μίας συγκεκριμένης κλάσης, ο πίνακας data\_tr, ο οποίος έχει ως γραμμές τα training δείγματα, χωρίζεται σε δύο υποπίνακες. Ο ένας υποπίνακας περιέχει όλα τα δείγματα που ανήκουν στην κλάση αυτή ενώ ο άλλος υποπίνακας περιέχει τα υπόλοιπα δείγματα. Κατόπιν, κάθε batch δημιουργείται από τυχαία επιλογή batch size / 2, σε αριθμό, δειγμάτων του πρώτου υποπίνακα και batch size / 2, σε αριθμό, δειγμάτων του δεύτερου υποπίνακα. Με αυτόν τον τρόπο, το batch είναι ισορροπημένο, περιέχοντας ίσο αριθμό δειγμάτων με ετικέτα +1 και δειγμάτων με ετικέτα -1.

Η αλλαγή αυτή οδήγησε και σε διαφορετικό τρόπο υπολογισμού του training accuracy, καθώς κάθε σύνολο SVMs μίας κλάσης, λόγω της τυχαιότητας στην δημιουργία των batches, δεν εκπαιδεύεται στα ίδια batches training δειγμάτων. Επιλέχθηκε, λοιπόν, ο υπολογισμός του training accuracy να πραγματοποιείται μέσω ενός validation test. Συγκεκριμένα, δημιουργείται ένα batch 1000 δειγμάτων εκπαίδευσης, κάποια από τα οποία έχουν, ενδεχομένως (λόγω της τυχαιότητας), χρησιμοποιηθεί κατά την εκπαίδευση ενώ κάποια άλλα όχι. Η διαδικασία που ακολουθεί είναι ανάλογη με τον τρόπο υπολογισμού των προβλέψεων που αναλύθηκε παραπάνω για τον πρώτο κώδικα. Στο τέλος της εκπαίδευσης κάθε SVM, υπολογίζονται οι τιμές των προβλέψεων του για τα δείγματα του validation set. Για το σύνολο των SVMs που αντιστοιχούν σε μία κλάση, ως πρόβλεψη για κάθε τέτοιο δείγμα ορίζεται ο μέσος όρος των προβλέψεων όλων των SVMs αυτών. Μόλις ολοκληρωθεί αυτή η διαδικασία για τα SVMs όλων των κλάσεων, έχουμε 10 \* (πλήθος δειγμάτων του validation set), σε αριθμό, προβλέψεις. Σε κάθε δείγμα αντιστοιχούν 10 προβλέψεις, καθεμία από τις οποίες είναι ο μέσος όρος των προβλέψεων όλων των SVMs κάθε κλάσης. Ως τελική πρόβλεψη για κάθε validation δείγμα, επιλέγεται η πρόβλεψη με την μεγαλύτερη αλγεβρική τιμή και το validation δείγμα κατηγοριοποιείται στην κλάση της οποίας το σύνολο των SVMs πραγματοποίησε την πρόβλεψη αυτή.

Το μειονέκτημα αυτού του κώδικα είναι ότι, σε αντίθεση με τον πρώτο κώδικα, δεν χρησιμοποιούνται όλα τα δείγματα του training set κατά την διάρκεια της εκπαίδευσης. Λόγω της τυχαιότητας κατά την

δημιουργία των batches, είναι βέβαιο ότι ορισμένα training δείγματα δεν θα βρεθούν σε κανένα από τα batches, ώστε το συνολικό δίκτυο να εκπαιδευτεί σε αυτό.

## Παραδείγματα ορθής και εσφαλμένης κατηγοριοποίησης

Ορθή κατηγοριοποίηση:

1)

```
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 0:0.5715025830603944
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 1:0.42949388959624873
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 2:-0.30307686154780744
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 3:-0.41180491400859465
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 4:-2.0696190295912755
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 5:-0.5142618367482619
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 6:-1.3073200907576354
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 7:-1.320722844814095
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 8:2.8279507314107937
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 9:0.26694872726163393
Τελική πρόβλεψη για το δείγμα: Κλάση 8
Πραγματική κλάση του δείγματος: Κλάση 8
```

Παρατηρείται ότι, σε αντίθεση με τα αντίστοιχα παραδείγματα κατηγοριοποίησης του πρώτου κώδικα, τέσσερις τιμές πρόβλεψης είναι θετικές. Αυτό υποδεικνύει ότι υπήρξαν τέσσερα σύνολα SVMs που προέβλεψαν ότι το δείγμα ανήκει στην κλάση τους. Ως τελική πρόβλεψη επιλέχθηκε η κλάση 8 στην οποία αντιστοιχεί η μεγαλύτερη τιμή πρόβλεψης.

2)

```
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 0:-1.2383545147370532
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 1:-0.7442702523132347
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 2:-0.6697064551747325
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 3:-0.20934490508792036
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 4:-0.8951539205984985
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 5:-0.31608591825492804
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 6:2.811287360946135
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 7:-0.34218103917444787
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 8:-2.7243489289547993
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 9:1.254202052915679
Τελική πρόβλεψη για το δείγμα: Κλάση 6
Πραγματική κλάση του δείγματος: Κλάση 6
```

Στο παράδειγμα αυτό, μόνο δύο σύνολα SVMs προέβλεψαν ότι το δείγμα ανήκει στην κλάση τους. Η τιμή που αντιστοιχεί στην κλάση 6 είναι αρκετά μεγαλύτερη από αυτή της κλάσης 9, με αποτέλεσμα το δείγμα να κατηγοριοποιείται, ορθά, στην πρώτη.



3)

```
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 0:0.9561709022275009
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 1:-1.138474297138344
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 2:-0.594957173842338
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 3:0.15670125392097667
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 4:-1.48918138686152
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 5:-0.16672049719367332
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 6:-1.728534761709835
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 7:-2.9163919122654143
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 8:0.6167092357788234
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 9:-1.2645559187420699
Τελική πρόβλεψη για το δείγμα: Κλάση 0
Πραγματική κλάση του δείγματος: Κλάση 0
```

Οι τιμές των προβλέψεων που αντιστοιχούν στις κλάσεις 0, 3 και 8 ήταν οι μόνες θετικές. Μεταξύ των κλάσεων 0 και 8 οι τιμές είναι αρκετά κοντά. Αυτό σημαίνει ότι το δίκτυο δεν κατέληξε “με μεγάλη βεβαιότητα” στην σωστή κλάση 0.

Εσφαλμένη κατηγοριοποίηση:

1)

```
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 0:-0.18278936806036142
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 1:1.5451218954865922
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 2:1.5750006557493819
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 3:1.8526785169126883
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 4:-1.8362124551449268
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 5:0.4032593204823691
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 6:3.487766541084475
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 7:-1.4409275185247727
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 8:-0.0910064088402763
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 9:-1.0020003727793576
Τελική πρόβλεψη για το δείγμα: Κλάση 6
Πραγματική κλάση του δείγματος: Κλάση 3
```

Παρατηρείται ότι υπάρχουν πολλές θετικές τιμές προβλέψεων, με αυτήν για την κλάση 6 να είναι με διαφορά η μεγαλύτερη. Η τιμή πρόβλεψης για την κλάση 3, που είναι και η σωστή, είναι θετική και, μάλιστα, η δεύτερη μεγαλύτερη, γεγονός που υποδεικνύει ότι το δίκτυο βρέθηκε αρκετά κοντά στην σωστή πρόβλεψη.

2)

```
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 0:-0.2703825092373924
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 1:0.7249107510565771
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 2:0.8849144637739463
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 3:0.23581624659926037
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 4:-0.6923277802656244
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 5:-0.67623360401791
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 6:-3.256535539471241
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 7:0.10499632430521622
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 8:1.6223223759979526
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 9:-0.42282870424343716
Τελική πρόβλεψη για το δείγμα: Κλάση 8
Πραγματική κλάση του δείγματος: Κλάση 0
```

Σε αυτό το παράδειγμα, η τιμή πρόβλεψης που αντιστοιχεί στην σωστή κλάση 0 είναι αρνητική. Ως συμπέρασμα, το σύνολο SVMs της κλάσης 0 δεν προέβλεψε ότι το δείγμα ανήκει στην κλάση αυτή.

3)

```
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 0:-1.0775839932204994
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 1:0.3305803831879081
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 2:-0.03667579041141458
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 3:4.280433800644776
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 4:-4.54835869047795
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 5:2.1357282308776457
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 6:3.659000948820061
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 7:-0.762554899835073
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 8:-2.447127068256578
Πρόβλεψη του συνόλου SVMs που αντιστοιχούν στην κλάση 9:-0.6906272470843622
Τελική πρόβλεψη για το δείγμα: Κλάση 3
Πραγματική κλάση του δείγματος: Κλάση 1
```

Σε αντίθεση με το 1<sup>ο</sup> παράδειγμα εσφαλμένης κατηγοριοποίησης, εδώ παρατηρείται ότι, αν και η τιμή πρόβλεψης για την σωστή κλάσης 1 είναι θετική, υπάρχουν πολλές κλάσεις των οποίων οι τιμές πρόβλεψης είναι θετικές και αρκετά μεγαλύτερες από την αντίστοιχη της κλάσης 1. Διαπιστώνεται, άρα, ότι το δίκτυο δεν βρέθηκε κοντά στην πραγματοποίηση σωστής τελικής πρόβλεψης.

## Πειράματα:

Όπως και παραπάνω, πραγματοποιήθηκαν διάφορα πειράματα με διαφορετικές τιμές των παραμέτρων σε μικρότερα τμήματα του data set για να μειωθεί ο συνολικός χρόνος εκπαίδευσης και να δημιουργηθεί μία γενική εικόνα για την συμπεριφορά του αλγορίθμου ανάλογα τις τιμές αυτές. Με βάση τις παρατηρήσεις στα πειράματα αυτά, επιλέχθηκαν ορισμένα για διεξαγωγή σε όλο το data set.

Παρακάτω δίνονται τα αποτελέσματα πειραμάτων με διαφορετικούς πυρήνες (Kernel συναρτήσεις), διαφορετική τιμή της παραμέτρου C και batch size = 1000(το οποίο αναδείχθηκε το καταλληλότερο όσον αφορά το test accuracy σε σύγκριση με τον συνολικό χρόνο εκπαίδευσης) για εκπαίδευση σε όλο το data set.

Πείραμα 1: Γραμμικός πυρήνας, C = 0.1, Batch size = 1000

Ποσοστό επιτυχίας εκπαίδευσης (training accuracy)	Ποσοστό επιτυχίας test (test accuracy)	Συνολικός χρόνος εκπαίδευσης (σε λεπτά)
38.1%	37.92%	203.81

Όπως και στο αντίστοιχο πυρήνα του πρώτου κώδικα, τα αποτελέσματα για γραμμικό πυρήνα δεν είναι ικανοποιητικά.

Πείραμα 2: Πολυωνυμικός πυρήνας 1<sup>ου</sup> βαθμού, C = 0.1, Batch size = 1000

Ποσοστό επιτυχίας εκπαίδευσης (training accuracy)	Ποσοστό επιτυχίας test (test accuracy)	Συνολικός χρόνος εκπαίδευσης (σε λεπτά)
40%	38.51%	230.35

Η αλλαγή σε πολυωνυμικό πυρήνα 1<sup>ου</sup> βαθμού επιφέρει μικρή αύξηση στα ποσοστά επιτυχίας του training και του test, με παράλληλη αύξηση του χρόνου εκτέλεσης. Αυτό είναι λογικό, δεδομένου ότι η χρήση πολυωνυμικού πυρήνα μεγαλύτερου βαθμού συμβάλλει στην δυνατότητα του δικτύου να “μαθαίνει” πιο πολύπλοκα data sets.

Πείραμα 3: Πολυωνυμικός πυρήνας 2<sup>ου</sup> βαθμού,  $C = 0.1$ , Batch size = 1000

Ποσοστό επιτυχίας εκπαίδευσης (training accuracy)	Ποσοστό επιτυχίας test (test accuracy)	Συνολικός χρόνος εκπαίδευσης (σε λεπτά)
51.3%	48.3%	296.42

Ο πολυωνυμικός πυρήνας 2<sup>ου</sup> βαθμού, όπως και στα αντίστοιχα πειράματα του πρώτου κώδικα, παρουσιάζει τα καλύτερα αποτελέσματα. Σε σύγκριση των πειραμάτων για αυτόν τον πυρήνα ( $C = 0.1$  και batch size = 1000), παρατηρείται ότι η υλοποίηση του SVM με τον δεύτερο κώδικα οδηγεί σε 2% μεγαλύτερο ποσοστό επιτυχίας στο test. Συμπερασματικά, τα ισορροπημένα batches, πιθανόν, βοηθούν στην καλύτερη εκπαίδευση, με αποτέλεσμα την βελτίωση της απόδοσης.

Πείραμα 4: Πολυωνυμικός πυρήνας 2<sup>ου</sup> βαθμού,  $C = 10$ , Batch size = 1000

Ποσοστό επιτυχίας εκπαίδευσης (training accuracy)	Ποσοστό επιτυχίας test (test accuracy)	Συνολικός χρόνος εκπαίδευσης (σε λεπτά)
52.6%	48,45%	288.48

Το πείραμα αυτό είναι παρόμοιο με το προηγούμενο, με μόνη διαφοροποίηση την αλλαγή της τιμής της παραμέτρου  $C$  σε 10. Όπως φαίνεται από τα αποτελέσματα, το training accuracy αυξήθηκε κατά 1.3%, ενώ το test accuracy μόνο κατά 0.15%. Οι παρατηρήσεις αυτές είναι λογικές, καθώς η αύξηση της τιμής του  $C$  οδηγεί σε επιδίωξη του δικτύου να επιτρέψει λιγότερα λάθη κατά την φάση της εκπαίδευσης (μειώνοντας το περιθώριο μεταξύ της διαχωριστικής υπερεπιφάνειας και των δειγμάτων) αυξάνοντας έτσι το training accuracy. Παράλληλα, το test accuracy ενδέχεται να αυξηθεί, να μειωθεί ή να μείνει σταθερό, καθώς η βελτίωση του training accuracy είναι πιθανό να συμβάλει στην αντίστοιχη βελτίωση του test accuracy ή να οδηγήσει σε overfitting, με αποτέλεσμα το ποσοστό επιτυχίας στο test να αρχίσει να μειώνεται. Στην περίπτωση του πειράματος αυτού, αποδεικνύεται ότι η τιμή  $C = 10$  δεν οδήγησε σε overfitting.

Πείραμα 5: Πολυωνυμικός πυρήνας 3<sup>ου</sup> βαθμού,  $C = 0.1$ , Batch size = 1000

Ποσοστό επιτυχίας εκπαίδευσης (training accuracy)	Ποσοστό επιτυχίας test (test accuracy)	Συνολικός χρόνος εκπαίδευσης (σε λεπτά)
46.2%	40.97%	260.05

Το συμπέρασμα του πειράματος αυτού συμβαδίζει με το αντίστοιχο του πρώτου κώδικα. Ο πολυωνυμικός πυρήνας 3<sup>ου</sup> βαθμού αποδεικνύεται λιγότερο αποδοτικός από τον πολυωνυμικό πυρήνα 2<sup>ου</sup> βαθμού.

Πείραμα 6: RBF πυρήνας,  $C = 1$ , Batch size = 1000

Ποσοστό επιτυχίας εκπαίδευσης (training accuracy)	Ποσοστό επιτυχίας test (test accuracy)	Συνολικός χρόνος εκπαίδευσης (σε λεπτά)
49.3%	45.49%	770.57

Ο RBF πυρήνας δείχνει να οδηγεί σε καλύτερη απόδοση από τον πολυωνυμικό 3<sup>ου</sup> βαθμού, αλλά χειρότερη από τον αντίστοιχο 2<sup>ου</sup> βαθμού. Ο συνολικός χρόνος εκπαίδευσης, ωστόσο, είναι πολύ υψηλότερος από τους αντίστοιχους των άλλων δύο πυρήνων, με αποτέλεσμα να καθίσταται ο λιγότερος αποδοτικός (αναλογία test accuracy προς χρόνο εκπαίδευσης).

Παρακάτω παρατίθεται ένας πίνακας σύγκρισης των χαρακτηριστικών των δύο κωδίκων που περιγράφηκαν και των οποίων τα αποτελέσματα αναλύθηκαν παραπάνω:

Χαρακτηριστικά	Αρχικός κώδικας (πρώτος)	Τροποποιημένος κώδικας (δεύτερος)
Δημιουργία batches	Μη ισορροπημένα	Ισορροπημένα
Training accuracy	Υπολογισμός σε όλα τα δεδομένα	Validation set
Πλεονεκτήματα	Χρήση όλων των δεδομένων	Καλύτερη ισορροπία προβλέψεων
Μειονεκτήματα	Αρνητικές προβλέψεις	Χρήση υποσυνόλου δεδομένων εκπαίδευσης

### Σύγκριση απόδοσης με άλλους αλγορίθμους κατηγοριοποίησης:

Η καλύτερη απόδοση του SVM σημειώθηκε στο πείραμα με πολυωνυμικό Kernel 2<sup>ου</sup> βαθμού, batch size = 2500, C = 0.1, στο οποίο το ποσοστό επιτυχίας στο test υπολογίστηκε 48,81%.

Όσον αφορά το MLP με ένα κρυφό επίπεδο που χρησιμοποιεί hinge loss function για την βελτιστοποίηση, το μεγαλύτερο ποσοστό επιτυχίας στο test ήταν 49.26% για 512 νευρώνες στο κρυφό επίπεδο και 300 εποχές.

Τα αντίστοιχα ποσοστά για τους κατηγοριοποιητές 1 και 3 γειτόνων καθώς και για τον κατηγοριοποιητή πλησιέστερου κέντρου κλάσης αντλήθηκαν από τα αποτελέσματα της Ενδιάμεσης Εργασίας.

Αναλυτικά:

	Ποσοστό επιτυχίας στο test	Χρόνος εκτέλεσης
SVM	48.81%	1093.18 λεπτά
MLP	49.26%	48.24 λεπτά
1-NN	35.39%	85 λεπτά
3-NN	35.61%	88 λεπτά
Nearest Centroid	27.74%	2.04 δευτερόλεπτα

Όπως είναι εμφανές από τα στοιχεία του παραπάνω πίνακα, το SVM δεν αποδεικνύεται ο καταλληλότερος αλγόριθμος κατηγοριοποίησης με βάση το ποσοστό επιτυχίας στο test. Συγκεκριμένα, το test accuracy του SVM είναι πολύ υψηλότερο από τα αντίστοιχα των αλγορίθμων 1-NN, 3-NN και Nearest Centroid, το οποίο είναι αναμενόμενο δεδομένου της αρκετά πολυπλοκότερης υλοποίησης του πρώτου σε σχέση με τους υπόλοιπους. Ωστόσο, όσον αφορά την σύγκριση SVM με MLP, το τελευταίο επιτυγχάνει λίγο υψηλότερο ποσοστό.

Παρά την σχετικά ικανοποιητική επίδοση του SVM αναφορικά με το test accuracy, αποτελεί με μεγάλη διαφορά την πιο χρονοβόρα επιλογή. Αξίζει να σημειωθεί ότι παρά την εξίσου μεγάλη πολυπλοκότητα του MLP και την επίτευξη test accuracy στα ίδια επίπεδα με το SVM (και ελαφρώς καλύτερη), ο χρόνος εκτέλεσης που απαιτεί είναι ο μικρότερος από όλους, εκτός από την περίπτωση του αλγορίθμου Nearest Centroid.