



ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΣΚΗΣΗΣ MLP ΔΙΚΤΥΟΥ

ΜΕΛΗ ΤΗΣ ΟΜΑΔΑΣ

Αγνή Παΐλα, ΑΜ: 4753
Νικολέτα Μπεράτη, ΑΜ: 4884

Μάθημα: Υπολογιστική Νοημοσύνη
Έτος: 2023-2024

Στην παρούσα άσκηση υλοποιούμε ένα πρόγραμμα ταξινόμησης βασισμένο στο πολυεπίπεδο perceptron με 3 κρυμμένα επίπεδα.

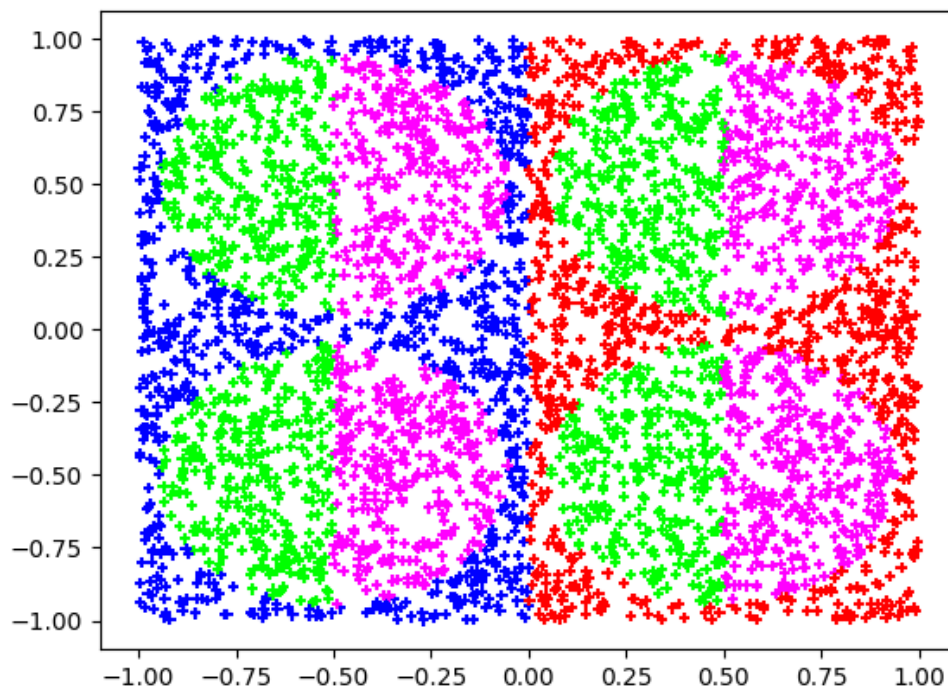
Το πρόγραμμα ταξινόμησης που κατασκευάσαμε δομείται από 4 κλάσεις:

- Την κλάση **MLP** η οποία περιέχει την main συνάρτηση του προγράμματος και υλοποιεί τους βασικούς αλγορίθμους, όπως είναι η διαδικασία της εκπαίδευσης και του υπολογισμού της ικανότητας γενίκευσης.
- Την κλάση **Layer** η οποία αναπαριστά τα επίπεδα του MLP δικτύου και υλοποιεί τις συναρτήσεις που αφορούν τα επίπεδα, όπως είναι οι forwardPass και backPropagation.
- Την κλάση **Neuron** η οποία αναπαριστά του νευρώνες του MLP δικτύου και υλοποιεί τους βασικούς υπολογισμούς που αφορούν τους νευρώνες, όπως είναι ο υπολογισμός της εξόδου ενός νευρώνα, ο υπολογισμός του δέλτα, τα βάρη κτλ.
- Την κλάση Point η οποία αναπαριστά ένα παράδειγμα της μορφής [x1, x2, category], όπως το εξάγουμε από το σύνολο εκπαίδευσης και το σύνολο ελέγχου (training_set, test_set).

Για την μεταγλώττιση και την εκτέλεση του προγράμματος, είναι απαραίτητο να εκτελεστούν οι παρακάτω εντολές:

- **javac *.java**
- **java MLP**

Τα αρχεία που περιέχουν τα δεδομένα μας για την εκπαίδευση και τον έλεγχο του δικτύου είναι σε μορφή .csv (training_set.csv, test_set.csv) και δημιουργήθηκαν μέσα από μια java κλάση. Προκειμένου να τα οπτικοποιήσουμε, δημιουργήσαμε ένα αρχείο Python, του οποίου η εκτέλεση απεικονίζεται ως εξής:



Ο αριθμός των εισόδων (d), το πλήθος των κατηγοριών (K), ο αριθμός νευρώνων στο πρώτο κρυμμένο επίπεδο ($h1$), ο αριθμός νευρώνων στο δεύτερο κρυμμένο επίπεδο ($h2$), ο αριθμός νευρώνων στο τρίτο κρυμμένο ($h3$) και το είδος συνάρτησης ενεργοποίησης (λογιστική, \tanh ή relu) για τα κρυμμένα επίπεδα και το επίπεδο εξόδου καθορίζονται στην αρχή της κλάσης MLP. Επιπλέον, στο σημείο εκείνο καθορίζεται και ο ρυθμός μάθησης (learningRate), ο αριθμός των batches (numBatches) και το κατώφλι τερματισμού ($\text{terminateThreshold}$).

ΑΛΓΟΡΙΘΜΟΣ ΕΚΠΑΙΔΕΥΣΗΣ:

Ο αλγόριθμος **train** που υλοποιήσαμε, διατρέχει όλα τα παραδείγματα του συνόλου εκπαίδευσης, είτε σειριακά είτε σε ομάδες ανάλογα με τον αριθμό του numBatches . Αποθηκεύει, λοιπόν, για κάθε παράδειγμα, τόσο το σημείο ($x1, x2$) όσο και την κατηγορία στην οποία ανήκει.

Εκτελεί, έπειτα, τον αλγόριθμο *forwardPass*, κάνει του απαραίτητους υπολογισμούς και παράγει το αποτέλεσμα στο επίπεδο εξόδου.

Επόμενο βήμα είναι η εκτέλεση του αλγορίθμου *backpropagation*, ο οποίος κάνει τους απαραίτητους υπολογισμούς ώστε να πραγματοποιηθεί η οπισθοδιάδοση του σφάλματος.

Ύστερα, υπολογίζεται το συνολικό σφάλμα, ελέγχοντας την διαφορά της παραγόμενης τιμής εξόδου με την αναμενόμενη.

Στο τέλος εκτέλεσης των παραπάνω διαδικασιών, ενημερώνονται τα βάρη των νευρώνων με σκοπό να επιτευχθεί η εκπαίδευση, χρησιμοποιώντας την μέθοδο *gradient descent*.

Η εκπαίδευση τελειώνει όταν το συνολικό σφάλμα μίας εποχής έχει διαφορά με το συνολικό σφάλμα της προηγούμενης εποχής ένα κατώτατο κατώφλι, το οποίο έχουμε ορίσει στην μεταβλητή $\text{terminateThreshold}$, αρκεί να έχουν περάσει 700 εποχές.

ΠΕΙΡΑΜΑΤΑ

Υλοποιήσαμε πειράματα και καταγράψαμε σε πίνακες πως μεταβάλλεται το ποσοστό επιτυχίας στο σύνολο ελέγχου θεωρώντας:

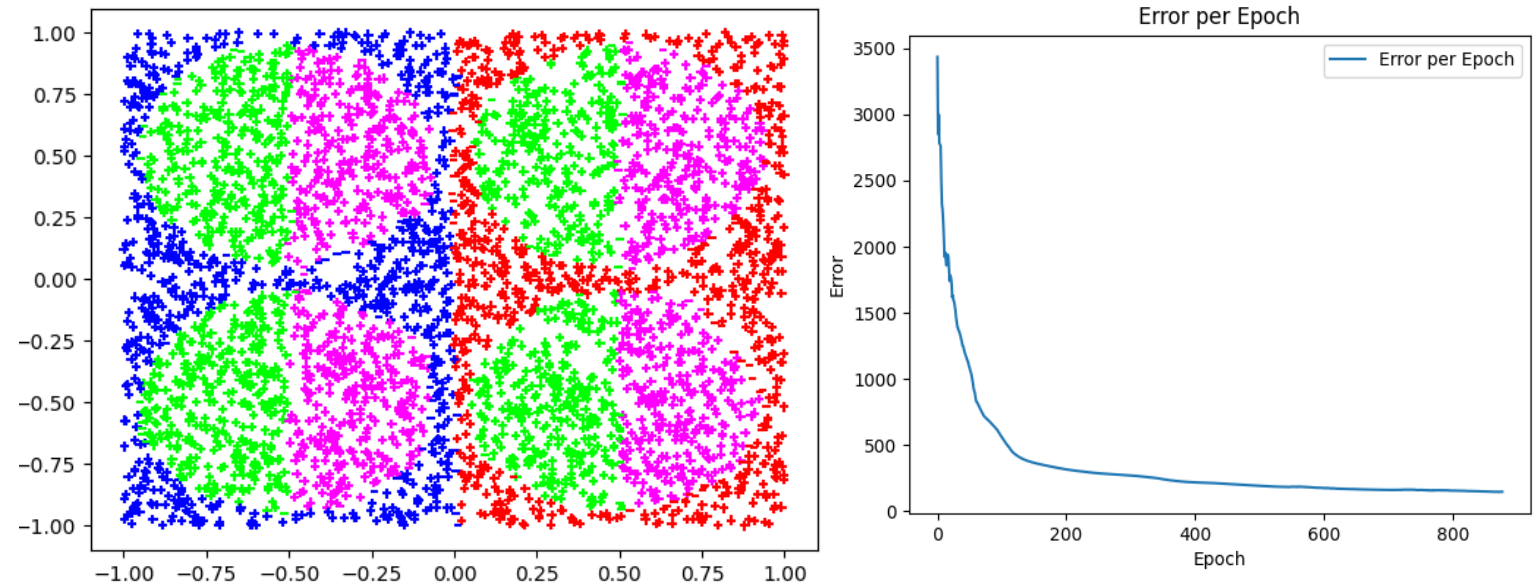
- Διάφορους συνδυασμούς τιμών των νευρώνων στα κρυμμένα επίπεδα.
- Συνάρτηση ενεργοποίησης στους κρυμμένους νευρώνες την \tanh , την relu ή την sigmoid .
- Αριθμό ομάδων $\text{numBatches}=40$ ή $\text{numBatches}=400$.

Για όλα τα πειράματα χρησιμοποιήθηκε **learningRate=0.005**, ενώ στο επίπεδο εξόδου χρησιμοποιήθηκε η συνάρτηση **sigmoid**, εκτός από τα πειράματα με το **numBatches=40** για την συνάρτηση tanh όπου χρησιμοποιήθηκε ως συνάρτηση εξόδου ξανά η *tanh* η οποία φάνηκε να λειτουργεί πιο ομαλά.

ΣΥΝΑΡΤΗΣΗ TANH

Hidden Function	Output Function	H1	H2	H3	Batches	Gener. Percentage
tanh	sigmoid	3	3	3	40	70.7%
tanh	sigmoid	6	6	6	40	80.2%
tanh	sigmoid	9	9	9	40	90.6%
tanh	sigmoid	12	12	12	40	94.4%
tanh	sigmoid	15	15	15	40	93.1%
tanh	sigmoid	18	18	18	40	95.2%
tanh	sigmoid	21	21	21	40	96.0%
tanh	sigmoid	21	18	18	40	95.4%
tanh	tanh	1	1	1	40	48%
tanh	tanh	3	3	3	40	69.2%
tanh	tanh	6	6	6	40	75.3%
tanh	tanh	9	9	9	40	92.6%
tanh	tanh	12	12	12	40	92.3%
tanh	tanh	15	15	15	40	93.8%
tanh	tanh	18	18	18	40	95.1%
tanh	tanh	21	21	21	40	95.5%
tanh	tanh	21	18	18	40	96.1%
tanh	tanh	21	15	18	40	96%

Παρατηρούμε πως όταν η συνάρτηση τόσο των κρυμμένων επιπέδων, όσο και του επιπέδου εξόδου είναι η tanh, το δίκτυο έχει την καλύτερη γενικευτική ικανότητα όταν ο αριθμός των νευρώνων στα κρυμμένα επίπεδα είναι **h1 = 21, h2 = 18 και h3 = 18**. Στην περίπτωση αυτή, το δίκτυο επιτυγχάνει **γενικευτική ικανότητα 96.1%**. Για το δίκτυο αυτό, απεικονίζουμε παρακάτω τα παραδείγματα τα οποία είτε ταξινομήθηκαν σωστά (+) είτε όχι (-). Παρακάτω φαίνεται επίσης το διάγραμμα που παρουσιάζει πως μεταβάλλεται το σφάλμα όσο περνάνε οι εποχές:



ΣΥΝΑΡΤΗΣΗ RELU

Hidden Function	Output Function	H1	H2	H3	Batches	Gener. Percentage
relu	sigmoid	1	1	1	40	31.7%
relu	sigmoid	3	3	3	40	31.7%
relu	sigmoid	6	6	6	40	31.5%
relu	sigmoid	9	9	9	40	31.7%
relu	sigmoid	12	12	12	40	31.7%
relu	sigmoid	15	15	15	40	31.5%

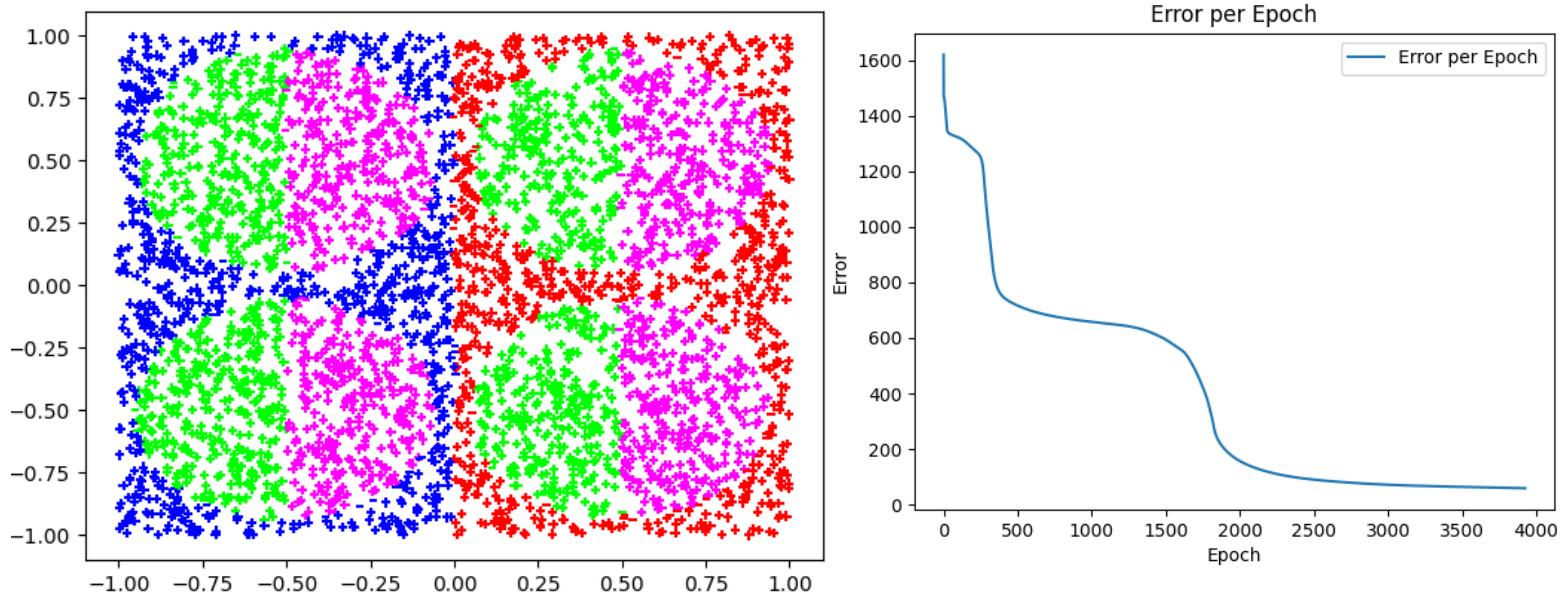
Παρατηρούμε πως όταν η συνάρτηση των κρυμμένων επιπέδων είναι η relu και του επιπέδου εξόδου είναι η sigmoid, η γενικευτική ικανότητα του δικτύου παραμένει πολύ

χαμηλή και σχεδόν σταθερή στο ποσοστό 31.7%. Το γεγονός αυτό είναι αναμενόμενο και οφείλεται στο γεγονός ότι η relu συνάρτηση παρότι δεν είναι γραμμική, η παράγωγός της για αρνητικές τιμές ισούται με μηδέν, για αυτό το δίκτυο στα αρνητικά τεταρτημόρια δεν μπορεί να μάθει.

ΣΥΝΑΡΤΗΣΗ SIGMOID

Hidden Function	Output Function	H1	H2	H3	Batches	Gener. Percentage
sigmoid	sigmoid	1	1	1	40	31.7%
sigmoid	sigmoid	3	3	3	40	71.1%
sigmoid	sigmoid	6	6	6	40	78.4%
sigmoid	sigmoid	9	9	9	40	94.2%
sigmoid	sigmoid	12	12	12	40	96.9%
sigmoid	sigmoid	15	15	15	40	97.7%
sigmoid	sigmoid	18	18	18	40	97.9%
sigmoid	sigmoid	21	21	21	40	97.9%

Παρατηρούμε πως όταν η συνάρτηση τόσο των κρυμμένων επιπέδων, όσο και του επιπέδου εξόδου είναι η sigmoid, το δίκτυο έχει την καλύτερη γενικευτική ικανότητα όταν ο αριθμός των νευρώνων στα κρυμμένα επίπεδα είναι **h1 = 21, h2 = 21 και h3 = 21**. Στην περίπτωση αυτή, το δίκτυο επιτυγχάνει **γενικευτική ικανότητα 97.3%**. Για το δίκτυο αυτό, απεικονίζουμε παρακάτω τα παραδείγματα τα οποία είτε ταξινομήθηκαν σωστά (+) είτε όχι (-). Παρακάτω φαίνεται επίσης το διάγραμμα που παρουσιάζει πως μεταβάλλεται το σφάλμα όσο περνάνε οι εποχές:

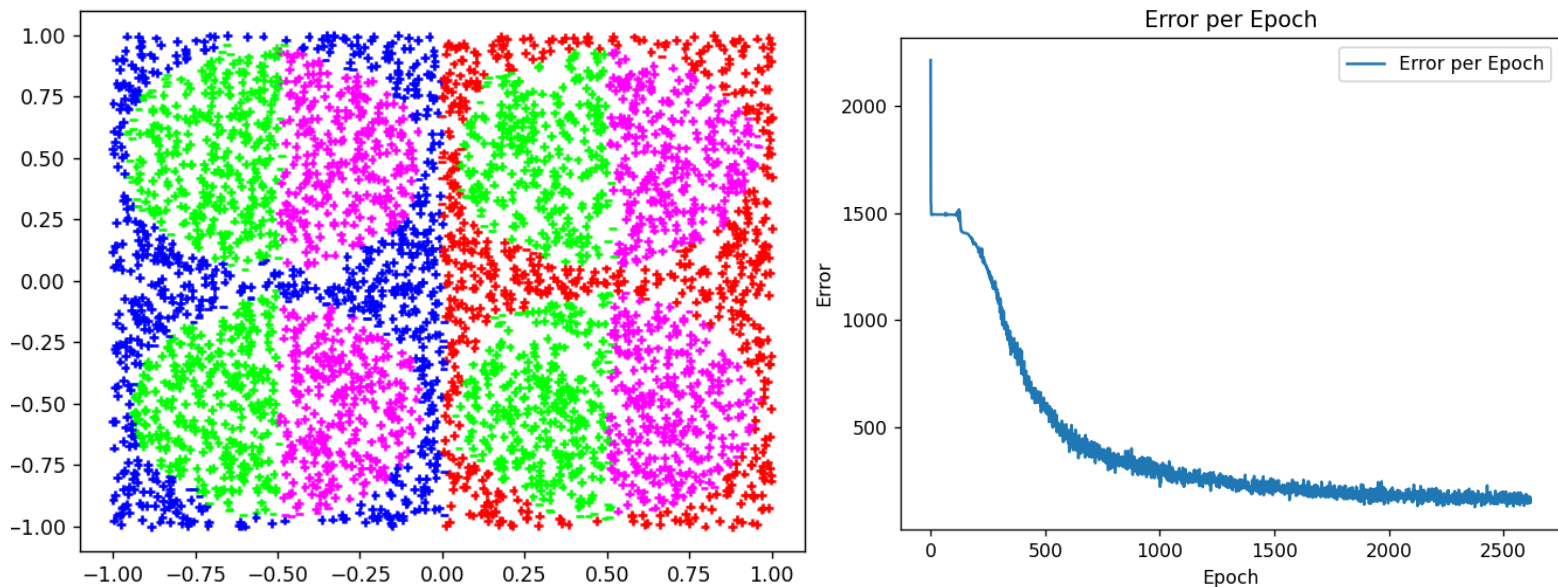


Στα παρακάτω πειράματα έχουμε θέσει των αριθμό των batches **numBatches = 400**

ΣΥΝΑΡΤΗΣΗ TANH

Hidden Function	Output Function	H1	H2	H3	Batches	Gener. Percentage
tanh	sigmoid	1	1	1	400	47.4%
tanh	sigmoid	3	3	3	400	31.8%
tanh	sigmoid	6	6	6	400	80.7%
tanh	sigmoid	9	9	9	400	87.0%
tanh	sigmoid	12	12	12	400	89.0%
tanh	sigmoid	15	15	15	400	88.6%
tanh	sigmoid	18	18	18	400	91.9%
tanh	sigmoid	21	21	21	400	95.6%
tanh	sigmoid	21	15	18	400	95.7%

Παρατηρούμε πως όταν η συνάρτηση των κρυμμένων επιπέδων είναι η tanh και του επιπέδου εξόδου είναι η sigmoid, το δίκτυο έχει την καλύτερη γενικευτική ικανότητα όταν ο αριθμός των νευρώνων στα κρυμμένα επίπεδα είναι **h1 = 21, h2 = 15** και **h3 = 18**. Στην περίπτωση αυτή, το δίκτυο επιτυγχάνει **γενικευτική ικανότητα 95.7%**. Για το δίκτυο αυτό, απεικονίζουμε παρακάτω τα παραδείγματα τα οποία είτε ταξινομήθηκαν σωστά (+) είτε όχι (-). Παρακάτω φαίνεται επίσης το διάγραμμα που παρουσιάζει πως μεταβάλλεται το σφάλμα όσο περνάνε οι εποχές:



ΣΥΝΑΡΤΗΣΗ RELU

Hidden Function	Output Function	H1	H2	H3	Batches	Gener. Percentage
relu	sigmoid	1	1	1	400	31.8%
relu	sigmoid	3	3	3	400	31.6%
relu	sigmoid	6	6	6	400	31.6%
relu	sigmoid	9	9	9	400	31.8%
relu	sigmoid	12	12	12	400	31.8%
relu	sigmoid	15	15	15	400	31.8%
relu	sigmoid	18	18	18	400	31.6%
relu	sigmoid	21	21	21	400	31.6%
relu	sigmoid	21	15	18	400	31.8%

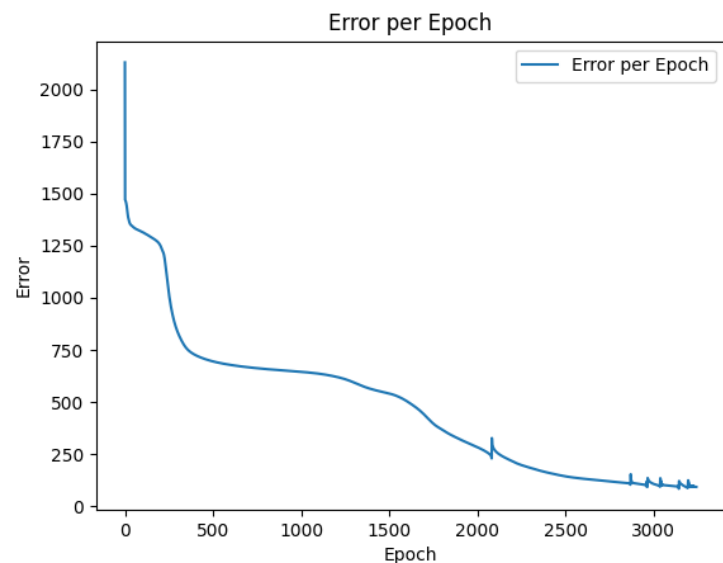
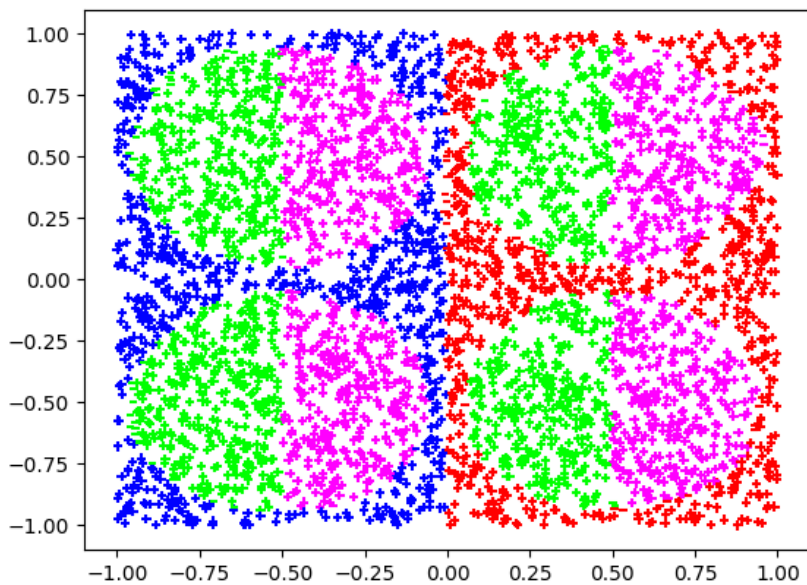
Παρατηρούμε πως όταν η συνάρτηση των κρυμμένων επιπέδων είναι η relu και του επιπέδου εξόδου είναι η sigmoid, η γενικευτική ικανότητα του δικτύου παραμένει πολύ χαμηλή και σχεδόν σταθερή στο ποσοστό 31.7%. Το γεγονός αυτό είναι αναμενόμενο και οφείλεται στο γεγονός ότι η relu συνάρτηση παρότι δεν είναι γραμμική, η παράγωγός της για αρνητικές τιμές ισούται με μηδέν, για αυτό το δίκτυο στα αρνητικά τεταρτημόρια δεν μπορεί να μάθει.

ΣΥΝΑΡΤΗΣΗ SIGMOID

Hidden Function	Output Function	H1	H2	H3	Batches	Gener. Percentage
sigmoid	sigmoid	1	1	1	400	31.7%
sigmoid	sigmoid	3	3	3	400	69.4%
sigmoid	sigmoid	6	6	6	400	78.9%

sigmoid	sigmoid	9	9	9	400	78.9%
sigmoid	sigmoid	12	12	12	400	78.7%
sigmoid	sigmoid	15	15	15	400	96%
sigmoid	sigmoid	18	18	18	400	93.2%
sigmoid	sigmoid	21	21	21	400	97.3%
sigmoid	sigmoid	21	15	15	400	95.5%

Παρατηρούμε πως όταν η συνάρτηση τόσο των κρυμμένων επιπέδων, όσο και του επιπέδου εξόδου είναι η sigmoid, το δίκτυο έχει την καλύτερη γενικευτική ικανότητα όταν ο αριθμός των νευρώνων στα κρυμμένα επίπεδα είναι **$h_1 = 21$, $h_2 = 21$** και **$h_3 = 21$** . Στην περίπτωση αυτή, το δίκτυο επιτυγχάνει **γενικευτική ικανότητα 97.3%**. Για το δίκτυο αυτό, απεικονίζουμε παρακάτω τα παραδείγματα τα οποία είτε ταξινομήθηκαν σωστά (+) είτε όχι (-). Παρακάτω φαίνεται επίσης το διάγραμμα που παρουσιάζει πως μεταβάλλεται το σφάλμα όσο περνάνε οι εποχές:



ΣΥΜΠΕΡΑΣΜΑ

Προκύπτει, λοιπόν, το συμπέρασμα ότι το δίκτυο με την καλύτερη γενικευτική ικανότητα είναι εκείνο το οποίο έχει συνάρτηση τόσο στα κρυμμένα επίπεδα, όσο και στο επίπεδο εξόδου την **sigmoid** και **numBatches=40**. Το δίκτυο αυτό με **18 νευρώνες** σε κάθε κρυμμένο επίπεδο, παρουσιάζει ποσοστό επιτυχίας στο σύνολο ελέγχου **97.9%**.