ΥΠΟΛΟΓΙΣΤΙΚΗ ΝΟΗΜΟΣΥΝΗ

Ακαδημαϊκό Έτος 2024-2025

Εργαστηριακή Άσκηση

Μέρος Α΄

Ον/μο: Φίλιππος Μινώπετρος

ΑΜ: 1093431

Εξάμηνο: 8ο

Email: up1093431@ac.upatras.gr

Κώδικας: https://github.com/F1114/computational-intelligence-project

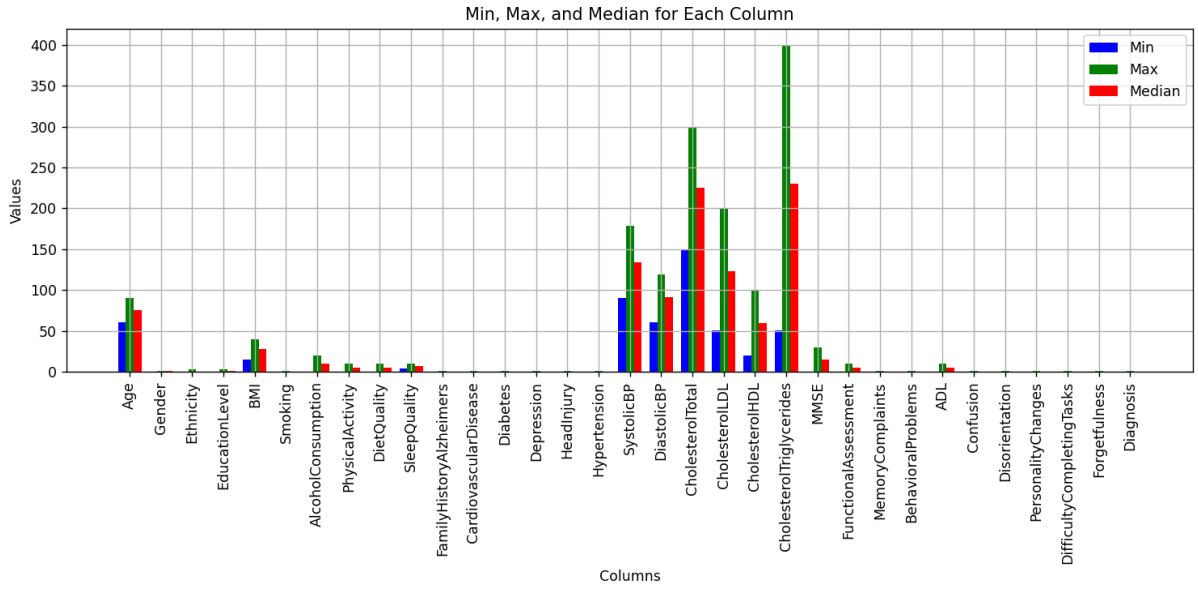
Περιεχόμενα

Α1. Προεπεξεργασία και Προετοιμασία δεδομένων

Α) Προτού ξεκινήσουμε την εκπαίδευση του νευρωνικού δικτύου κάνουμε τις κατάλληλες προσαρμογές στο δοσμένο dataset. Αρχικά στα κατηγορικά δεδομένα Ethnicity, EducationLevel, Gender εφαρμόζουμε One hot encoding ώστε να αποφευχθεί η έννοια της σειράς για αυτές τις πληροφορίες. Στη συνέχεια καθώς δεν υπάρχουν γραμμές με ελλιπή δεδομένα και ορισμένες στήλες είναι ήδη σε δυαδική αναπαράσταση κάνω τυποποίηση των υπόλοιπων στηλών. Επιλέγω τυποποίηση έναντι κανονικοποίησης καθώς η τυποποίηση μπορεί να παρέχει καλύτερη διαχείριση ακραίων τιμών αφού εξαρτάται και από τη μέση τιμή αλλά και από την τυπική απόκλιση. Το κεντράρισμα ομοίως με την κανονικοποίηση δεν θα παρέχει την καλύτερη διαχείριση.

$$X_{std} = \frac{X - \mu}{\sigma}$$

Β) Για τον διαχωρισμό του dataset σε folds θα χρησιμοποιήσουμε το KFold και ειδικότερα το Stratified KFold που φροντίζει κάθε Fold να είναι ισορροπημένο σε σχέση με ολόκληρο το dataset.



Α2. Επιλογή αρχιτεκτονικής

Α) Για την εκπαίδευση του νευρωνικού δικτύου θέλουμε μία μετρική που θα βοηθάει το δίκτυο όχι μόνο να καταλάβει αν κάνει μία σωστή ή λάθος πρόβλεψη, αλλά αν είναι λάθος θα πρέπει να γνωρίζει κατά πόσο θα διορθώσει. Το Accuracy δείχνει μόνο σωστό ή λάθος, πράγμα που όπως εξηγήσαμε δε θέλουμε.

$$\text{Accuracy} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(y_i = \hat{y}_i)$$

Το MSE αποτελεί μία καλύτερη μετρική ωστόσο οι μεταβολές του είναι πολύ μικρές καθώς περιέχει μόνο την απόσταση του προβλεπόμενου από το επιθυμητό και τιμές του κυμαίνονται στο διάστημα 0 έως 1.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Η καλύτερη μετρική στην περίπτωση ενός προβλήματος Classification είναι η Cross Entropy καθώς δίνει μικρά αποτελέσματα για σωστή πρόβλεψη , άρα και μικρή διόρθωση, και μεγάλα αποτελέσματα για λάθος πρόβλεψη, άρα και μεγάλη διόρθωση. Επίσης το διάστημα που κυμαίνονται οι τιμές της είναι μεγαλύτερο από του MSE.

$$\text{CrossEntropy} = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right]$$

Β) Για το πρόβλημα ταξινόμησης σε δύο κλάσεις αρκεί ένας νευρώνας στο επίπεδο εξόδου, ο οποίος αντιστοιχεί στη διάγνωση ή μη του ασθενούς.

Γ)

Ξεκινάμε αναλύοντας τις δοθέντες συναρτήσεις ενεργοποίησης. Αρχικά η Tanh έχει το πρόβλημα των vanishing gradients. Δηλαδή όσο πιο κοντά βρίσκεται η είσοδος στις max & min τιμές του διαστήματος που εκτείνεται [-1, 1] τόσο

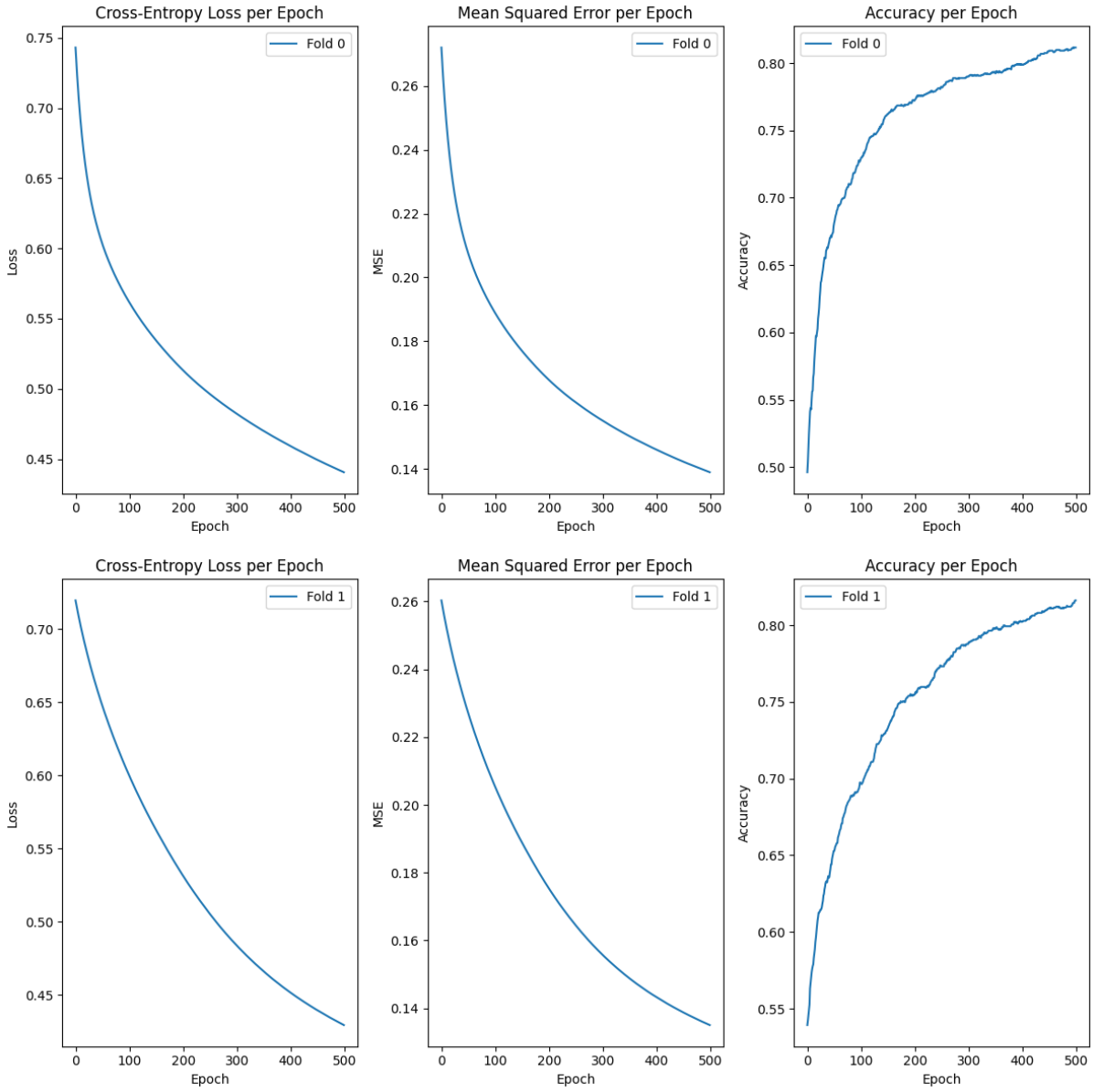
μικρότερο είναι και το gradient άρα η ταχύτητα μάθηση. Η SiLU δεν αντιμετωπίζει το ίδιο πρόβλημα με την Tanh καθώς , εκτείνεται από το μείον άπειρο έως άπειρο ωστόσο είναι υπολογιστικά ακριβή και δυσχεραίνει την εκπαίδευση. Τέλος, η ReLU είναι πιο υπολογιστικά φθηνή καθώς το gradient της είναι είτε 0 είτε 1 πράγμα που την θέτει ικανοποιητική καθώς αν η είσοδος της είναι μικρότερη από 0 τότε ο νευρώνας τίθεται για την συγκεκριμένη στιγμή ως ανενεργός ενώ αν είναι μεγαλύτερη από 0 τότε τίθεται ενεργός και διορθώνονται τα βάρη του, συνεπώς είναι και η καλύτερη επιλογή.

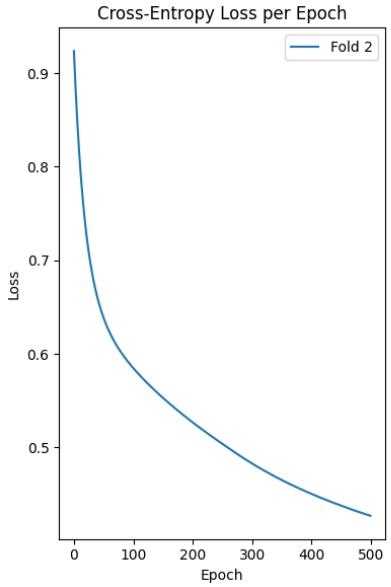
Δ) Όσον αφορά την συνάρτηση ενεργοποίησης του επιπέδου εξόδου, η γραμμική συνάρτηση παρέχει έξοδο που είναι γραμμικός συνδυασμός της εισόδου συνεπώς δεν είναι κατάλληλη για ένα πρόβλημα classification αλλά περισσότερο για την πρόβλεψη συνεχών τιμών όπως . Η Softmax μπορεί να λύσει ένα πρόβλημα classification ωστόσο λειτουργεί καλύτερα για περισσότερες από 2 κλάσεις καθώς περιγράφει την πιθανότητα να ανήκει σε όλες που σε αυτή την περίπτωση δεν είναι ιδανικό αφού θέλουμε να είμαστε σίγουροι για το αποτέλεσμα. Αναλύοντας και τη σιγμοειδή βλέπουμε ότι από τη στιγμή που κυμαίνεται στο διάστημα 0 έως 1 μπορεί να λειτουργήσει ικανοποιητικά για το πρόβλημα του classification και ορίζοντας ένα threshold μπορεί να απαντήσει με σιγουριά σε ποια από της δύο κλάσεις θα ανήκει το αποτέλεσμα. Αναζητώντας περισσότερες συναρτήσεις φαίνεται πως πολλές είναι παραλλαγές των προηγούμενων, συνεπώς για τη συγκεκριμένη περίπτωση η σιγμοειδής θα είναι αρκετή.

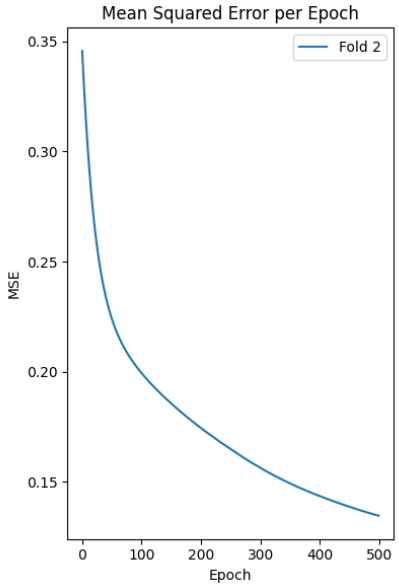
Ε) Για ορίσουμε το διάστημα [I/2, 2I] θέτουμε I ίσο με τον αριθμό των εισόδων , δηλαδή των αριθμών των στηλών του dataset εκτός από την τελευταία. Οπότε I = 39 , I Ε [20, 78]

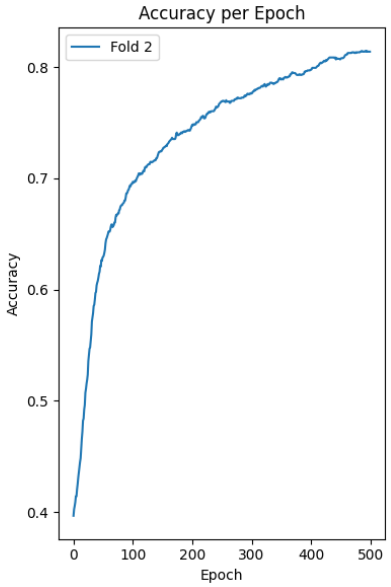
Αριθμός νευρώνων στο κρυφό επίπεδο	**CE loss**	**MSE**	**Acc**
$H_1 = I/2 = 20$	0.46	0.15	0.79
$H_2 = 2I/3 = 26$	0.46	0.15	0.79
$H_3 = I = 39$	0.45	0.14	0.80
$H_4 = 2I = 78$	0.43	0.14	0.82

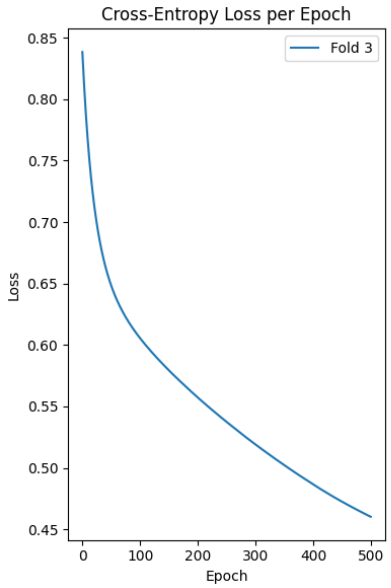
1)

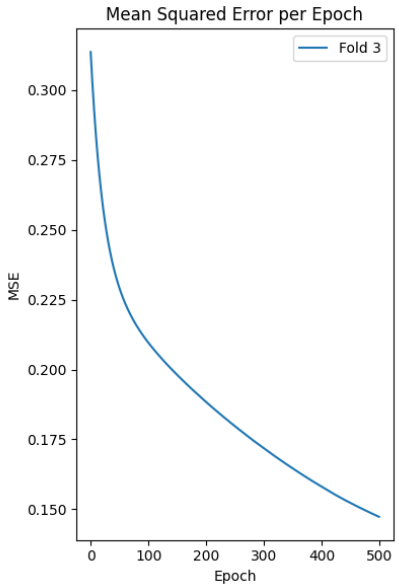

Cross-Entropy Loss per Epoch
Fold 0
Loss
Epoch
Mean Squared Error per Epoch
Fold 0
MSE
Epoch
Accuracy per Epoch
Fold 0
Accuracy
Epoch
Cross-Entropy Loss per Epoch
Fold 1
Loss
Epoch
Mean Squared Error per Epoch
Fold 1
MSE
Epoch
Accuracy per Epoch
Fold 1
Accuracy
Epoch

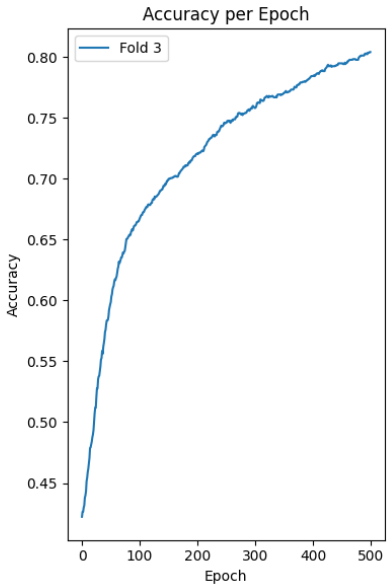

Cross-Entropy Loss per Epoch
Fold 2
Loss
Epoch

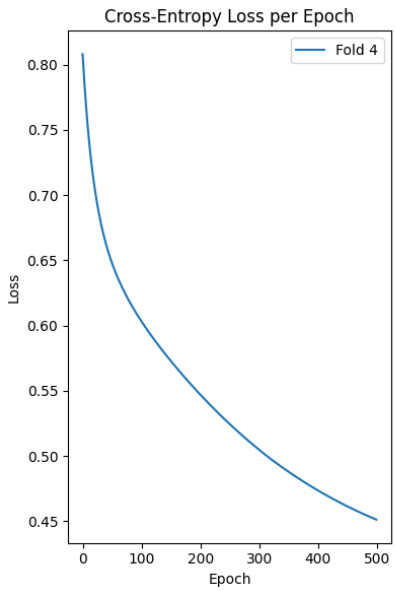

Mean Squared Error per Epoch
Fold 2
MSE
Epoch

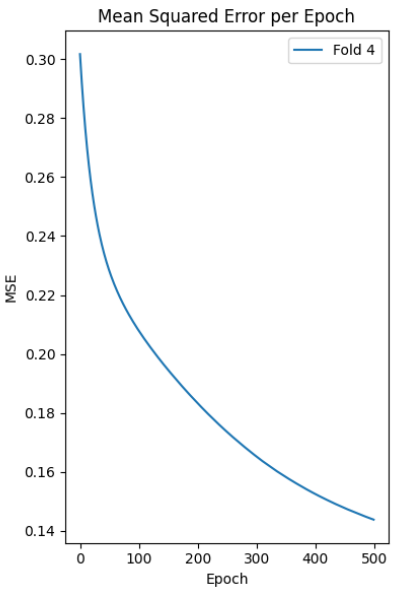

Accuracy per Epoch
Fold 2
Accuracy
Epoch

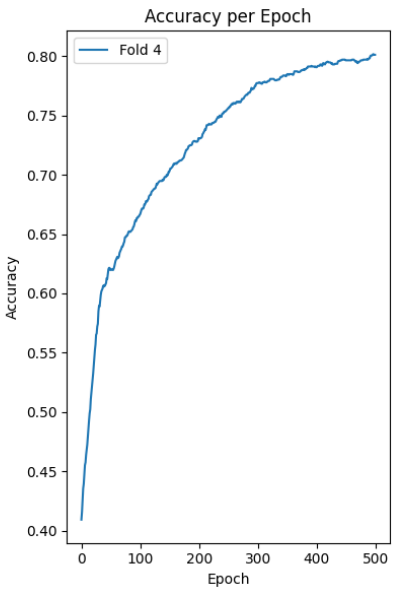

Cross-Entropy Loss per Epoch
Fold 3
Loss
Epoch


Mean Squared Error per Epoch
Fold 3
MSE
Epoch

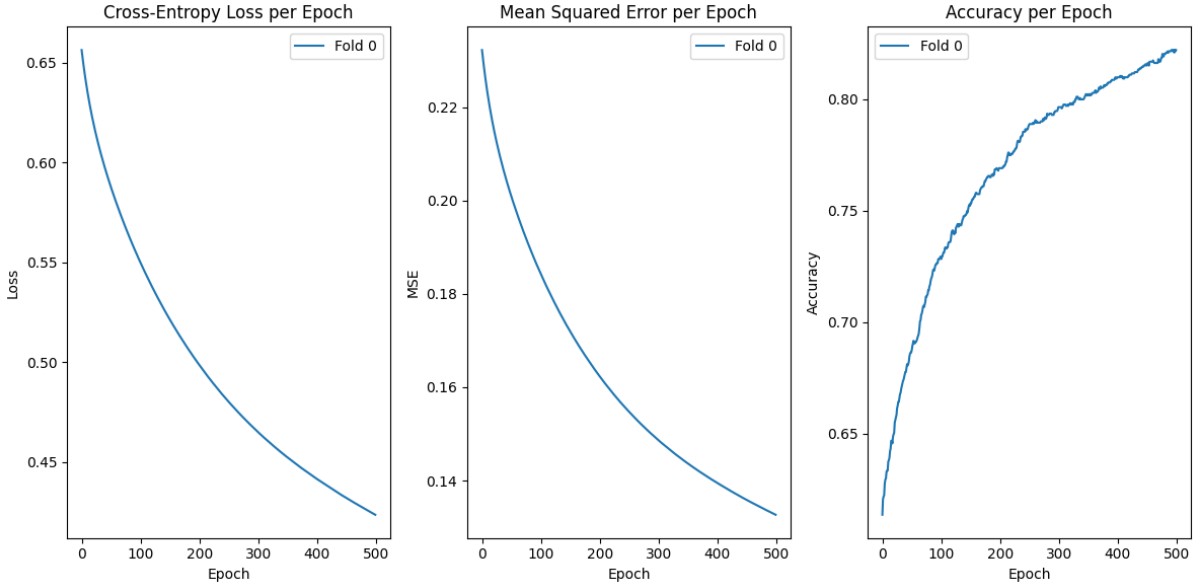

Accuracy per Epoch
Fold 3
Accuracy
Epoch

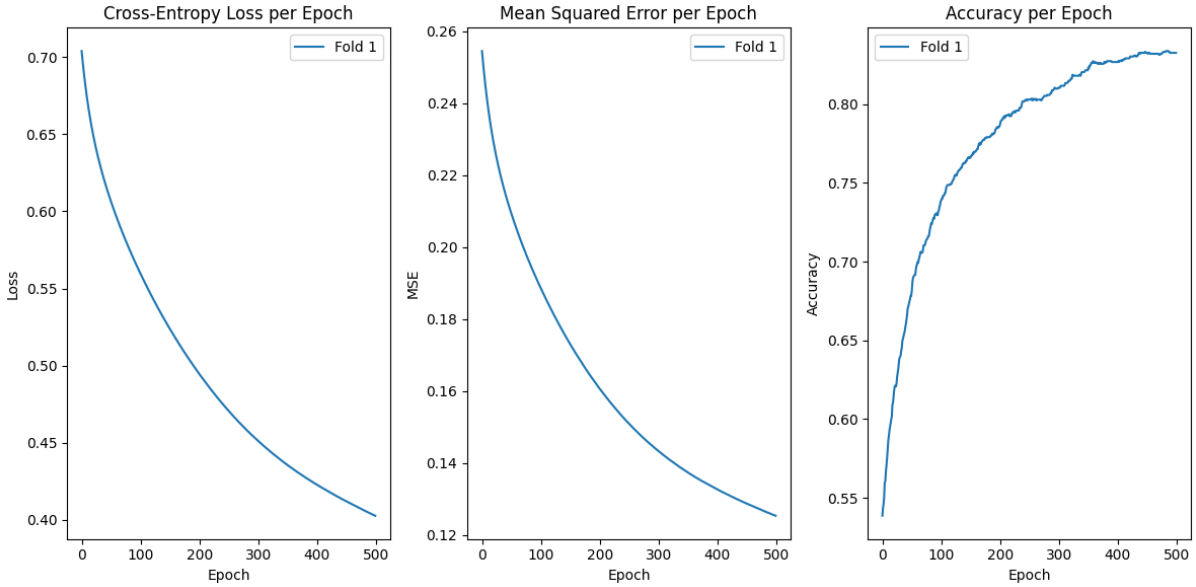

Cross-Entropy Loss per Epoch
Fold 4
Loss
Epoch

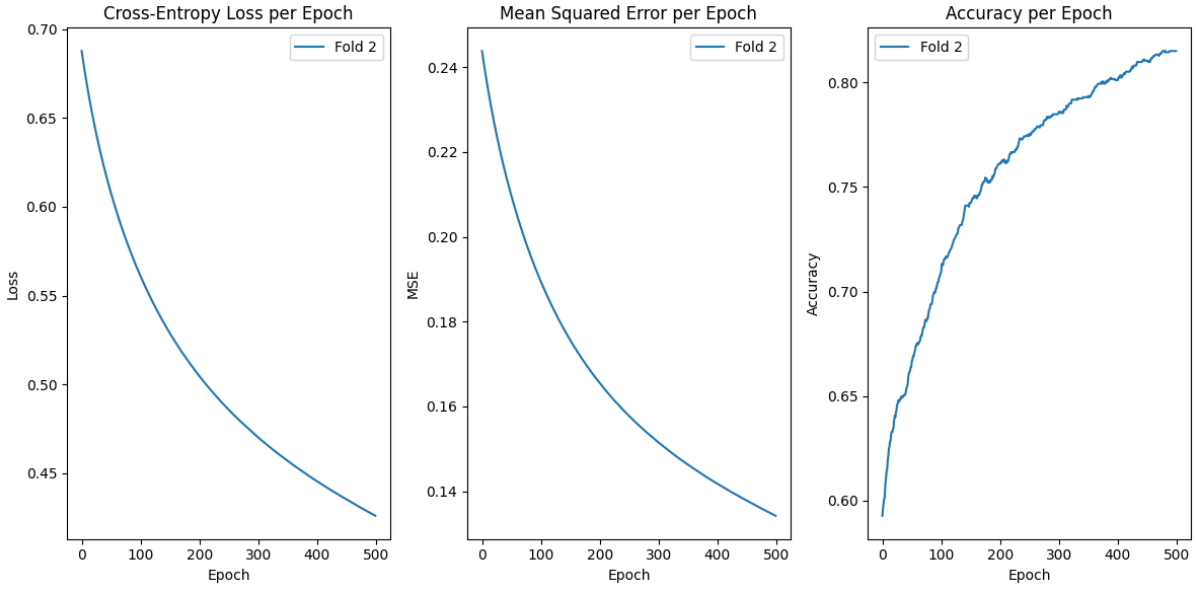

Mean Squared Error per Epoch
Fold 4
MSE
Epoch

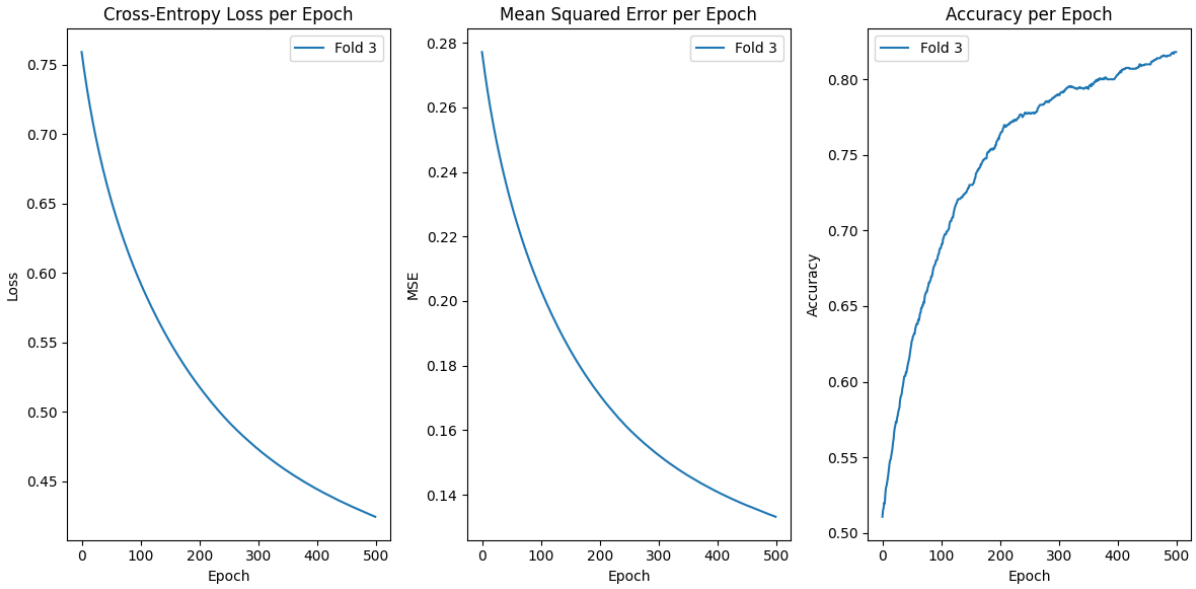

Accuracy per Epoch
Fold 4
Accuracy
Epoch

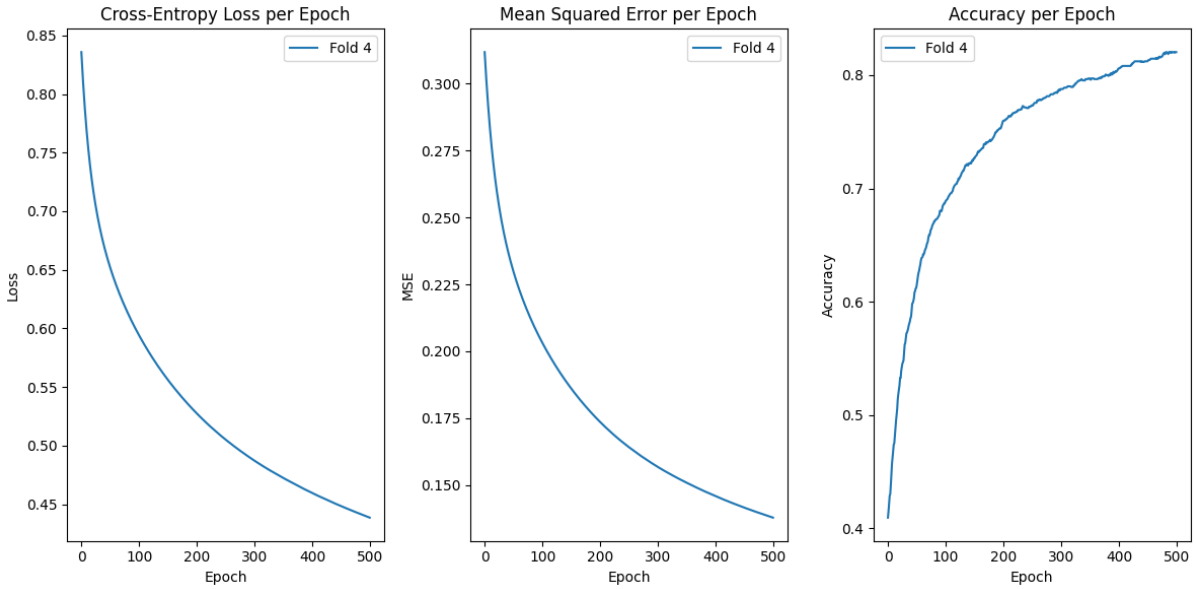
2)


Cross-Entropy Loss per Epoch
Fold 0
Loss
Epoch
Mean Squared Error per Epoch
Fold 0
MSE
Epoch
Accuracy per Epoch
Fold 0
Accuracy
Epoch

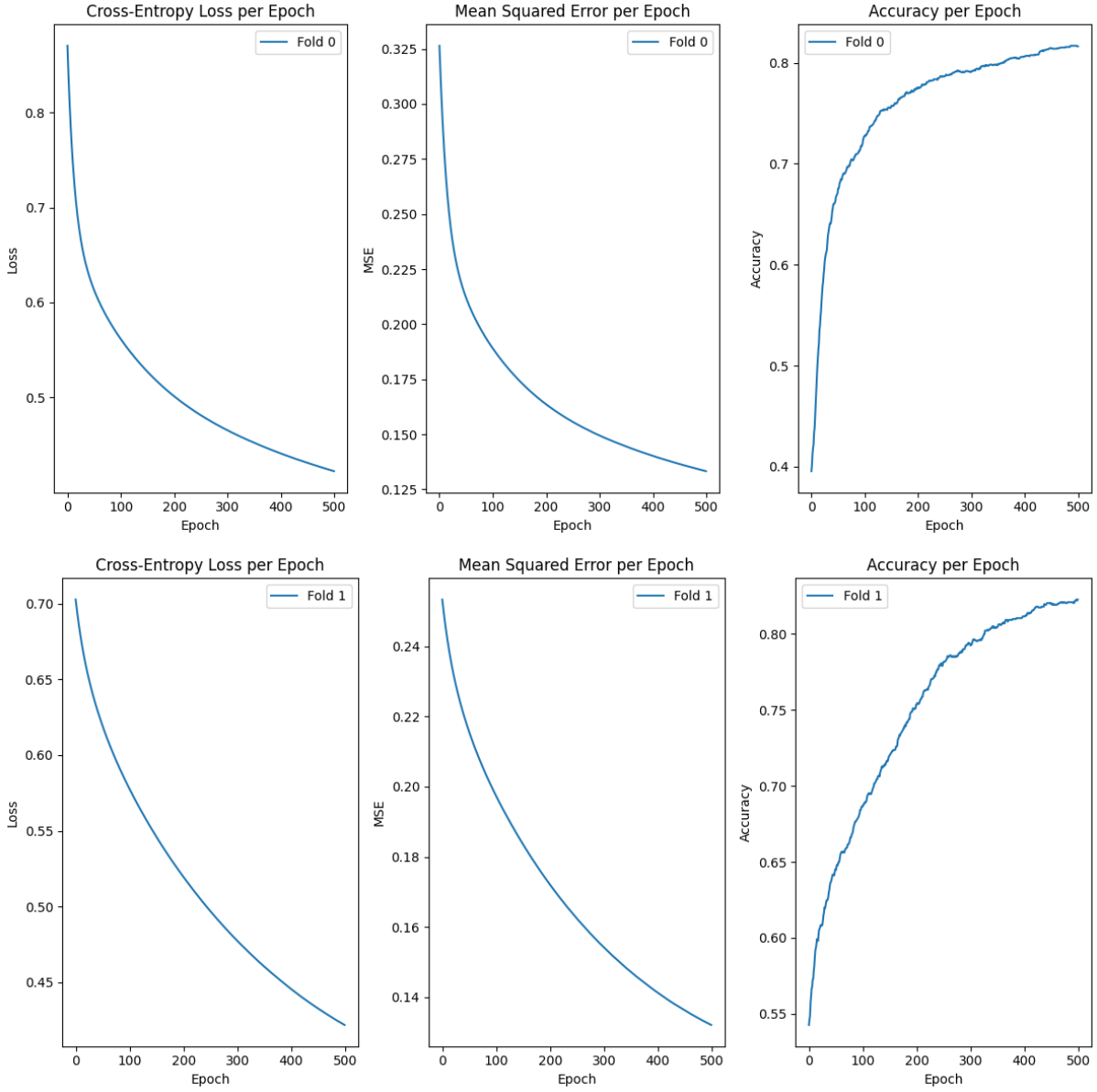

Cross-Entropy Loss per Epoch
Fold 1
Loss
Epoch
Mean Squared Error per Epoch
Fold 1
MSE
Epoch
Accuracy per Epoch
Fold 1
Accuracy
Epoch

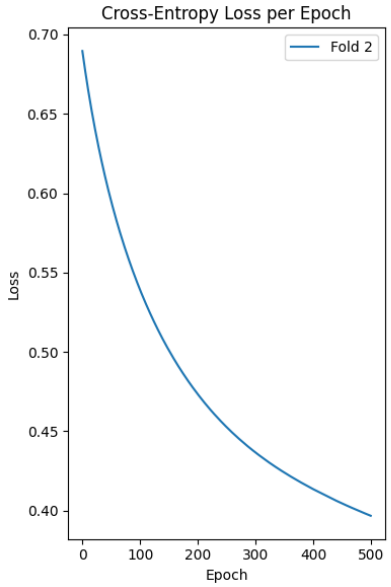

Cross-Entropy Loss per Epoch
Fold 2
Loss
Epoch
Mean Squared Error per Epoch
Fold 2
MSE
Epoch
Accuracy per Epoch
Fold 2
Accuracy
Epoch

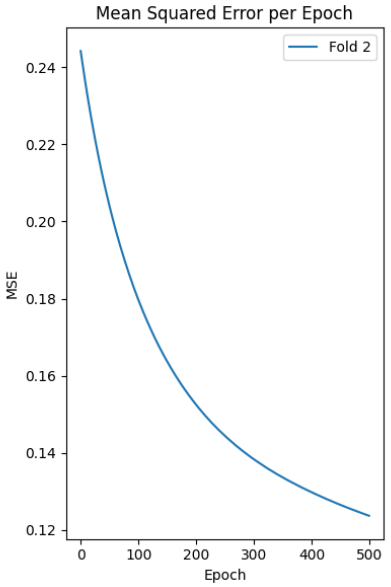

Cross-Entropy Loss per Epoch
Fold 3
Loss
Epoch
Mean Squared Error per Epoch
Fold 3
MSE
Epoch
Accuracy per Epoch
Fold 3
Accuracy
Epoch

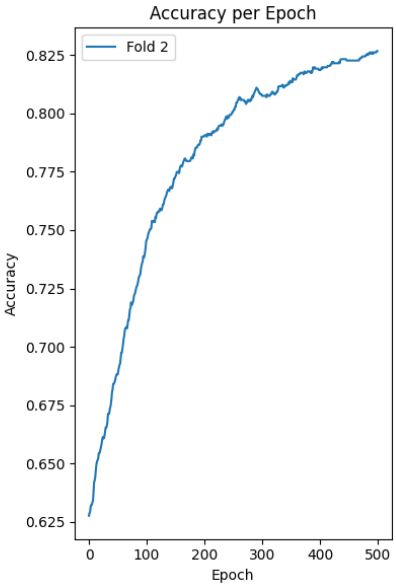

Cross-Entropy Loss per Epoch
Fold 4
Loss
Epoch
Mean Squared Error per Epoch
Fold 4
MSE
Epoch
Accuracy per Epoch
Fold 4
Accuracy
Epoch

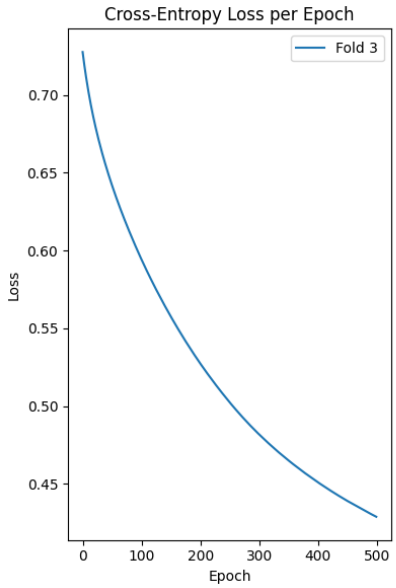
3)

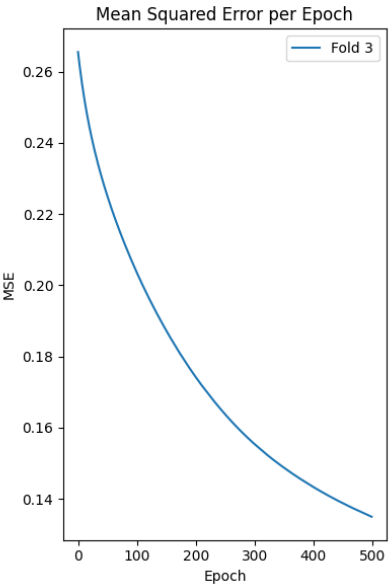

Cross-Entropy Loss per Epoch
Fold 0
Loss
Epoch
Mean Squared Error per Epoch
Fold 0
MSE
Epoch
Accuracy per Epoch
Fold 0
Accuracy
Epoch
Cross-Entropy Loss per Epoch
Fold 1
Loss
Epoch
Mean Squared Error per Epoch
Fold 1
MSE
Epoch
Accuracy per Epoch
Fold 1
Accuracy
Epoch

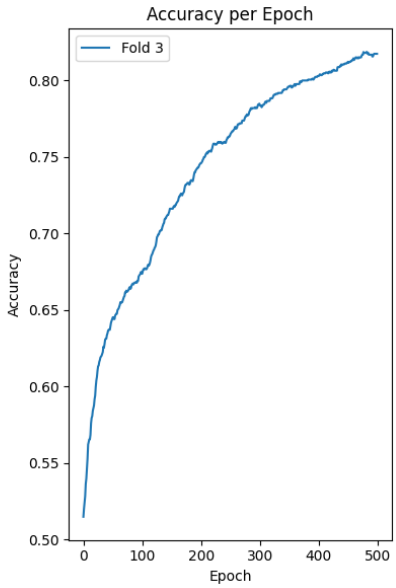

Cross-Entropy Loss per Epoch
Fold 2
Loss
Epoch

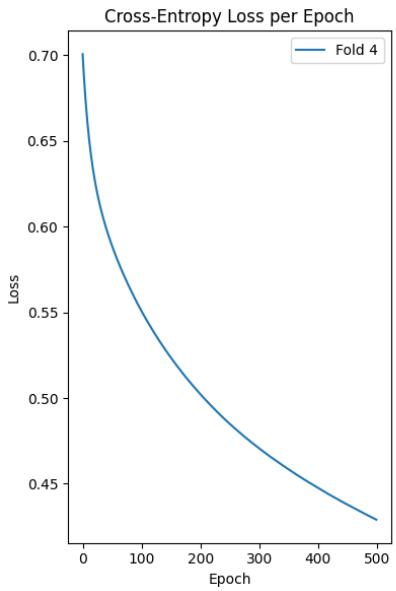

Mean Squared Error per Epoch
Fold 2
MSE
Epoch

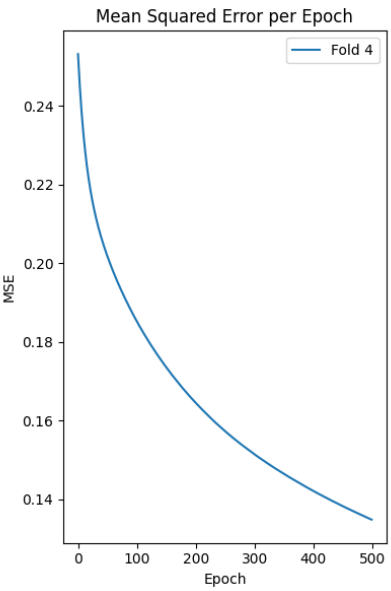

Accuracy per Epoch
Fold 2
Accuracy
Epoch

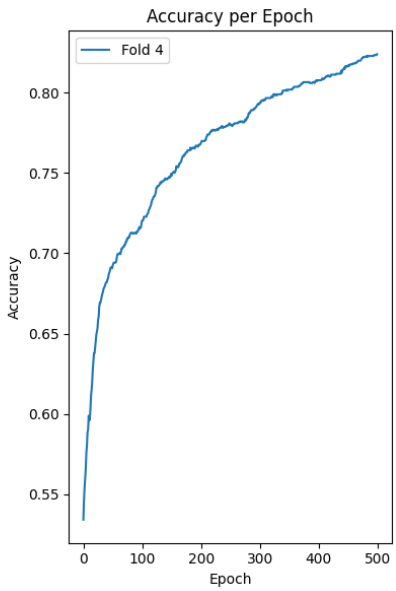

Cross-Entropy Loss per Epoch
Fold 3
Loss
Epoch


Mean Squared Error per Epoch
Fold 3
MSE
Epoch

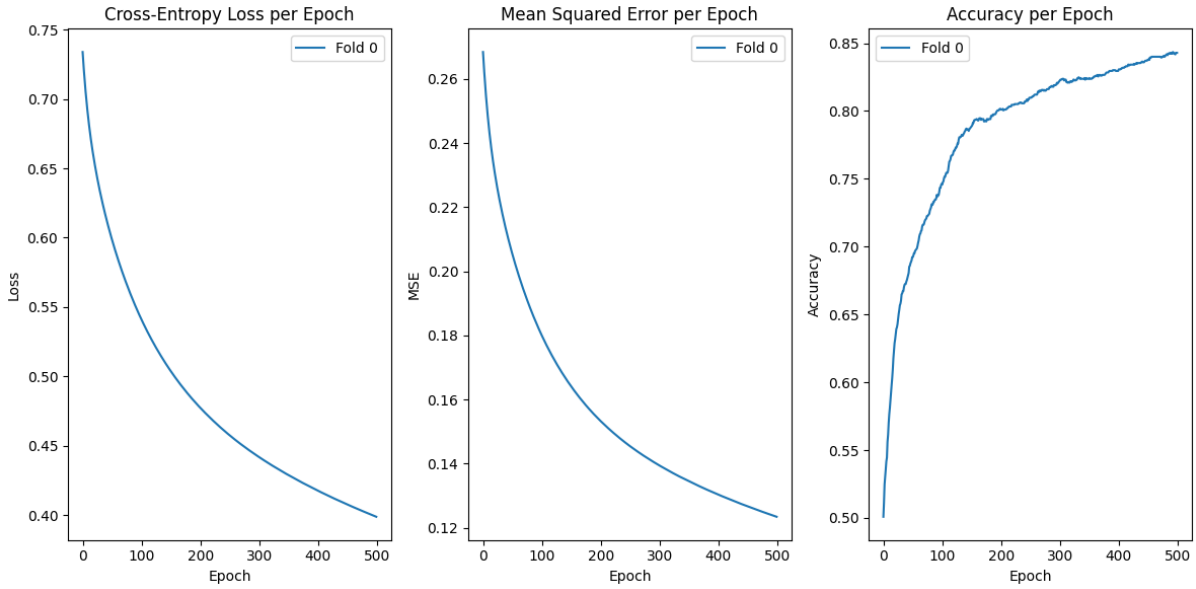

Accuracy per Epoch
Fold 3
Accuracy
Epoch

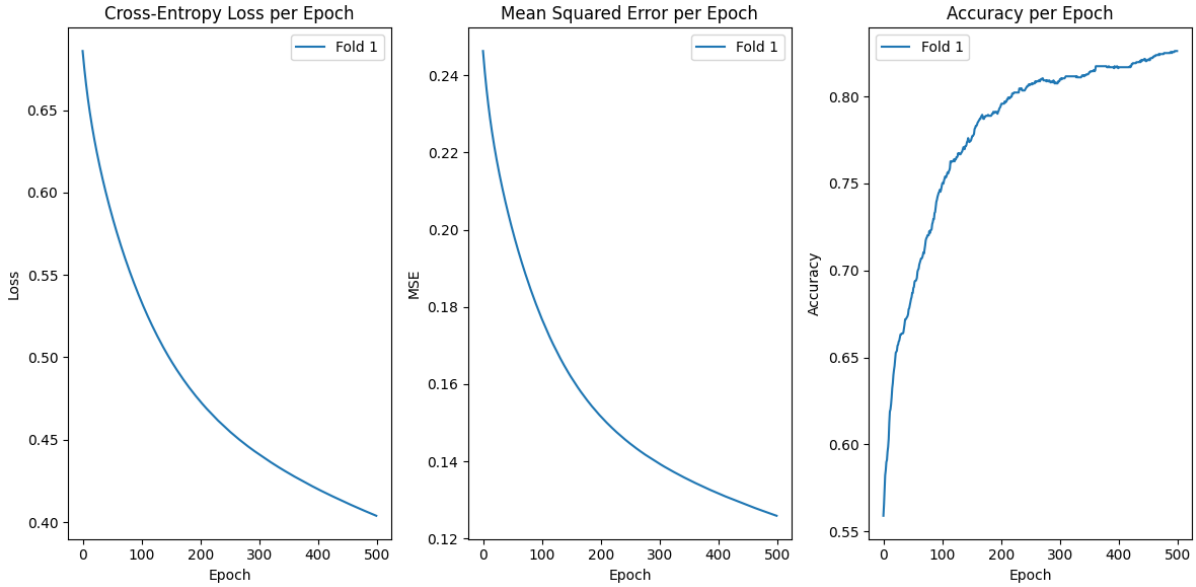

Cross-Entropy Loss per Epoch
Fold 4
Loss
Epoch

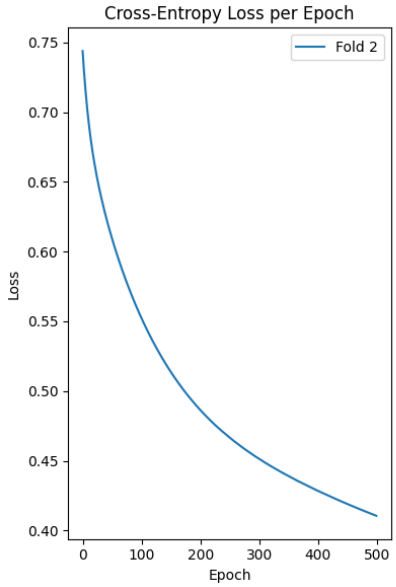

Mean Squared Error per Epoch
Fold 4
MSE
Epoch

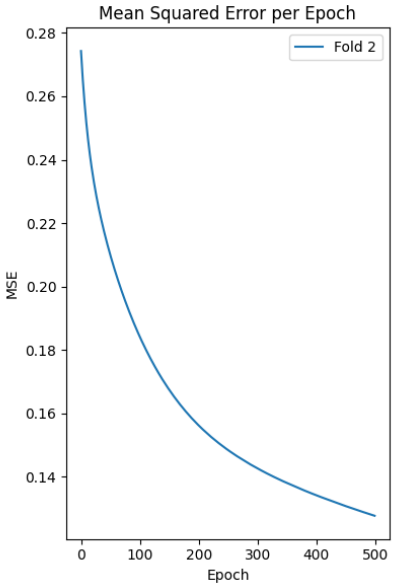

Accuracy per Epoch
Fold 4
Accuracy
Epoch

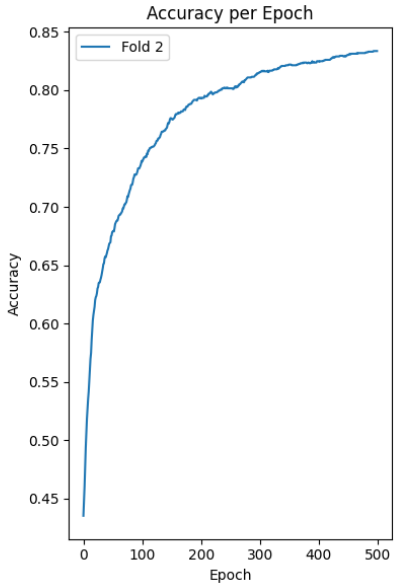
4)

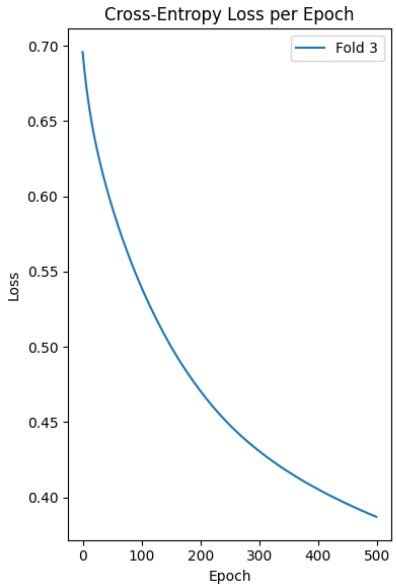

Cross-Entropy Loss per Epoch
Fold 0
Loss
Epoch
Mean Squared Error per Epoch
Fold 0
MSE
Epoch
Accuracy per Epoch
Fold 0
Accuracy
Epoch

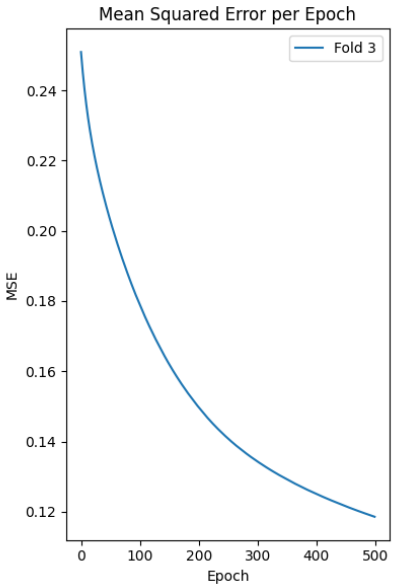

Cross-Entropy Loss per Epoch
Fold 1
Loss
Epoch
Mean Squared Error per Epoch
Fold 1
MSE
Epoch
Accuracy per Epoch
Fold 1
Accuracy
Epoch

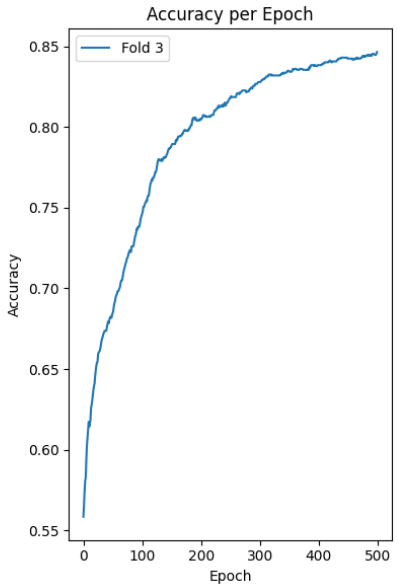

Cross-Entropy Loss per Epoch
Fold 2
Loss
Epoch

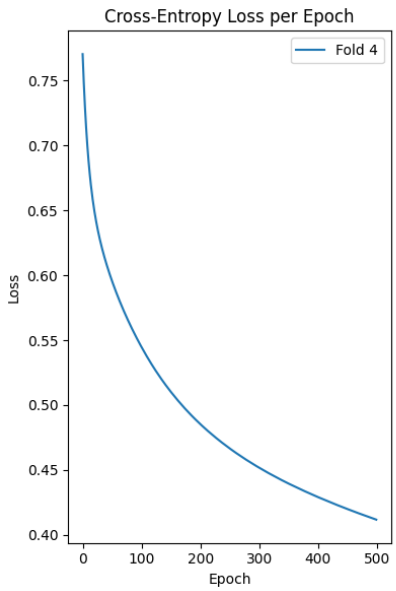

Mean Squared Error per Epoch
Fold 2
MSE
Epoch

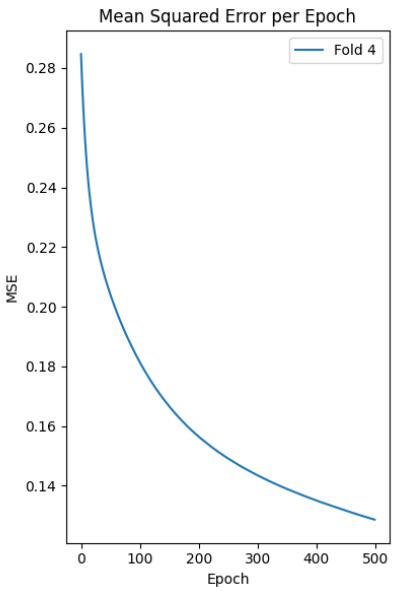

Accuracy per Epoch
Fold 2
Accuracy
Epoch

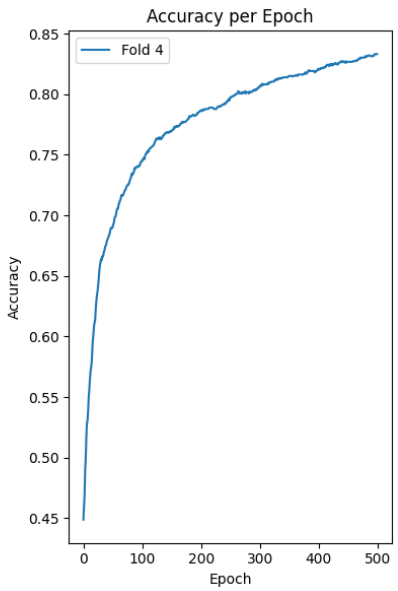

Cross-Entropy Loss per Epoch
Fold 3
Loss
Epoch


Mean Squared Error per Epoch
Fold 3
MSE
Epoch


Accuracy per Epoch
Fold 3
Accuracy
Epoch


Cross-Entropy Loss per Epoch
Fold 4
Loss
Epoch


Mean Squared Error per Epoch
Fold 4
MSE
Epoch


Accuracy per Epoch
Fold 4
Accuracy
Epoch

i) Παρατηρούμε ότι όσο αυξάνεται ο αριθμός των κόμβων του κρυφού επιπέδου έχουμε και αύξηση στην στην επιτυχία του δικτύου.
ii) Ως συνάρτηση κόστους για την εκπαίδευση του δικτύου χρησιμοποιήθηκε η binary cross entropy που ταιριάζει για το πρόβλημα του classification.
iii) Όπως αιτιολογήθηκε και σε προηγούμενο ερώτημα η συνάρτηση ενεργοποίησης που χρησιμοποιήθηκε για τους νευρώνες του κρυφού επιπέδου είναι η ReLU και η πειραματική παρατήρηση επιβεβαιώνει τη θεωρία.
iv) Μέσω των διαγραμμάτων παρατηρούμε ότι έχουμε πολύ αυξημένη ταχύτητα σύγκλισης κατά την αρχή της εκπαίδευσης, και όσο προχωράνε εποχές και το δίκτυο σταθεροποιείται γύρω από τη βέλτιστη λύση η ταχύτητα μειώνεται.

ΣΤ) Για τον τερματισμό χρησιμοποιείται early stopping που κάνει monitor τη cross entropy, που αν δεν μειωθεί για 10 εποχές τότε η εκπαίδευση σταματά και τα βάρη αντικαθιστούνται με τα βέλτιστα.

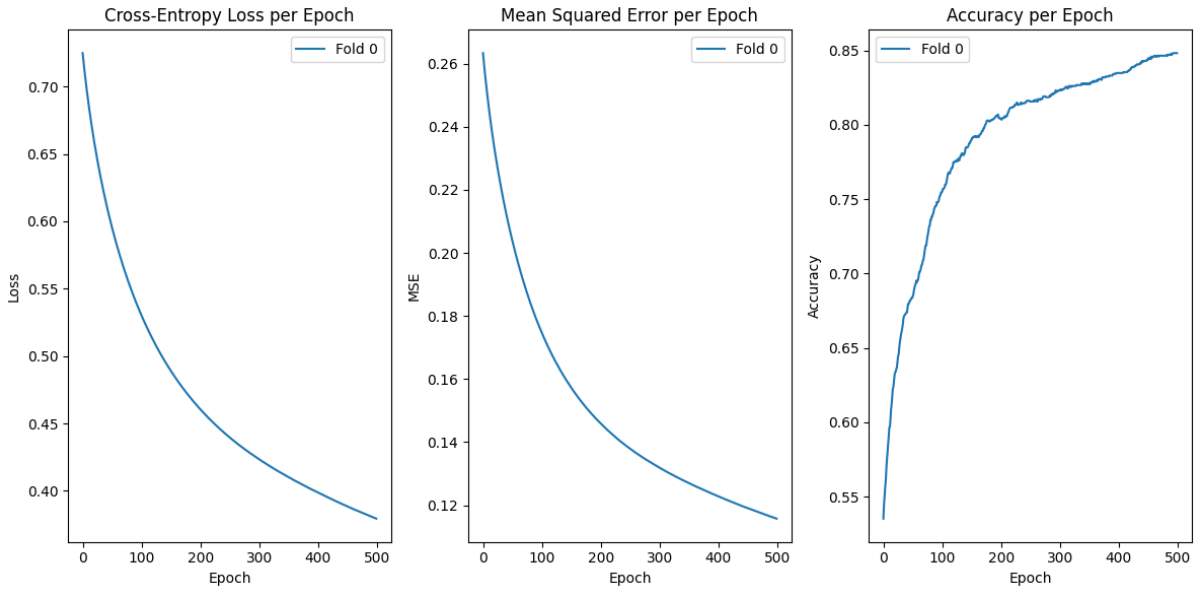
Α3. Μεταβολές στον ρυθμό εκπαίδευσης και σταθεράς ορμής

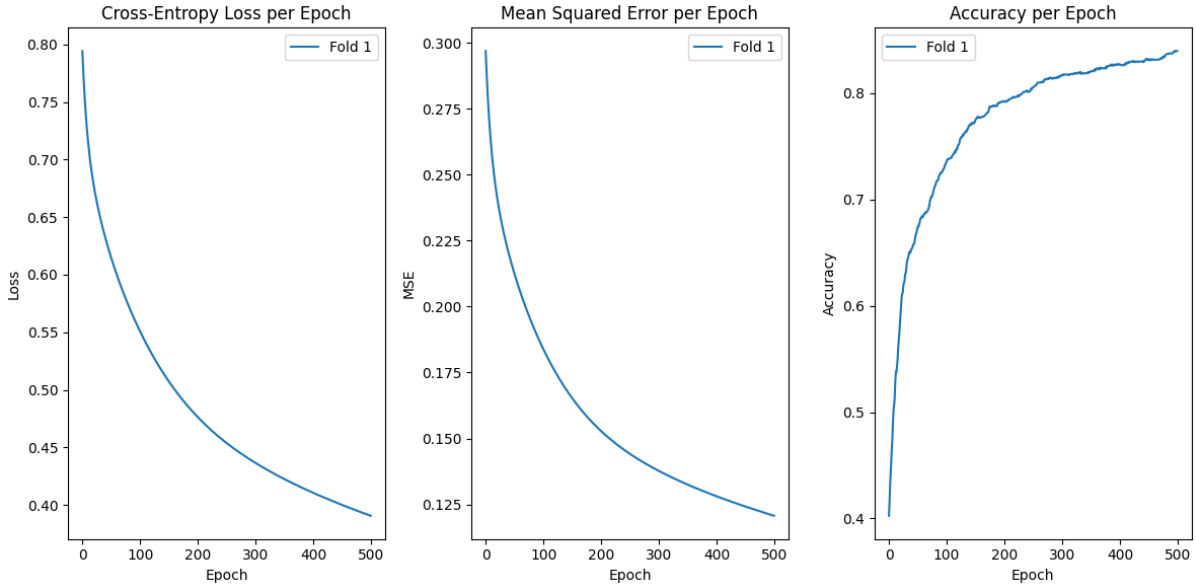
η	**m**	**CE loss**	**MSE**	**Acc**
0.001	0.2	0.41	0.13	0.82
0.001	0.6	0.38	0.12	0.84
0.05	0.6	0.96	0.16	0.81
0.1	0.6	1.18	0.16	0.82

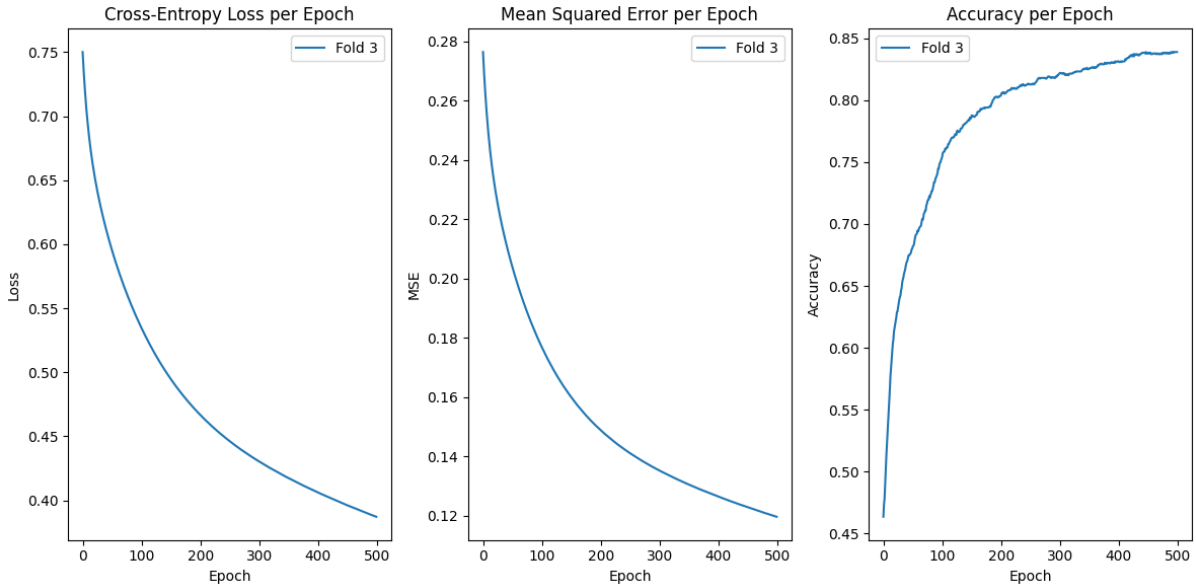
Η σταθερά ορμής δείχνει πόσο τα περασμένα gradients θα επηρεάσουν το παρόν βήμα. Αυξάνει την ταχύτητα μάθησης και μπορεί να οδηγήσει σε καλύτερα αποτελέσματα ωστόσο θέλουμε η ταχύτητα να μειώνεται ώστε το δίκτυο να μπορέσει να εντοπίσει και να σταθεροποιηθεί γύρω από την βέλτιστη λύση και για αυτό θέτουμε την σταθερά ορμής <1. Αν η σταθερά ορμής είναι >=1 αυτό σημαίνει ότι η ταχύτητα αντί να μειώνεται θα αυξάνεται, πράγμα που θα οδηγήσει στην ταλάντωση του δικτύου γύρω από τη λύση αντί για τη σταθεροποίησή του.

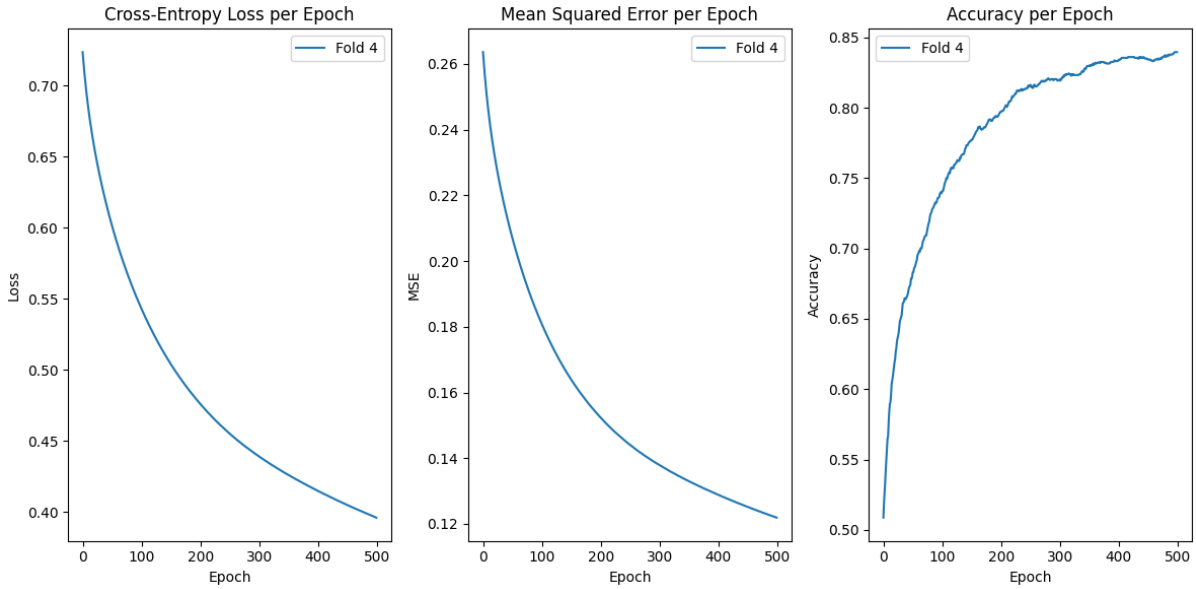
Όσον αφορά τις δοθείσες παραμέτρους περιμένουμε να έχουμε καλύτερα αποτελέσματα όταν δεν έχουμε ούτε πολύ μικρές ούτε πολύ μεγάλες τιμές για το ρυθμό μάθησης και τη σταθερά ορμής ώστε να μη σταθεροποιηθεί το δίκτυο πριν τη βέλτιστη λύση αλλά και να μη κάνει overshoot. Πράγματι, πειραματικά παρατηρούμε το μεγαλύτερο accuracy και το μικρότερο σφάλμα για η=0.001 και m=0.6 .

1)

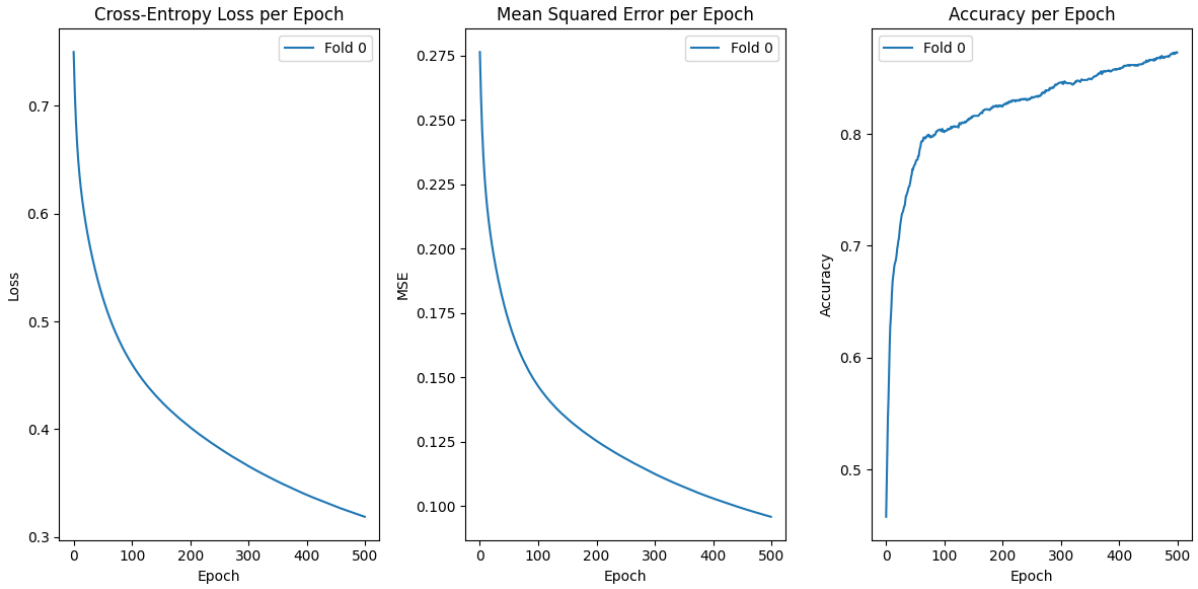


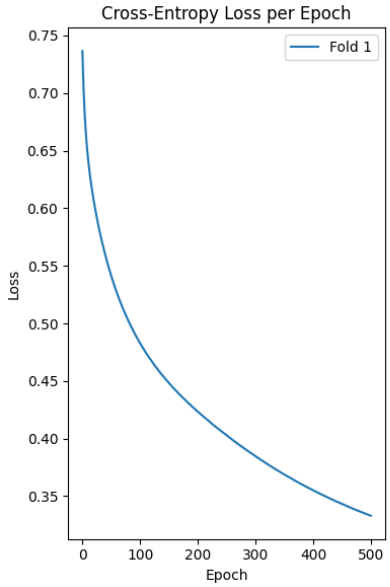

Cross-Entropy Loss per Epoch
Fold 1
Loss
Epoch
Mean Squared Error per Epoch
Fold 1
MSE
Epoch
Accuracy per Epoch
Fold 1
Accuracy
Epoch

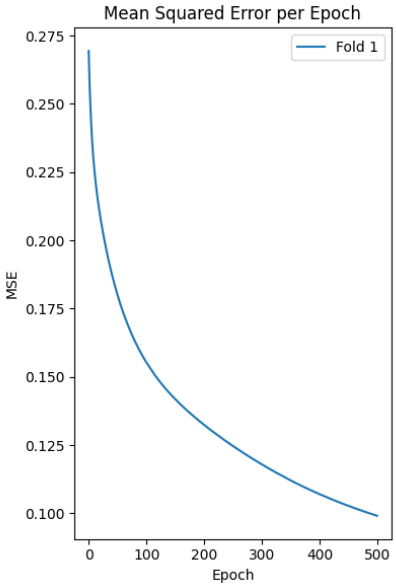

Cross-Entropy Loss per Epoch
Fold 3
Loss
Epoch
Mean Squared Error per Epoch
Fold 3
MSE
Epoch
Accuracy per Epoch
Fold 3
Accuracy
Epoch

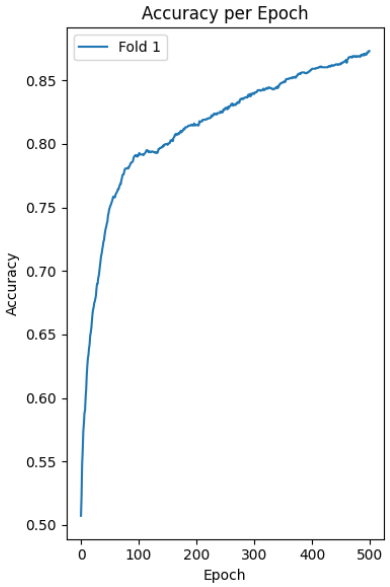

Cross-Entropy Loss per Epoch
Fold 4
Loss
Epoch
Mean Squared Error per Epoch
Fold 4
MSE
Epoch
Accuracy per Epoch
Fold 4
Accuracy
Epoch

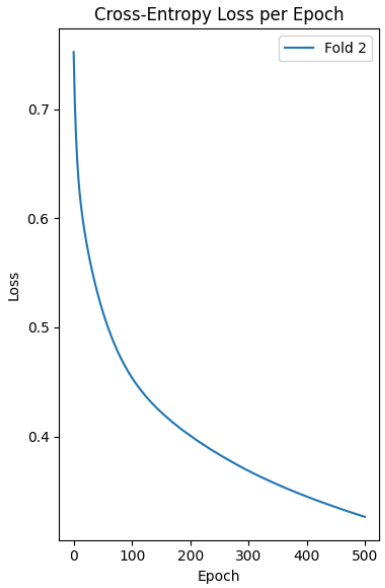
2)

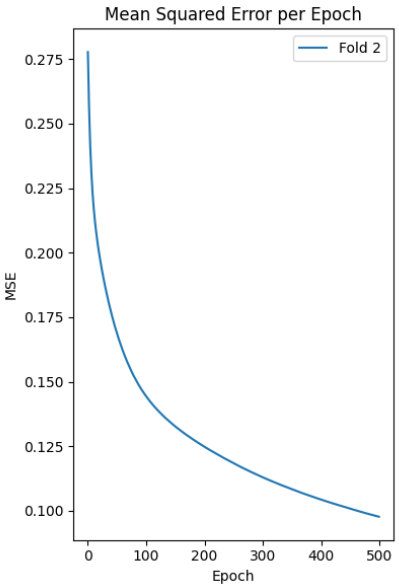

Cross-Entropy Loss per Epoch
Fold 0
Loss
Epoch
Mean Squared Error per Epoch
Fold 0
MSE
Epoch
Accuracy per Epoch
Fold 0
Accuracy
Epoch

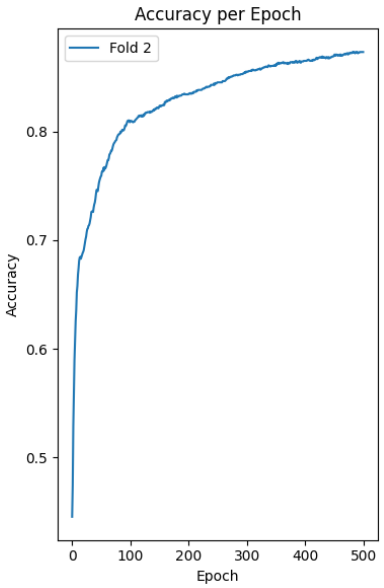

Cross-Entropy Loss per Epoch
Fold 1
Loss
Epoch

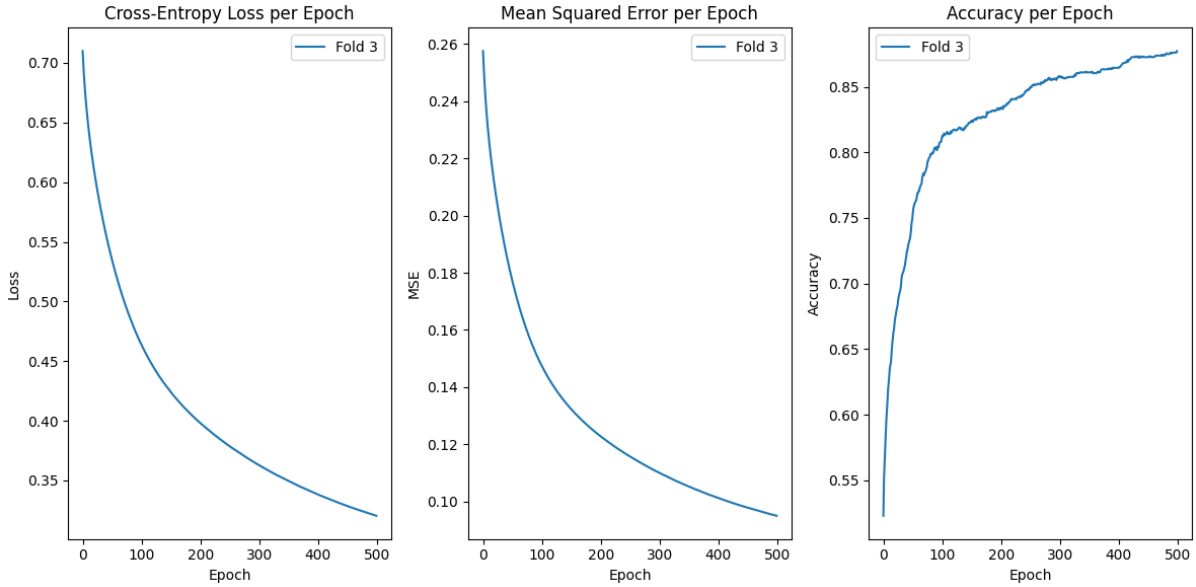

Mean Squared Error per Epoch
Fold 1
MSE
Epoch

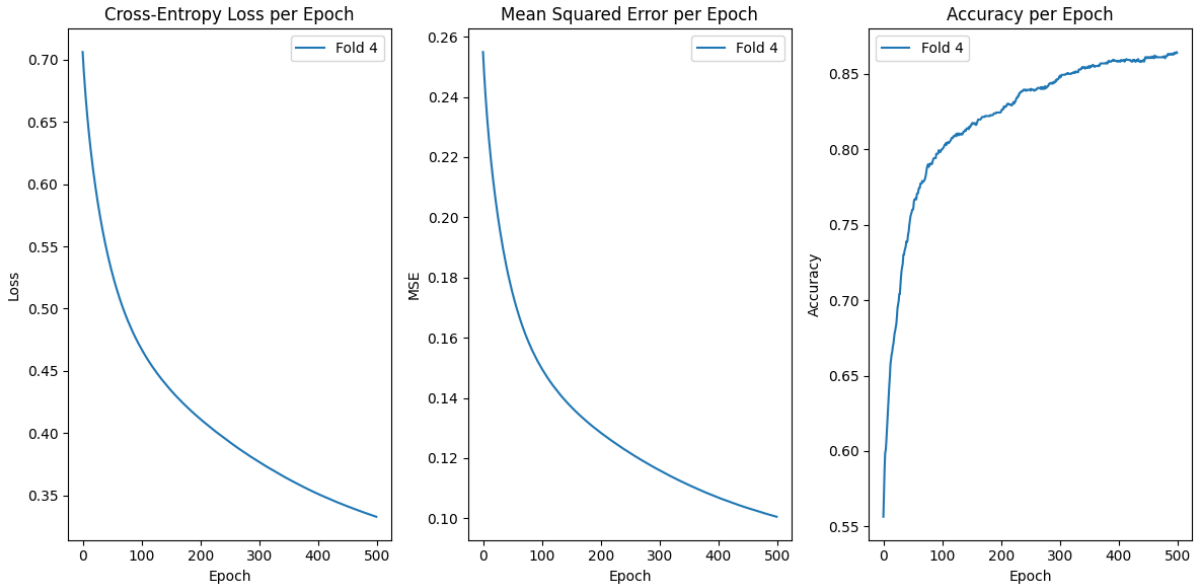

Accuracy per Epoch
Fold 1
Accuracy
Epoch


Cross-Entropy Loss per Epoch
Fold 2
Loss
Epoch

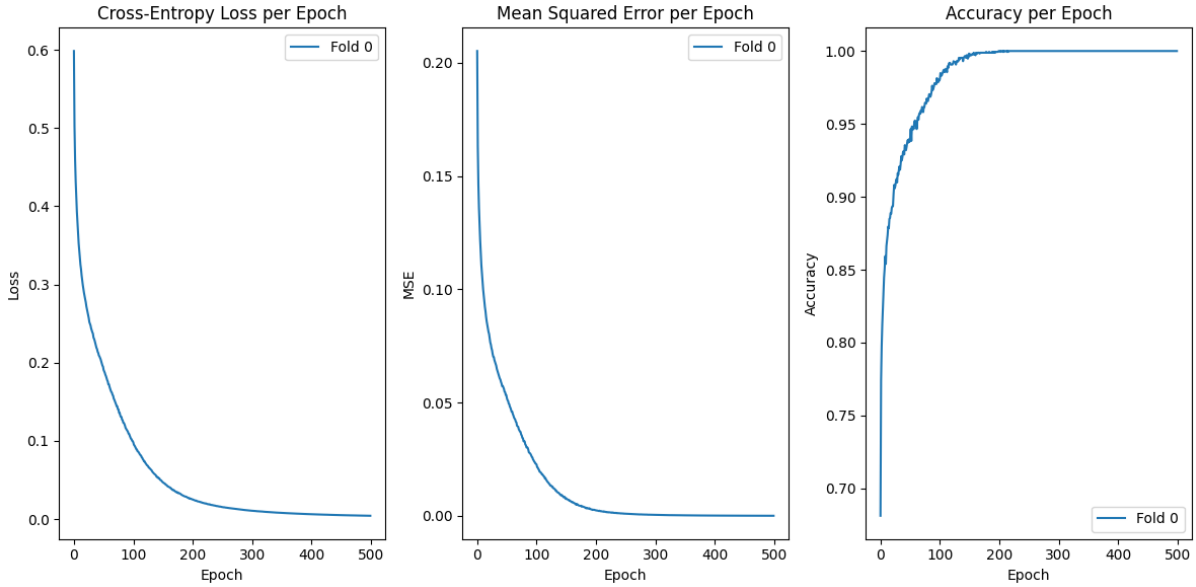

Mean Squared Error per Epoch
Fold 2
MSE
Epoch

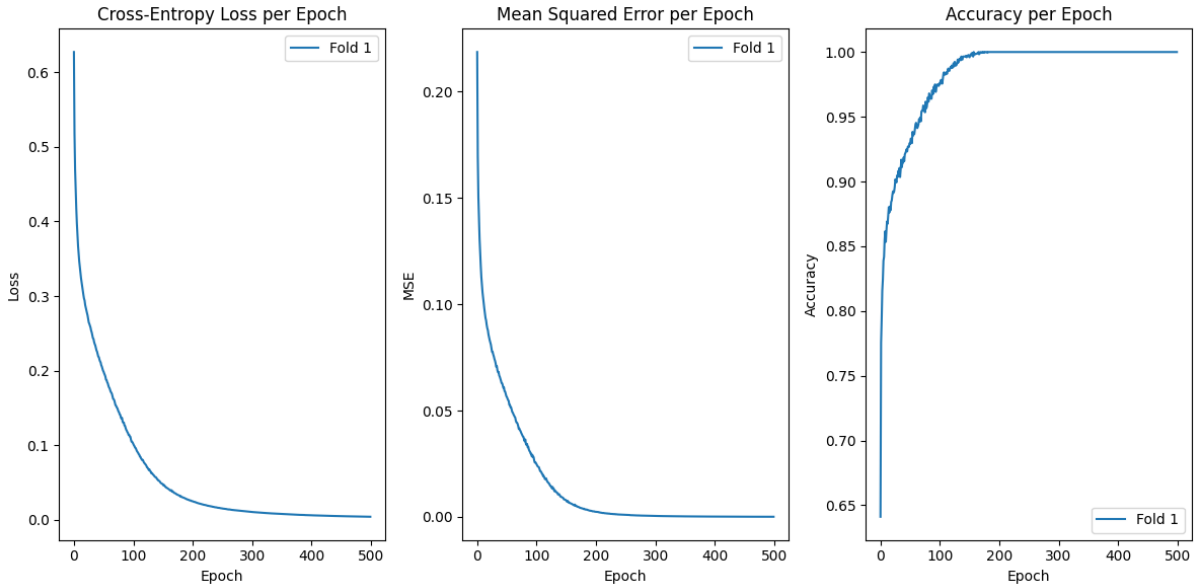

Accuracy per Epoch
Fold 2
Accuracy
Epoch

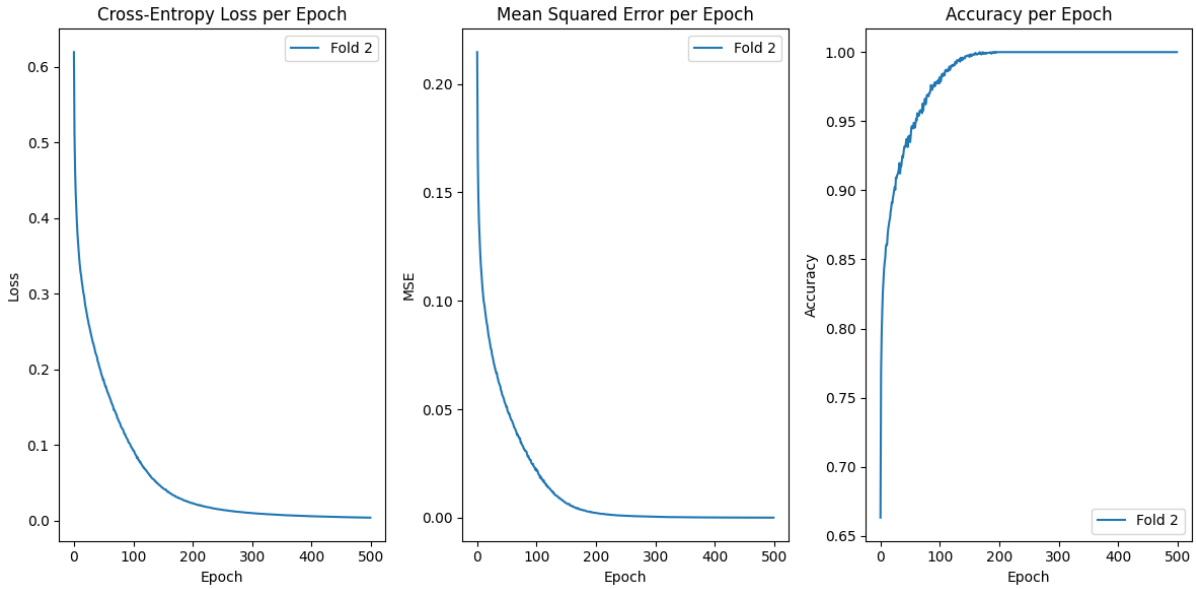

Cross-Entropy Loss per Epoch
Fold 3
Loss
Epoch
Mean Squared Error per Epoch
Fold 3
MSE
Epoch
Accuracy per Epoch
Fold 3
Accuracy
Epoch

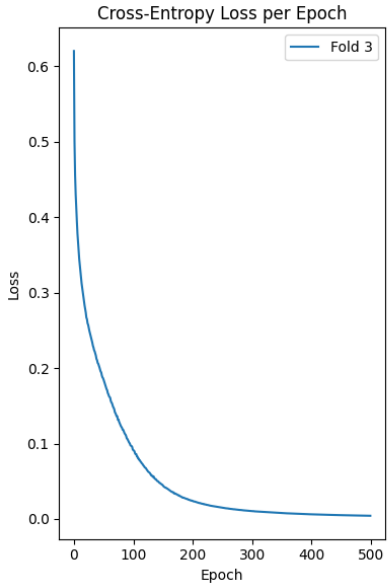

Cross-Entropy Loss per Epoch
Fold 4
Loss
Epoch
Mean Squared Error per Epoch
Fold 4
MSE
Epoch
Accuracy per Epoch
Fold 4
Accuracy
Epoch

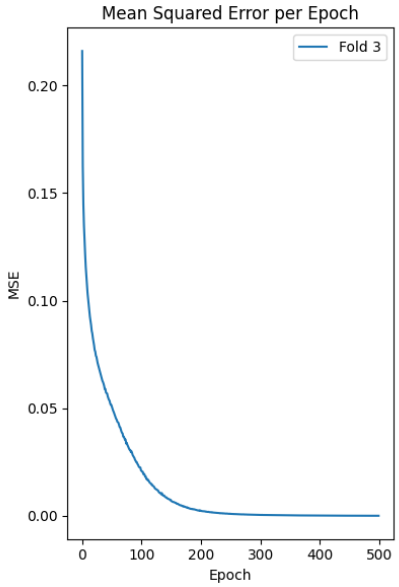
3)

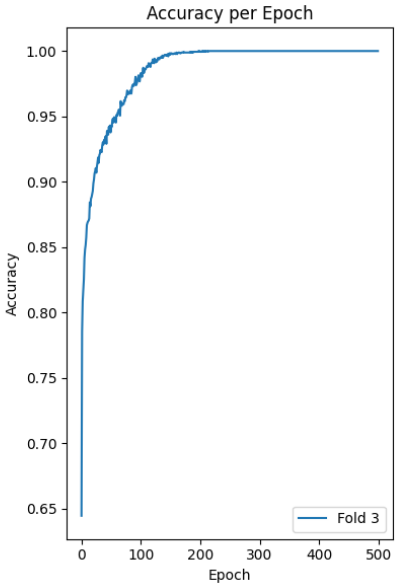

Cross-Entropy Loss per Epoch
Loss
Epoch
Fold 0
Mean Squared Error per Epoch
MSE
Epoch
Fold 0
Accuracy per Epoch
Accuracy
Epoch
Fold 0

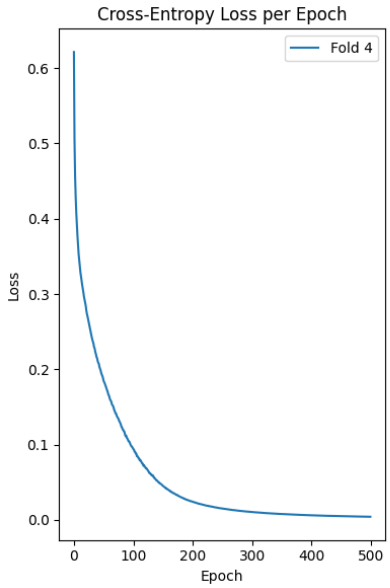

Cross-Entropy Loss per Epoch
Loss
Epoch
Fold 1
Mean Squared Error per Epoch
MSE
Epoch
Fold 1
Accuracy per Epoch
Accuracy
Epoch
Fold 1

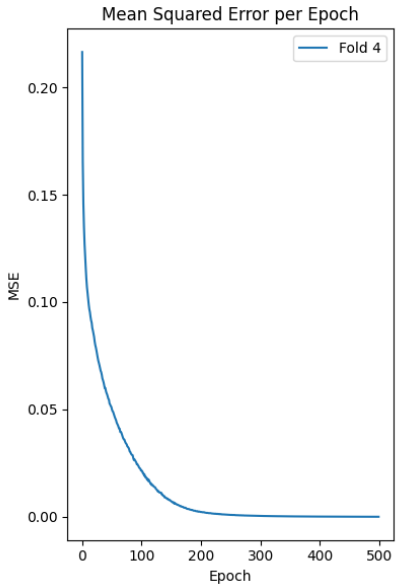

Cross-Entropy Loss per Epoch
Loss
Epoch
Fold 2
Mean Squared Error per Epoch
MSE
Epoch
Fold 2
Accuracy per Epoch
Accuracy
Epoch
Fold 2

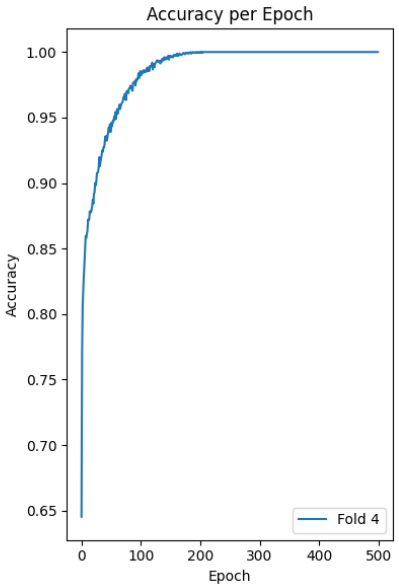

Cross-Entropy Loss per Epoch
Fold 3
Loss
Epoch


Mean Squared Error per Epoch
Fold 3
MSE
Epoch

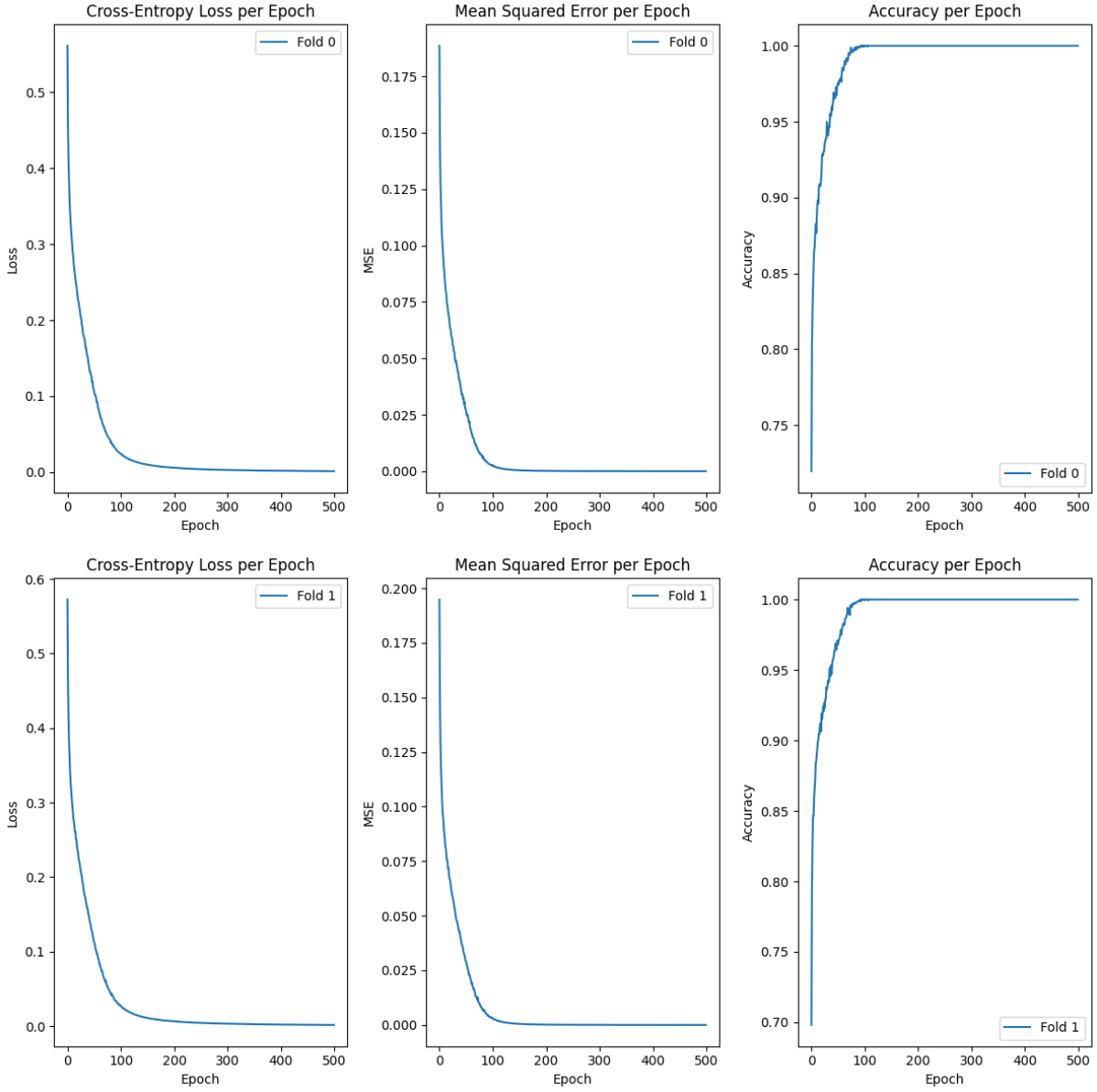

Accuracy per Epoch
Fold 3
Accuracy
Epoch

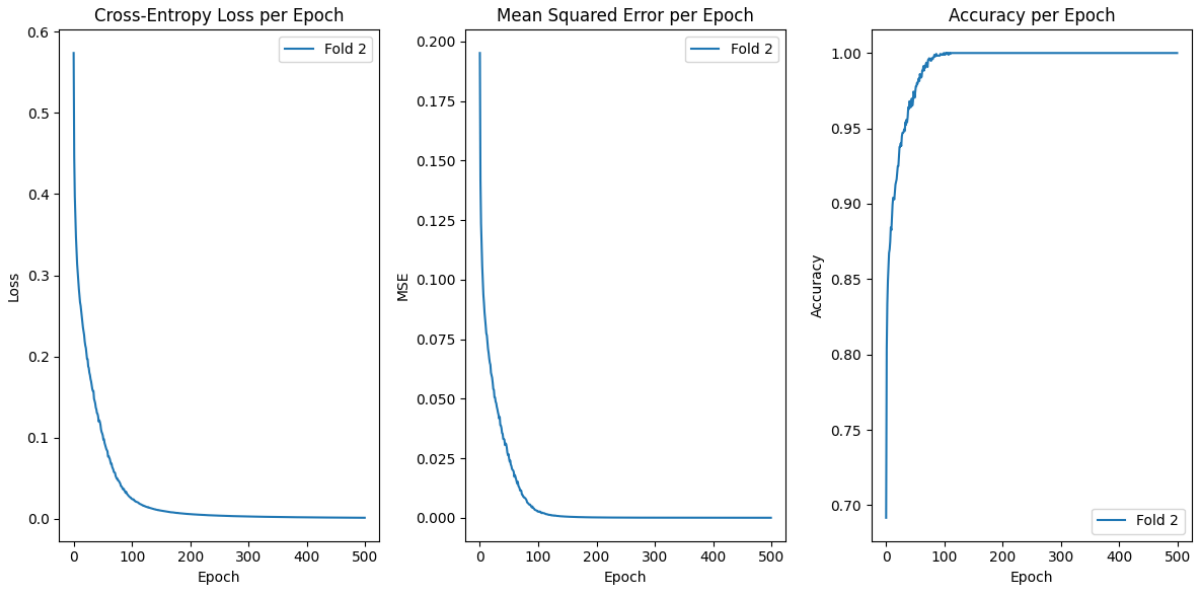

Cross-Entropy Loss per Epoch
Fold 4
Loss
Epoch

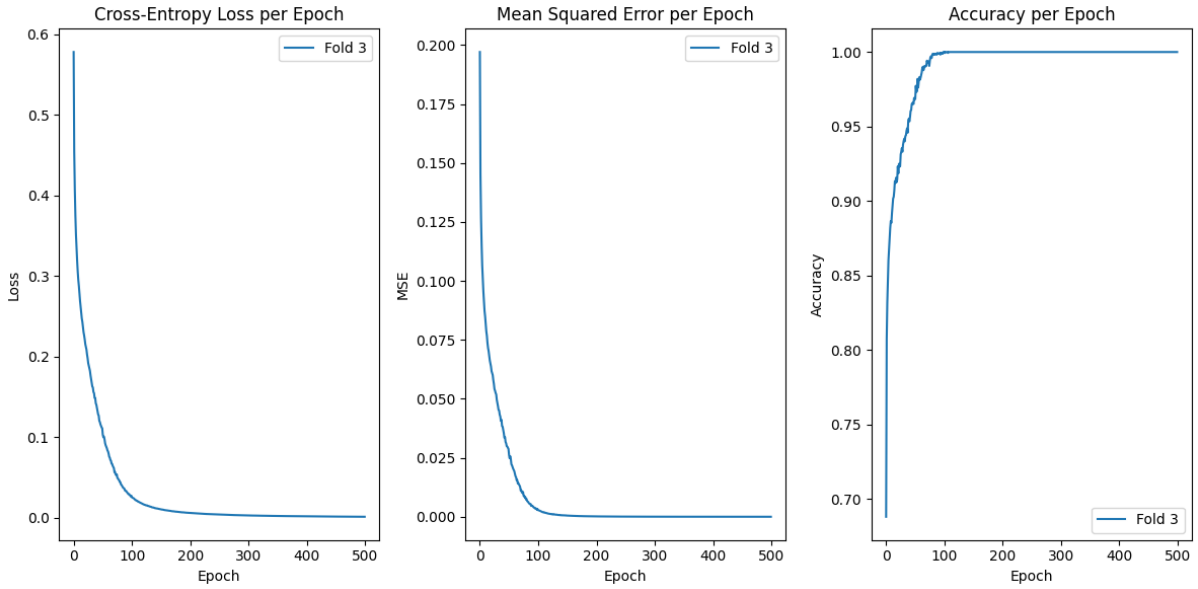

Mean Squared Error per Epoch
Fold 4
MSE
Epoch

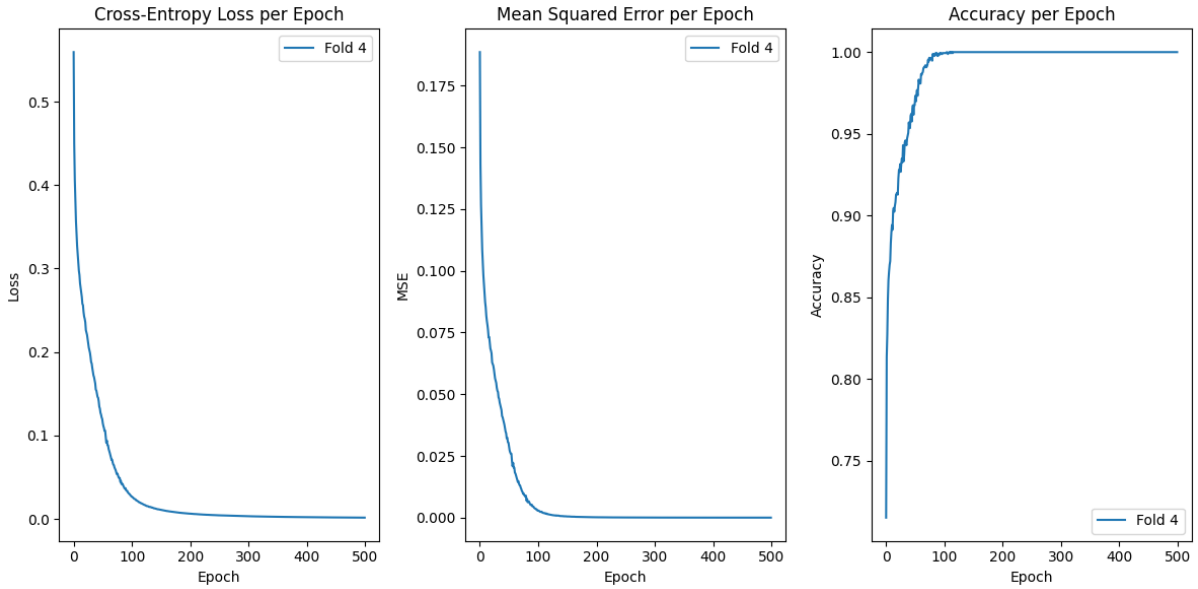

Accuracy per Epoch
Fold 4
Accuracy
Epoch

4)


Cross-Entropy Loss per Epoch
Fold 0
Loss
Epoch
Mean Squared Error per Epoch
Fold 0
MSE
Epoch
Accuracy per Epoch
Fold 0
Accuracy
Epoch
Cross-Entropy Loss per Epoch
Fold 1
Loss
Epoch
Mean Squared Error per Epoch
Fold 1
MSE
Epoch
Accuracy per Epoch
Fold 1
Accuracy
Epoch


Cross-Entropy Loss per Epoch
Loss
Epoch
Fold 2
Mean Squared Error per Epoch
MSE
Epoch
Fold 2
Accuracy per Epoch
Accuracy
Epoch
Fold 2


Cross-Entropy Loss per Epoch
Loss
Epoch
Fold 3
Mean Squared Error per Epoch
MSE
Epoch
Fold 3
Accuracy per Epoch
Accuracy
Epoch
Fold 3


Cross-Entropy Loss per Epoch
Loss
Epoch
Fold 4
Mean Squared Error per Epoch
MSE
Epoch
Fold 4
Accuracy per Epoch
Accuracy
Epoch
Fold 4

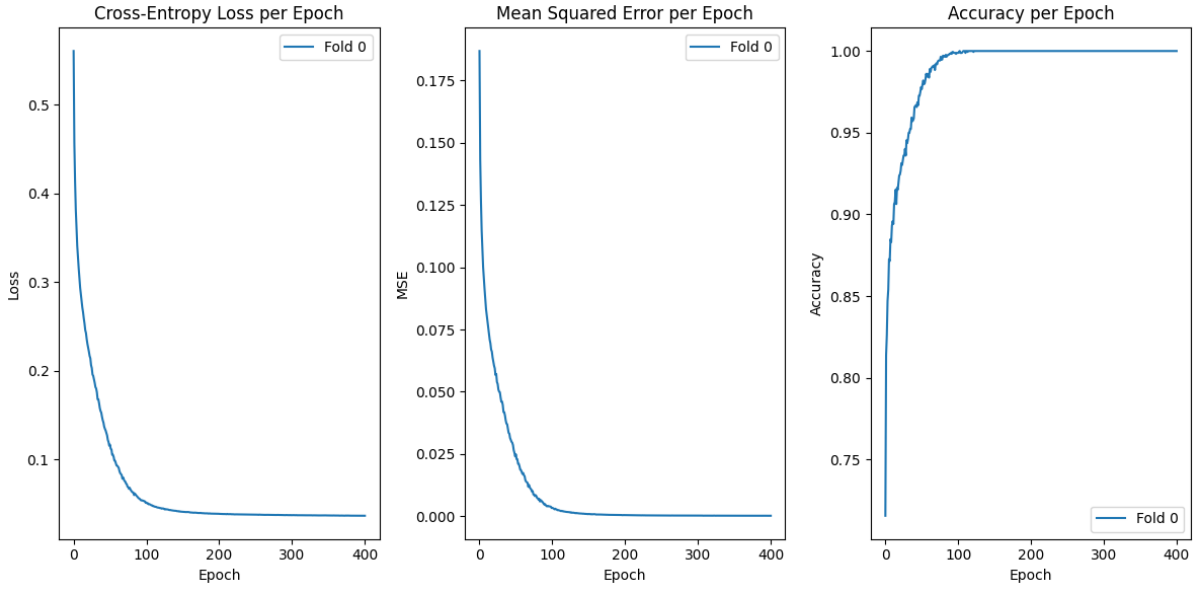
Α4. Ομαλοποίηση

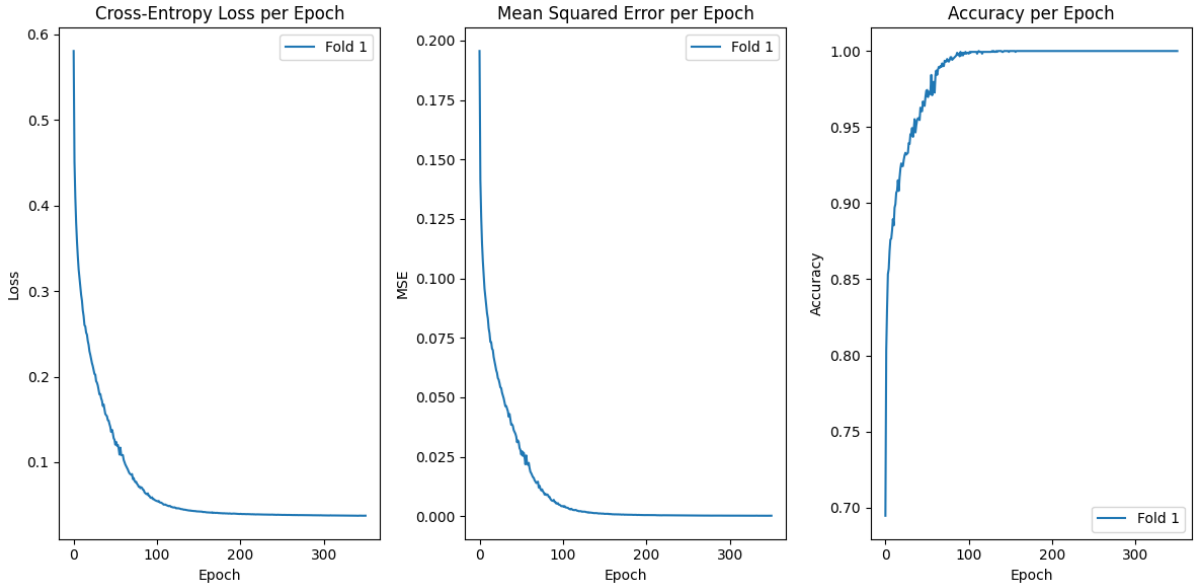
Συντελεστής r	**CE loss**	**MSE**	**Acc**
0.0001	0.89	0.15	0.82
0.001	0.58	0.13	0.83
0.01	0.45	0.11	0.85

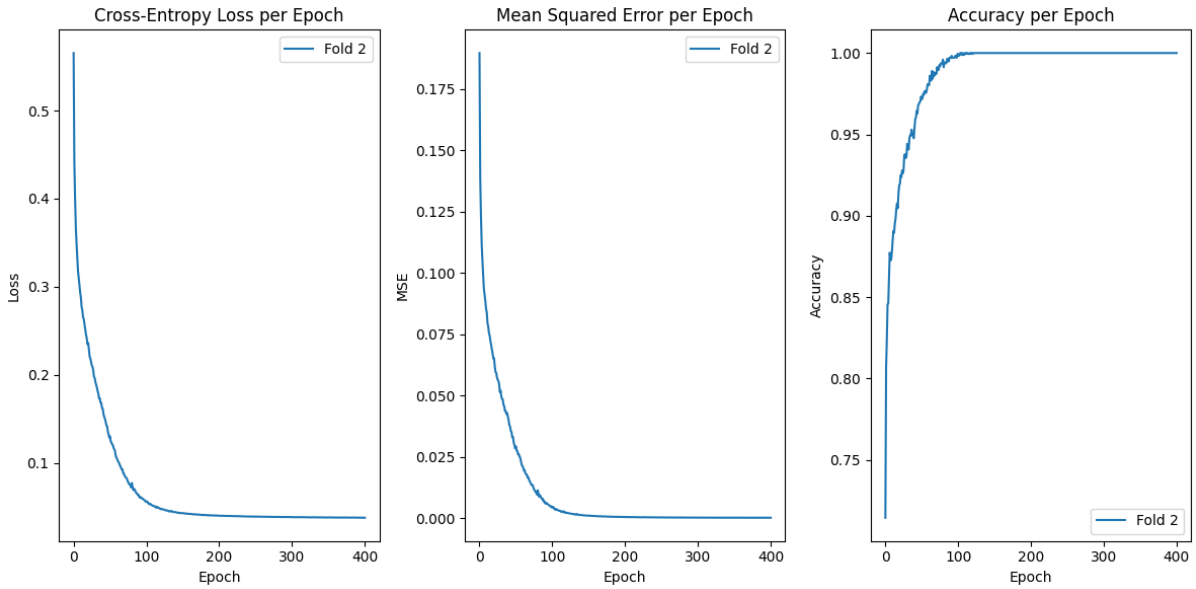
Η L1 νόρμα θα επιλέξει κάποια χαρακτηριστικά αυτόματα και τα υπόλοιπα θα τα θέσει ως αδιάφορα ορίζοντας τα βάρη στο μηδέν. Η L2 νόρμα καταφέρνει να αποφύγει και εκείνη το overfitting μειώνοντας τα μεγάλα βάρη αλλά διατηρώντας όλα τα χαρακτηριστικά, πράγμα που θα βελτιώσει την απόδοση του δικτύου σε άγνωστα δεδομένα επειδή δεν βασίζεται έντονα σε ορισμένα χαρακτηριστικά. Συνεπώς θα χρησιμοποιήσουμε την L2.

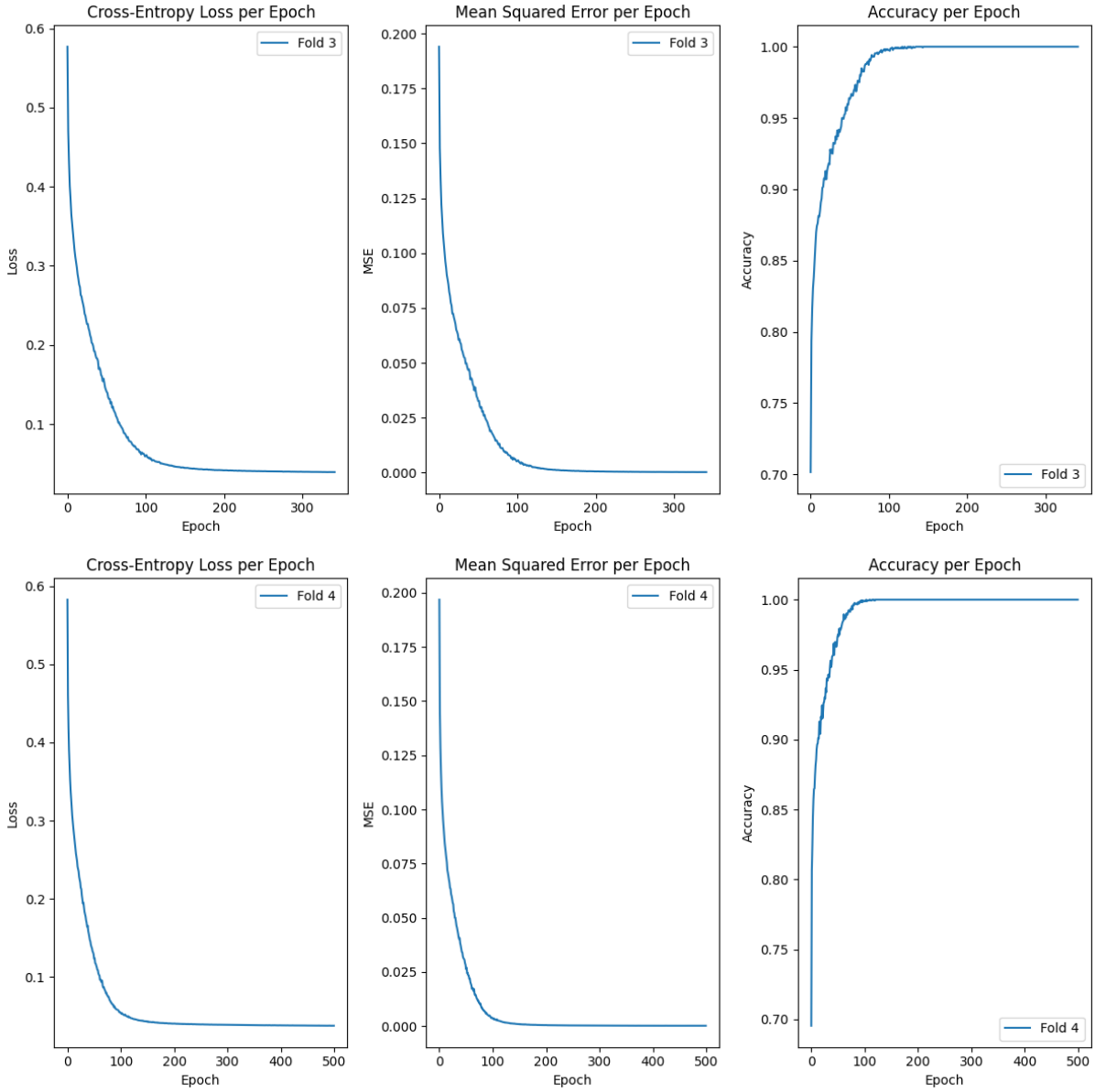
Για πολύ μικρή τιμή στο r δεν εκτελείται επαρκώς η γενίκευση, όσο μεγαλώνει η τιμή του βλέπουμε και καλύτερα αποτελέσματα στην απόδοση του δικτύου αφού εφαρμόζεται το penalty στα βάρη κρατώντας τα μικρά έχοντας έτσι πιο ομαλά gradients. Ωστόσο το r δεν πρέπει να είναι πολύ μεγάλο ώστε να μην προκληθεί underfitting στο δίκτυο.

1)

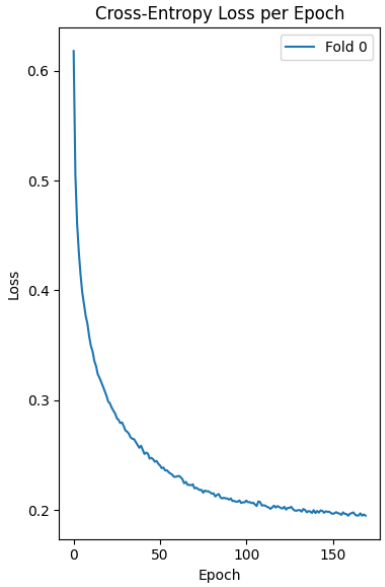

Cross-Entropy Loss per Epoch
Loss
Epoch
Fold 0
Mean Squared Error per Epoch
MSE
Epoch
Fold 0
Accuracy per Epoch
Accuracy
Epoch
Fold 0

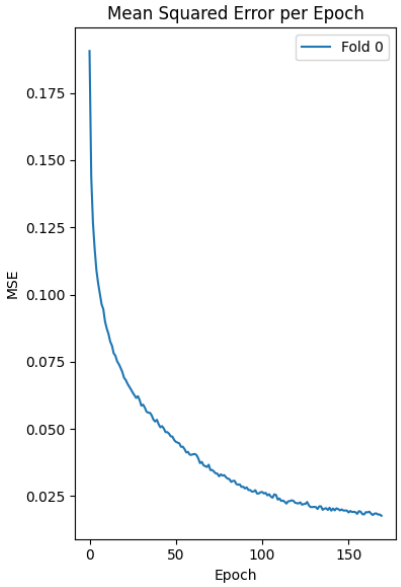

Cross-Entropy Loss per Epoch
Loss
Epoch
Fold 1
Mean Squared Error per Epoch
MSE
Epoch
Fold 1
Accuracy per Epoch
Accuracy
Epoch
Fold 1

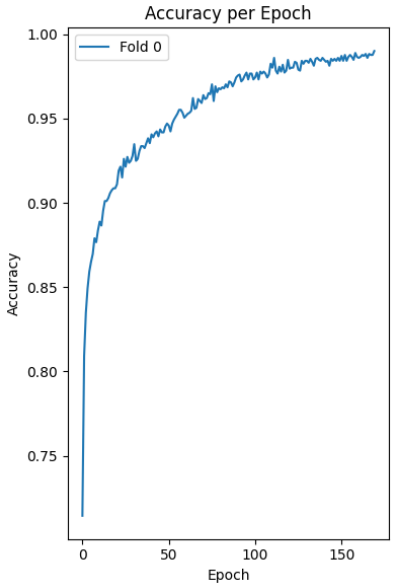

Cross-Entropy Loss per Epoch
Loss
Epoch
Fold 2
Mean Squared Error per Epoch
MSE
Epoch
Fold 2
Accuracy per Epoch
Accuracy
Epoch
Fold 2

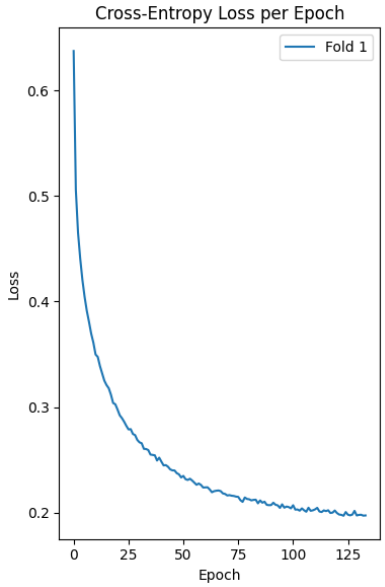

Cross-Entropy Loss per Epoch
Fold 3
Loss
Epoch
Mean Squared Error per Epoch
Fold 3
MSE
Epoch
Accuracy per Epoch
Fold 3
Accuracy
Epoch
Cross-Entropy Loss per Epoch
Fold 4
Loss
Epoch
Mean Squared Error per Epoch
Fold 4
MSE
Epoch
Accuracy per Epoch
Fold 4
Accuracy
Epoch

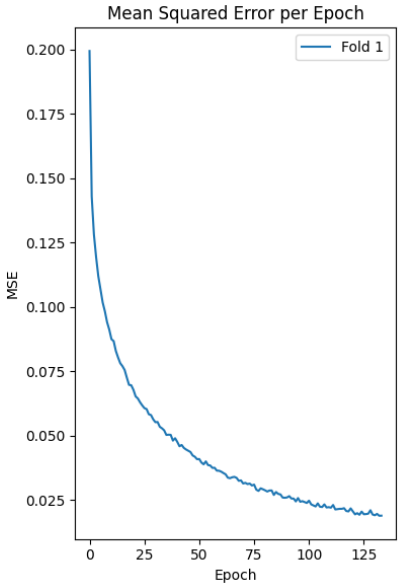
2)

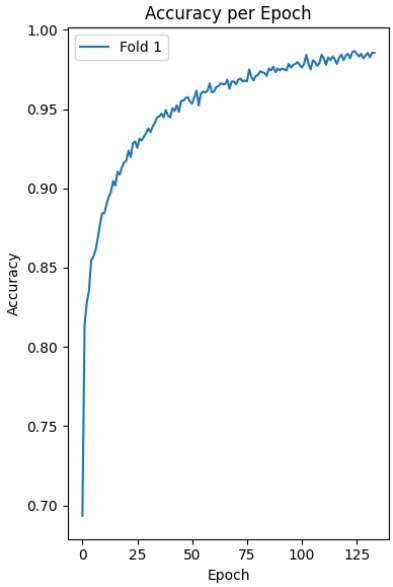

Cross-Entropy Loss per Epoch
Fold 0
Loss
Epoch

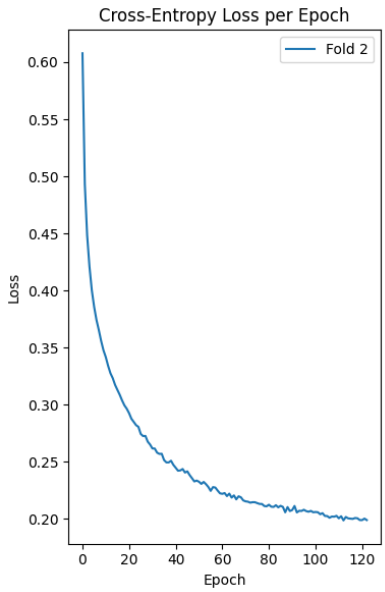

Mean Squared Error per Epoch
Fold 0
MSE
Epoch

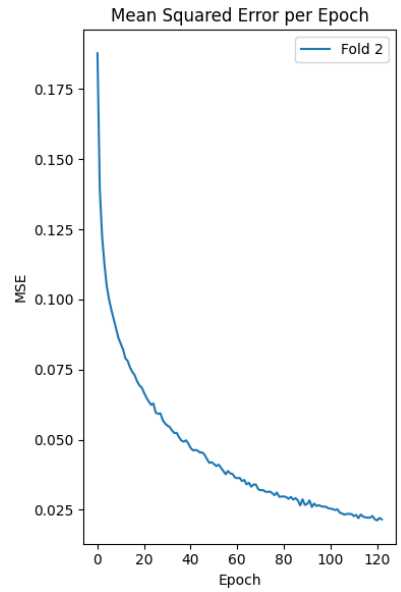

Accuracy per Epoch
Fold 0
Accuracy
Epoch

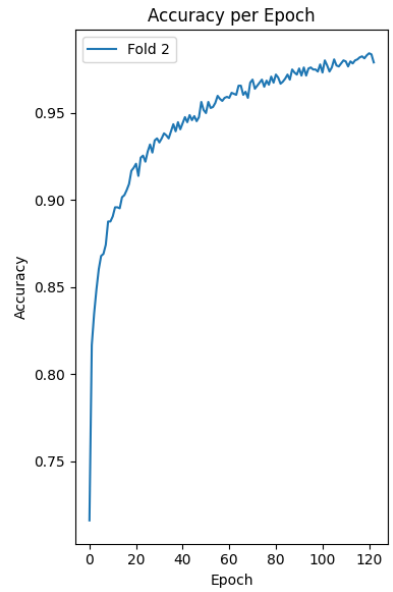

Cross-Entropy Loss per Epoch
Fold 1
Loss
Epoch

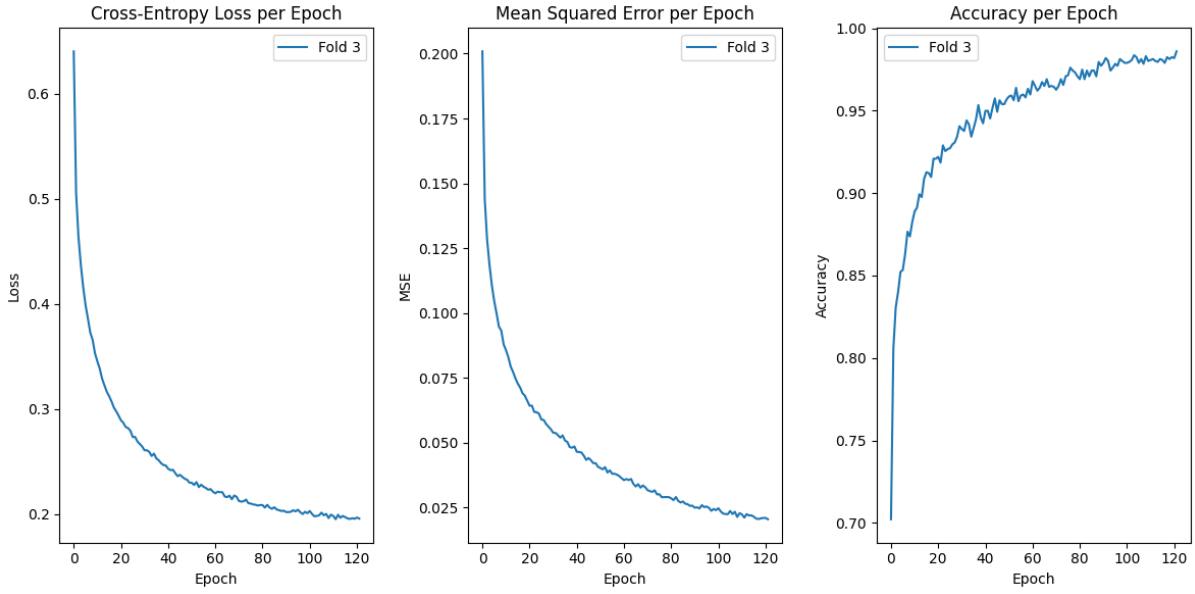

Mean Squared Error per Epoch
Fold 1
MSE
Epoch

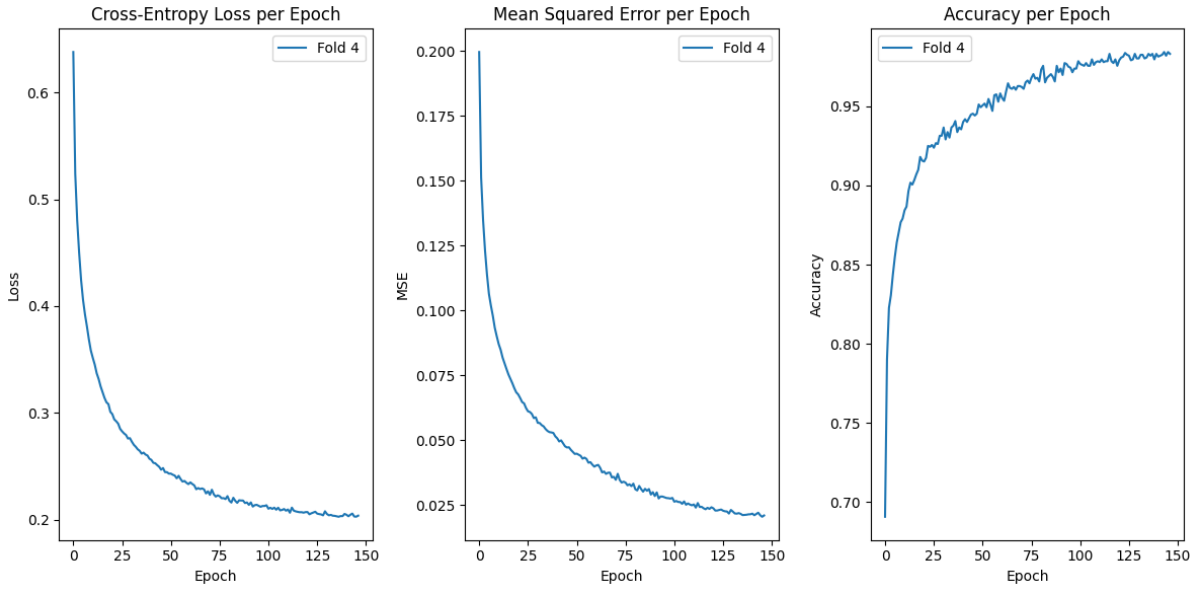

Accuracy per Epoch
Fold 1
Accuracy
Epoch


Cross-Entropy Loss per Epoch
Fold 2
Loss
Epoch

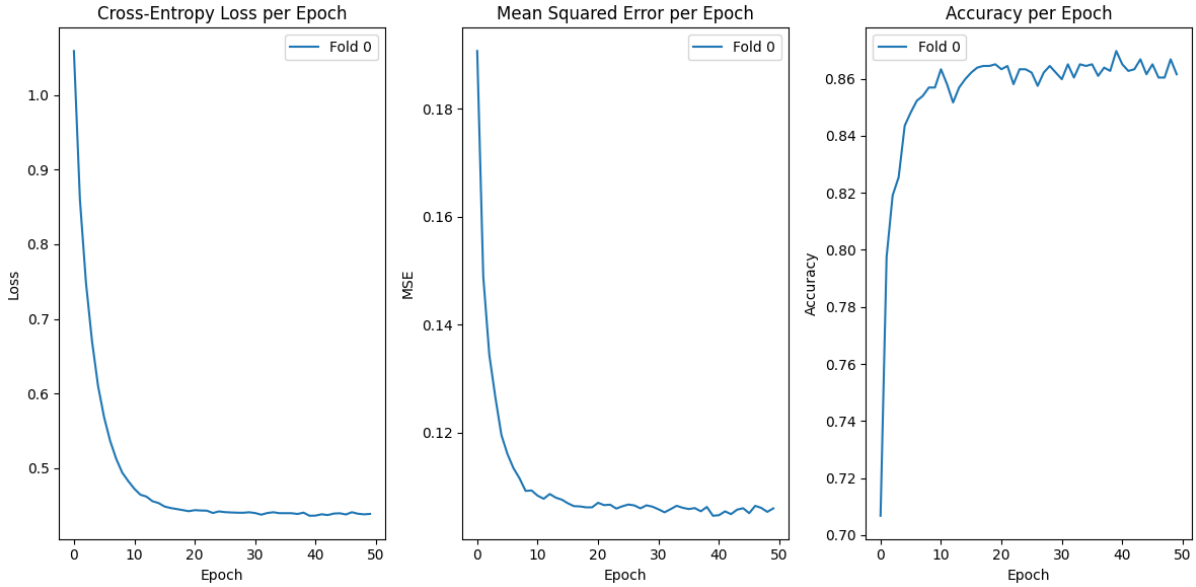

Mean Squared Error per Epoch
Fold 2
MSE
Epoch

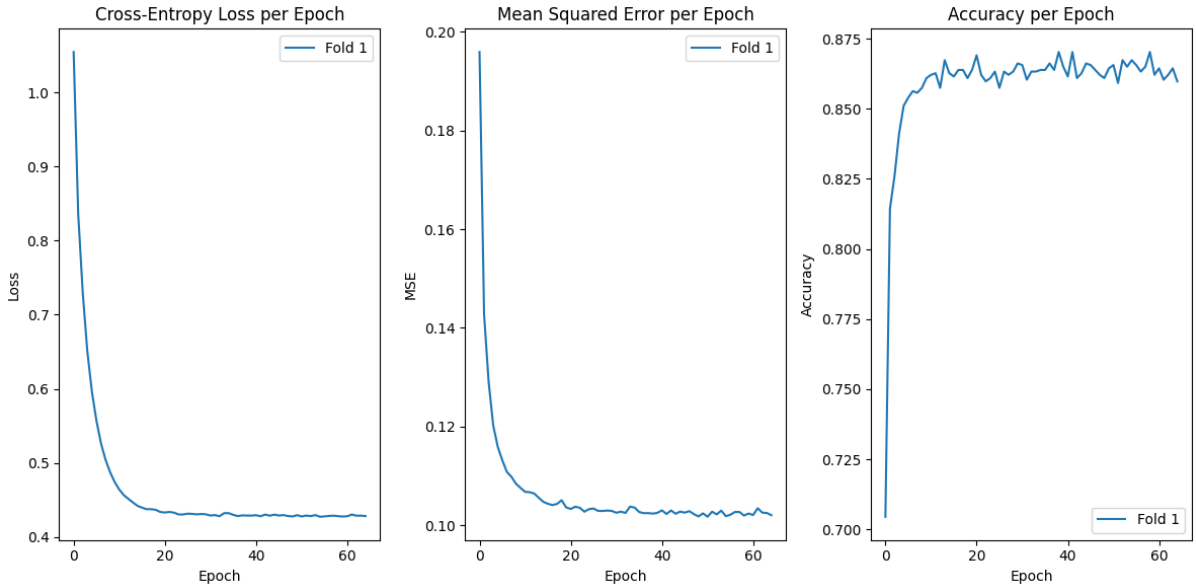

Accuracy per Epoch
Fold 2
Accuracy
Epoch

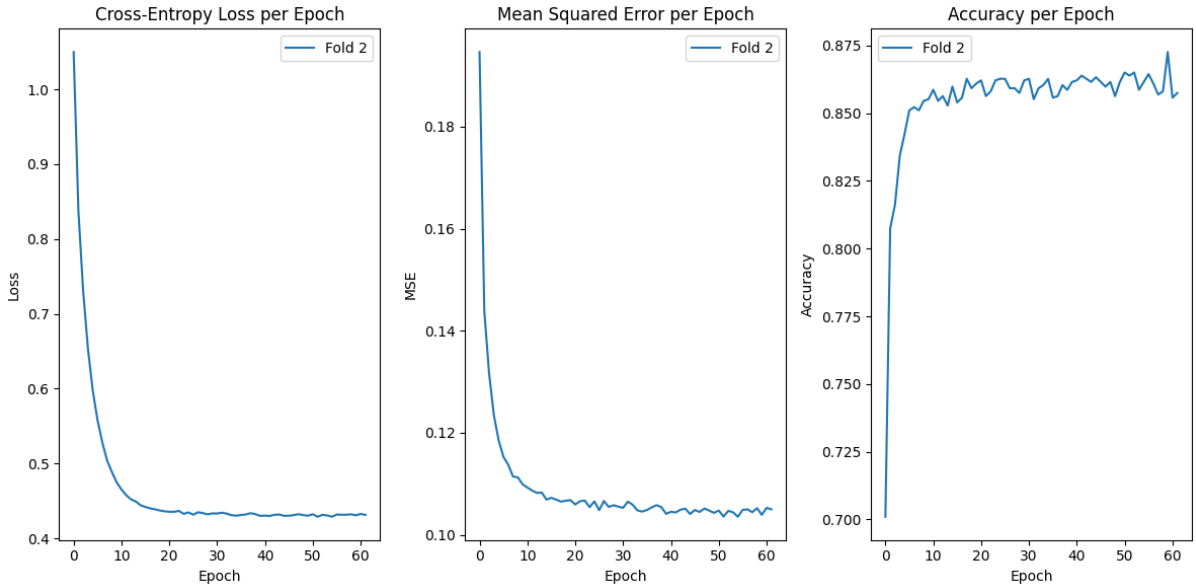

Cross-Entropy Loss per Epoch
Fold 3
Loss
Epoch
Mean Squared Error per Epoch
Fold 3
MSE
Epoch
Accuracy per Epoch
Fold 3
Accuracy
Epoch

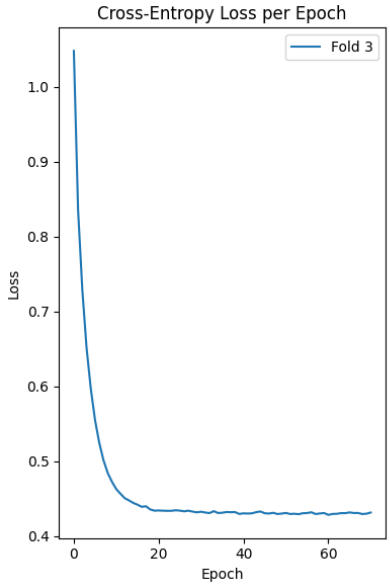

Cross-Entropy Loss per Epoch
Fold 4
Loss
Epoch
Mean Squared Error per Epoch
Fold 4
MSE
Epoch
Accuracy per Epoch
Fold 4
Accuracy
Epoch

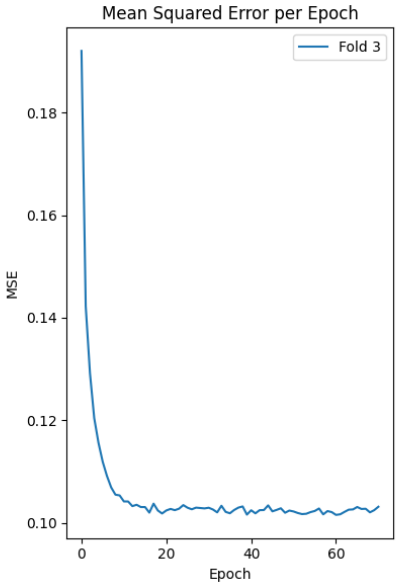
3)

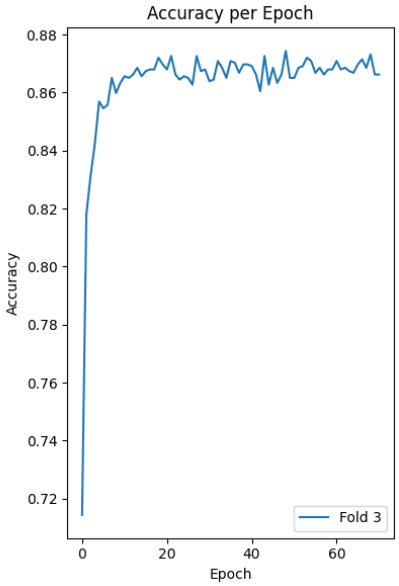

Cross-Entropy Loss per Epoch
Fold 0
Loss
Epoch
Mean Squared Error per Epoch
Fold 0
MSE
Epoch
Accuracy per Epoch
Fold 0
Accuracy
Epoch

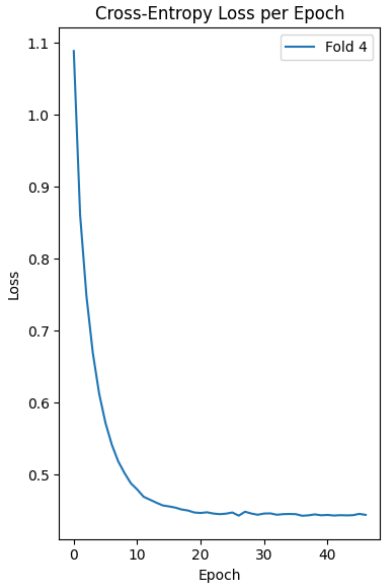

Cross-Entropy Loss per Epoch
Fold 1
Loss
Epoch
Mean Squared Error per Epoch
Fold 1
MSE
Epoch
Accuracy per Epoch
Fold 1
Accuracy
Epoch

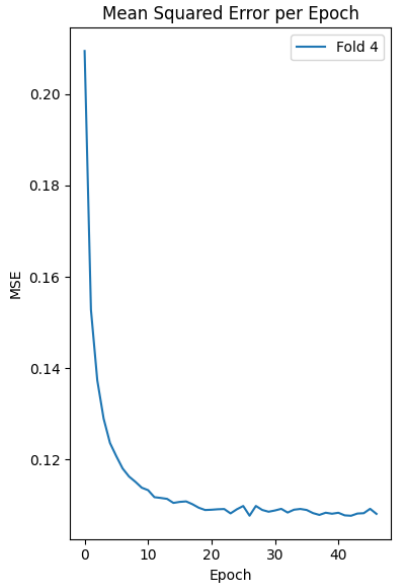

Cross-Entropy Loss per Epoch
Fold 2
Loss
Epoch
Mean Squared Error per Epoch
Fold 2
MSE
Epoch
Accuracy per Epoch
Fold 2
Accuracy
Epoch

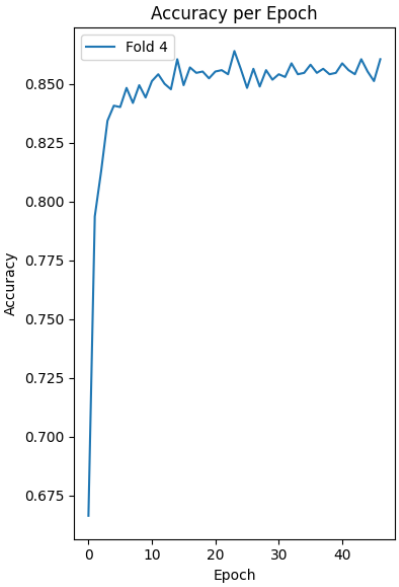

Cross-Entropy Loss per Epoch
Fold 3
Loss
Epoch


Mean Squared Error per Epoch
Fold 3
MSE
Epoch


Accuracy per Epoch
Fold 3
Accuracy
Epoch


Cross-Entropy Loss per Epoch
Fold 4
Loss
Epoch


Mean Squared Error per Epoch
Fold 4
MSE
Epoch


Accuracy per Epoch
Fold 4
Accuracy
Epoch

Α5. Βαθύ Νευρωνικό Δίκτυο

Αριθμός νευρώνων στα κρυφά επίπεδα	**CE loss**	**MSE**	**Acc**
78 | [78*(2/3)] =52	0.47	0.10	0.85
78 | 78	0.47	0.11	0.85
78 | [78*2] = 156	0.47	0.10	0.86

Μετά από έρευνα φτάνω στο συμπέρασμα ότι δεν υπάρχει κάποια προκαθορισμένη διαδικασία για τον καθορισμό του αριθμού των κρυφών επιπέδων ή του πλήθους νευρώνων που περιέχουν. Ωστόσο χρησιμοποιώντας το Rule of the thumb κρίνω πώς τα 2 κρυφά επίπεδα είναι αρκετά για την πολυπλοκότητα του προβλήματος της συγκεκριμένης άσκησης και ότι παραπάνω από 2 θα οδηγήσουν το δίκτυο σε overfitting και θα καθυστερήσουν την μάθηση. Όσον αφορά το πλήθος το νευρώνων, υποθέτω πως ο μειούμενος αριθμός θα παράγει καλύτερα αποτελέσματα καθώς θα γενικεύει την έξοδο του προηγούμενου επιπέδου. Πάνω σε αυτή την υπόθεση πειραματίζομαι με την ευρετική των 2/3 της εισόδου του πρώτου επιπέδου, όπου στο πρώτο επίπεδο ο αριθμός νευρώνων είναι εκείνος που παρείχε το καλύτερο αποτέλεσμα. Παρόλα αυτά ερευνώ και τις άλλες δύο πιθανότητες, των ίσων νευρώνων αλλά και των διπλάσιων. Η λοιπή αρχιτεκτονική του δικτύου παραμένει ίδια με τα προηγούμενα ερωτήματα.

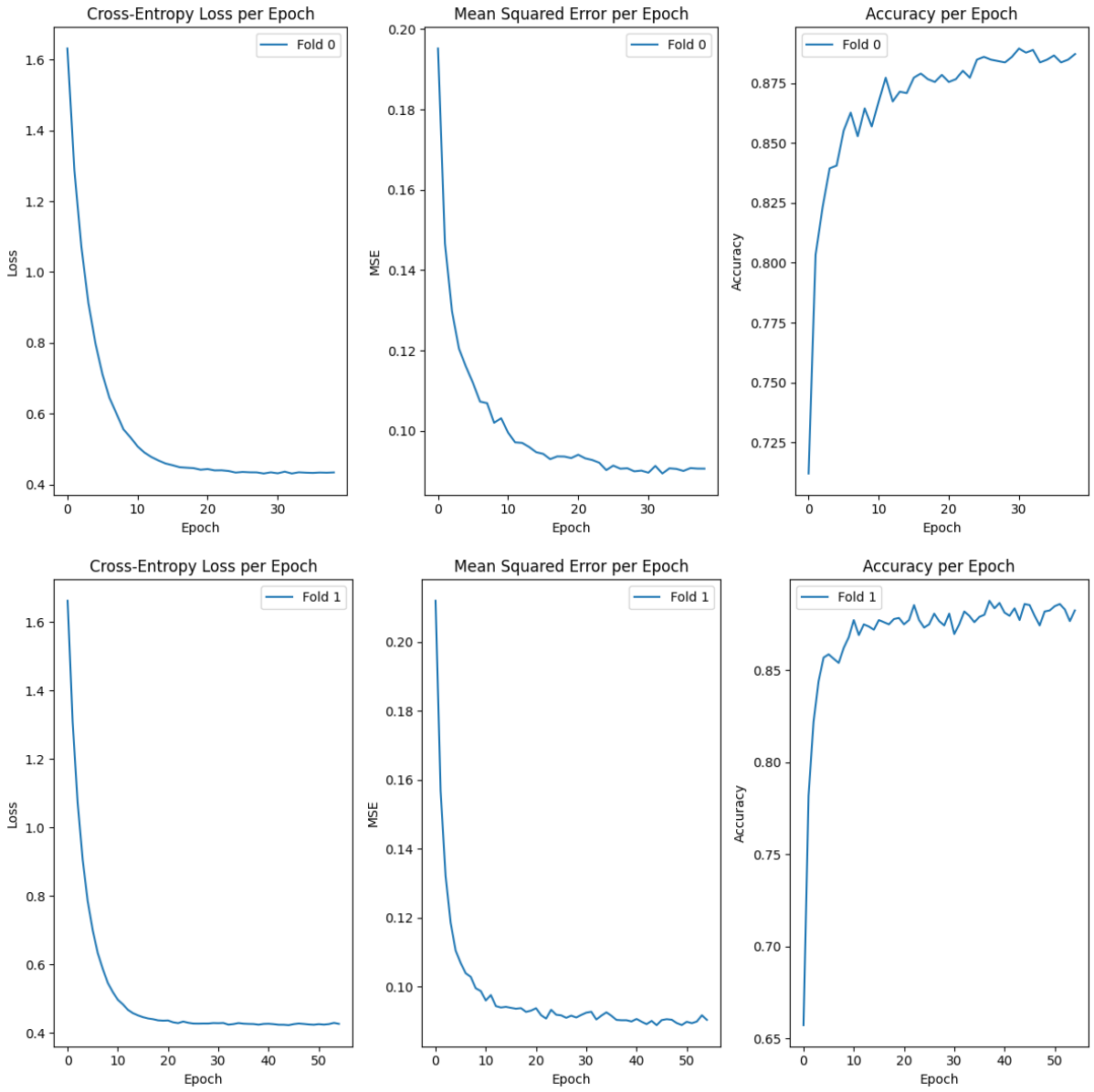
Οι μικρές αλλαγές στα πειραματικά αποτελέσματα δεν επιβεβαιώνουν την αρχική υπόθεση ωστόσο δείχνουν ότι το δίκτυο έχει ήδη αρκετό capacity και δεν χρειάζεται περαιτέρω αύξηση για το συγκεκριμένο πρόβλημα, ο πειραματισμός και με τρίτο κρυφό επίπεδο έφερε ίδιο συμπέρασμα. Η μείωση περαιτέρω των νευρώνων του δεύτερου επιπέδου οδήγησε σε underfitting.

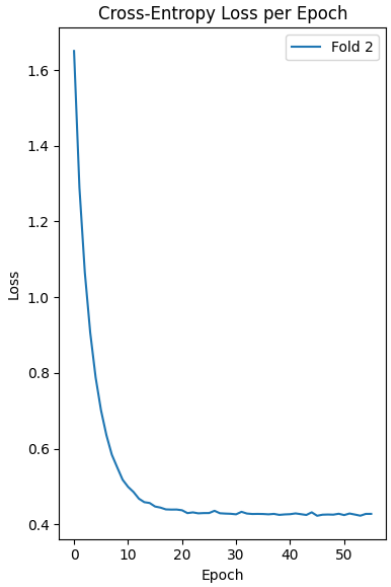
Συνεπώς αξιολογώντας τα νέα δεδομένα κρίνω πως η προσθήκη δεύτερου κρυφού επιπέδου βελτίωσε την απόδοση του δικτύου έναντι της απόδοσης από το ερώτημα Α2, ωστόσο το δίκτυο φαίνεται να επωφελείται ελάχιστα από την αρχιτεκτονική της διπλασίασης της εισόδου, χωρίς αυτό να είναι απόλυτα προφανές καθώς το πρόβλημα ίσως δεν είναι αρκετά περίπλοκο.

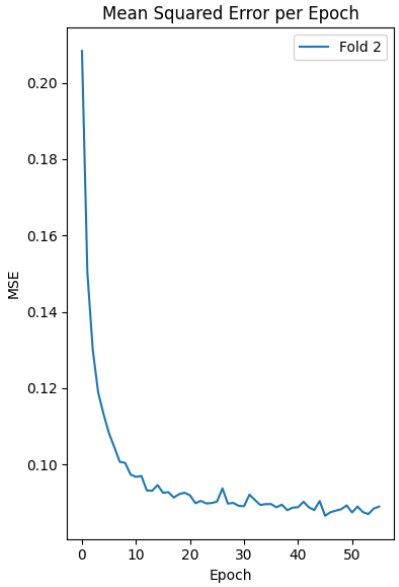
- https://medium.com/biased-algorithms/how-to-create-hidden-layers-in-neural-networks-31772414445f

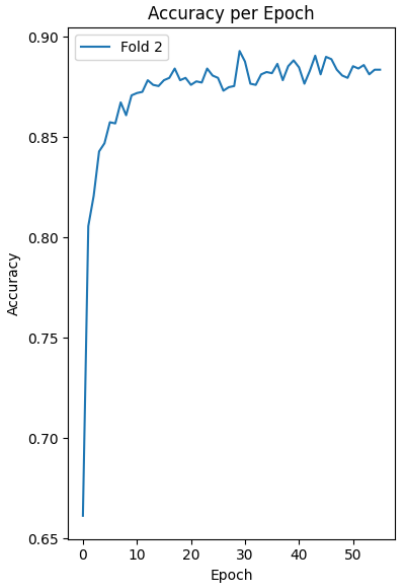
- https://medium.com/biased-algorithms/how-to-create-hidden-layers-in-neural-networks-31772414445f
- https://mljourney.com/how-to-decide-the-number-of-hidden-layers-in-a-neural-network/
- https://machinelearningmastery.com/how-to-configure-the-number-of-layers-and-nodes-in-a-neural-network/
- https://www.youtube.com/watch?v=bqBRET7tbiQ

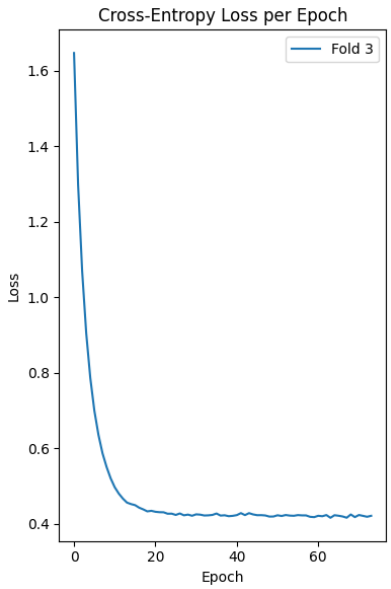
1)

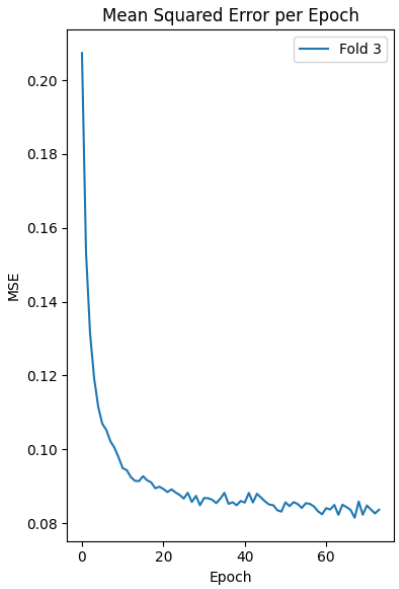


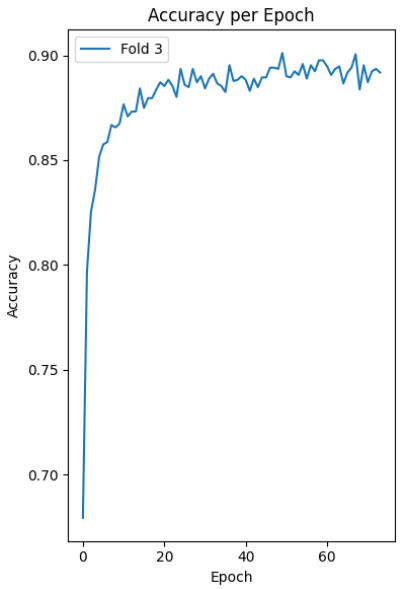

Cross-Entropy Loss per Epoch
Fold 2
Loss
Epoch

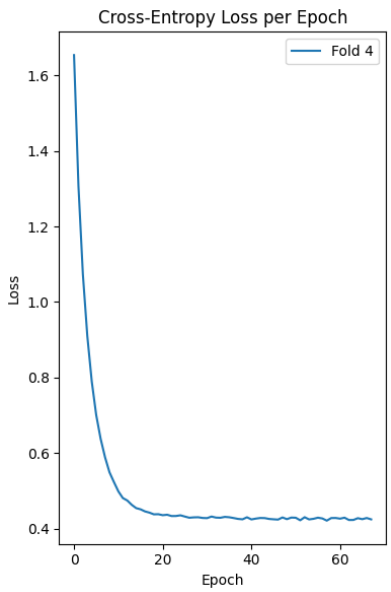

Mean Squared Error per Epoch
Fold 2
MSE
Epoch

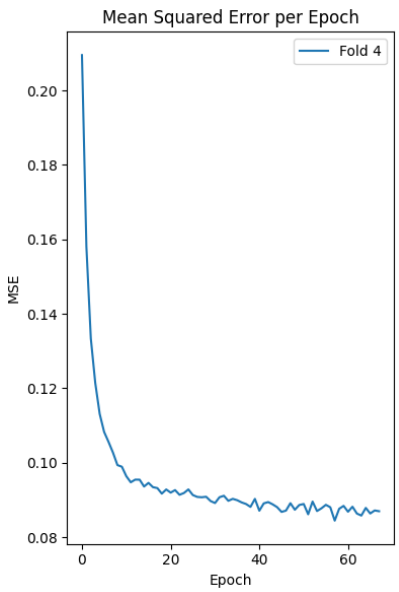

Accuracy per Epoch
Fold 2
Accuracy
Epoch

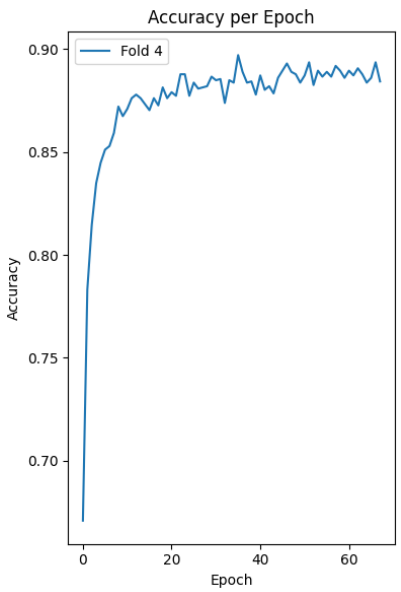

Cross-Entropy Loss per Epoch
Fold 3
Loss
Epoch


Mean Squared Error per Epoch
Fold 3
MSE
Epoch

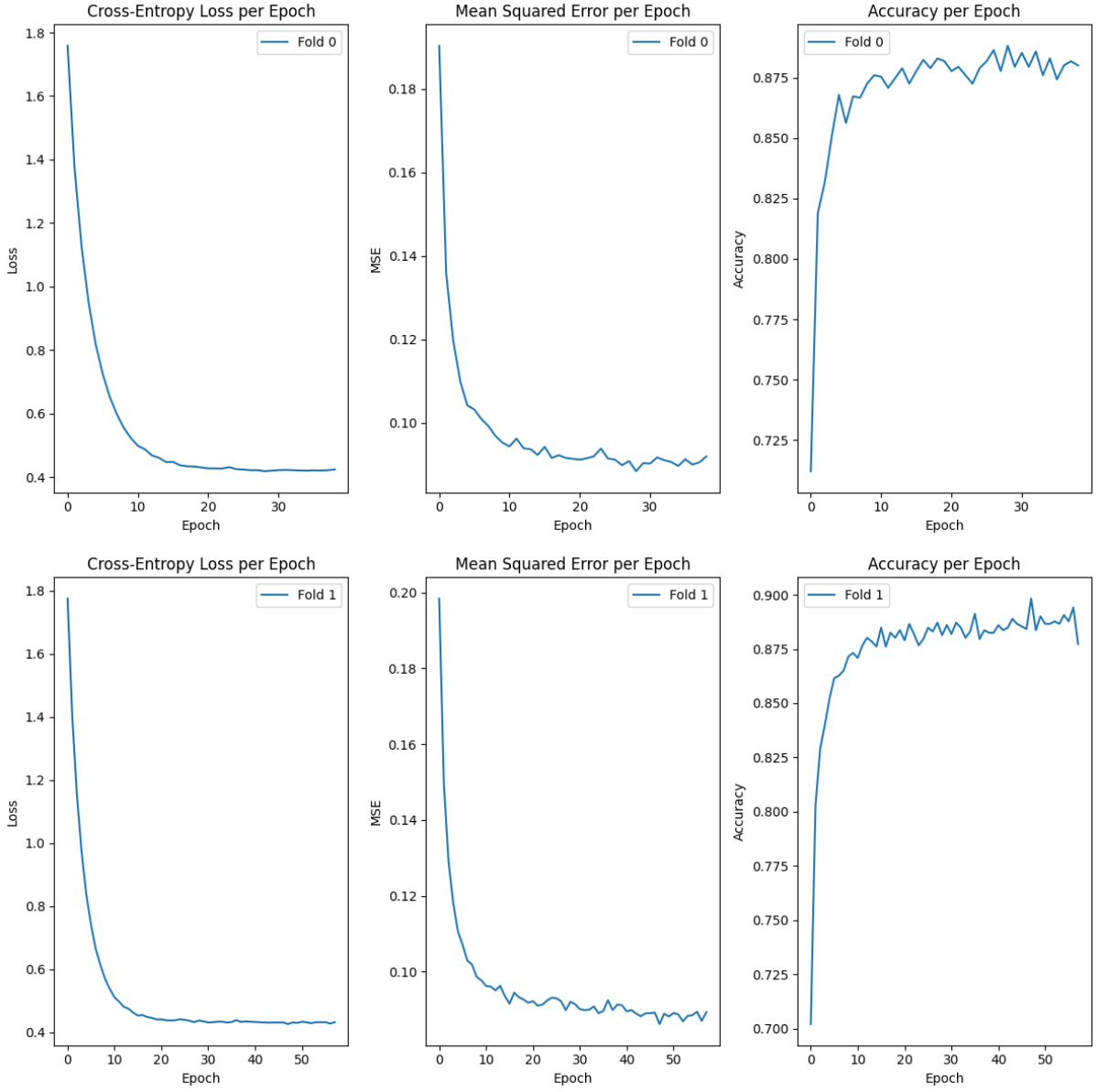

Accuracy per Epoch
Fold 3
Accuracy
Epoch

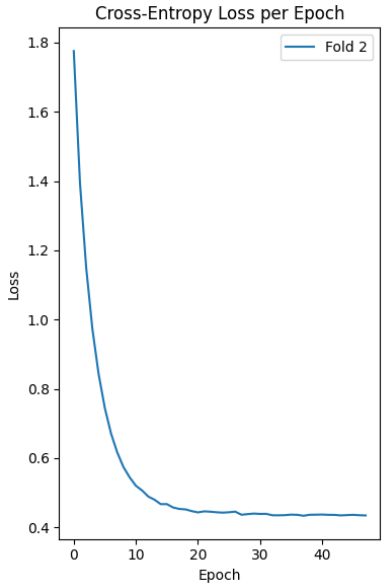

Cross-Entropy Loss per Epoch
Fold 4
Loss
Epoch

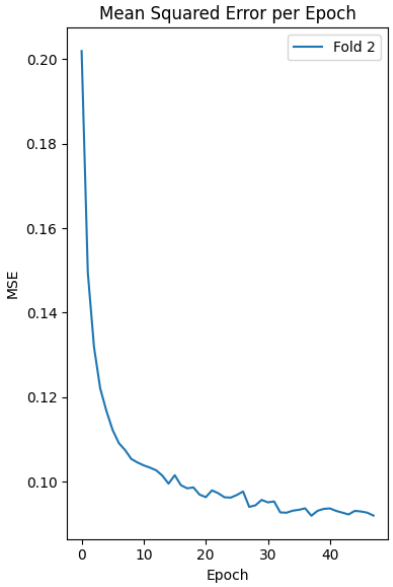

Mean Squared Error per Epoch
Fold 4
MSE
Epoch

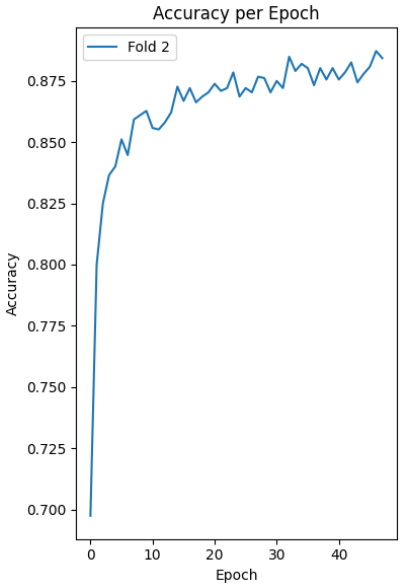

Accuracy per Epoch
Fold 4
Accuracy
Epoch

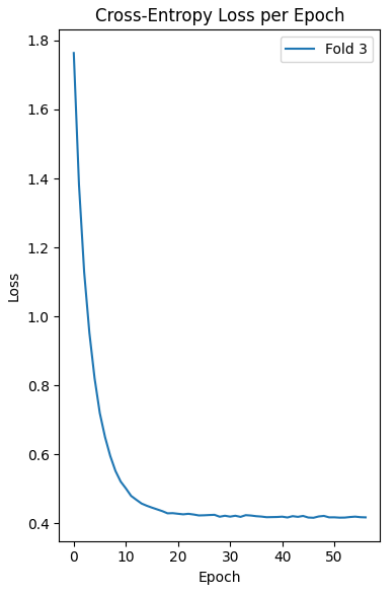
2)

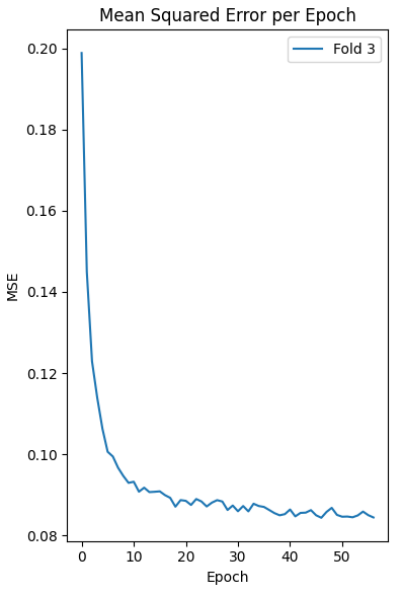

Cross-Entropy Loss per Epoch
Fold 0
Loss
Epoch
Mean Squared Error per Epoch
Fold 0
MSE
Epoch
Accuracy per Epoch
Fold 0
Accuracy
Epoch
Cross-Entropy Loss per Epoch
Fold 1
Loss
Epoch
Mean Squared Error per Epoch
Fold 1
MSE
Epoch
Accuracy per Epoch
Fold 1
Accuracy
Epoch

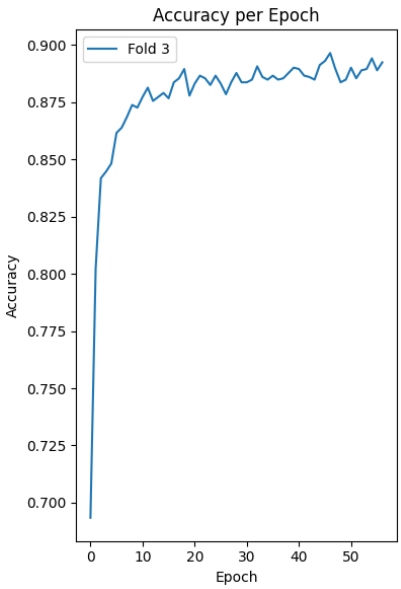

Cross-Entropy Loss per Epoch
Fold 2
Loss
Epoch

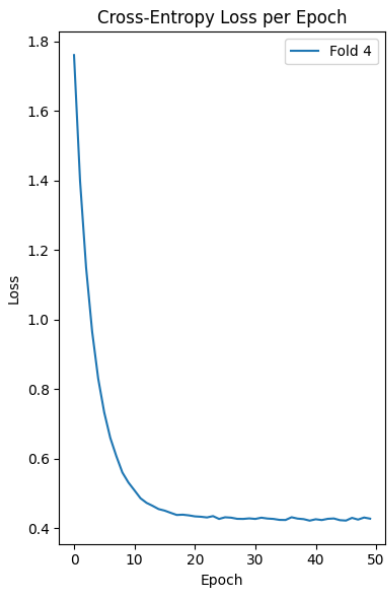

Mean Squared Error per Epoch
Fold 2
MSE
Epoch

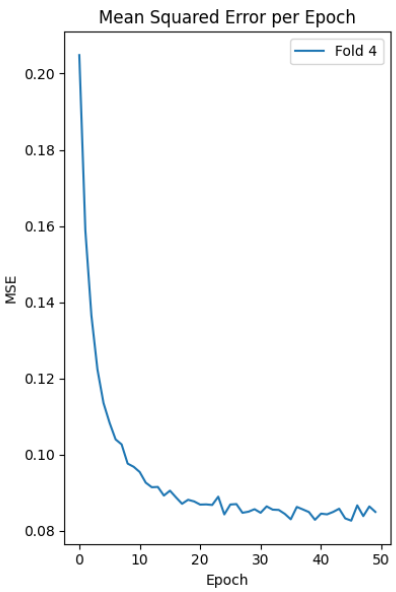

Accuracy per Epoch
Fold 2
Accuracy
Epoch

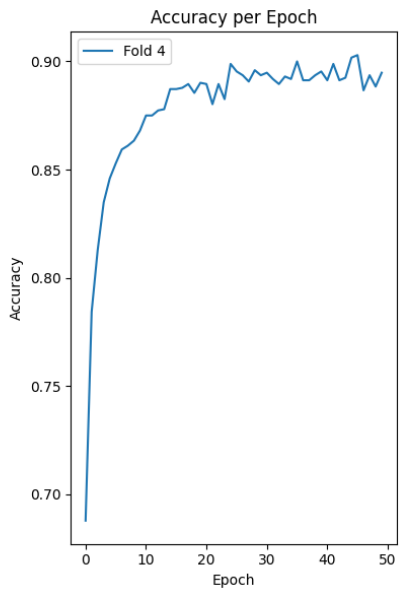

Cross-Entropy Loss per Epoch
Fold 3
Loss
Epoch


Mean Squared Error per Epoch
Fold 3
MSE
Epoch


Accuracy per Epoch
Fold 3
Accuracy
Epoch


Cross-Entropy Loss per Epoch
Fold 4
Loss
Epoch


Mean Squared Error per Epoch
Fold 4
MSE
Epoch


Accuracy per Epoch
Fold 4
Accuracy
Epoch