Final Project Artificial Neural Networks

Nikos Periklis Chourdsas -Tp4774

[nikos.hourdas@gmail.com](mailto:nikos.hourdas@gmail.com)

**INDEX**

Introduction

Step 1. Data Visualization

Step 2. Dataset Verification and Repairing

Step 3. Data pre-processing

Step 4. Neural Network Creation

Step 5. Visualization of the trained Neural Network

Step 6. Neural Network Optimization

Step 7. Test Results

Step 8. Kohonen Network, elbow and silhuette method

Step 9. Conclusion

**Introduction**

Για την υλοποιηση του τελικου προτζεκτ επελεξα το Heberman’s Survival Data set το οποιο περιεχει τις περιπτωσεις θνησιμοτητας ασθενων οι οποιοι εκαναν χειρουργικη επεμβαση για τον καρκινο του μαστού . Το συγκεκριμενο dataset αποτελειται από 3 + 1 στηλες ,3 για τα δεδομενα του και 1 για το αποτελεσμα (target ) . Κάθε στηλη ονομαστικε αναλογα ‘Age’ , ‘Year Operation’, ‘Axillary Nodes Detected’ , και για το target ‘Survival status’ .

<https://archive.ics.uci.edu/ml/datasets/haberman's+survival>

**Step 1**

Στο παρακατω γραφιμα προβαλουμαι όλα τα δεδομενα σε ένα διαγραμμα plot [1.1.1], καθως και μια αρχικη ποσοστιαια απεικονηση [1.2.1] των τιμων που περιεχονται .

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

1.2

1.1 1

Όπως Παρατηρουμε στο διαγραμμα [1.3.1] τυπου scatter plot ο στοχος του δικτυου αποτελειται από 1 και 2 , για να το χρησιμποποιησουμε στο θα πρεπει να μετατραπει στο διαδικο συστημα . Ετσι τα 1 θα γινουν 001 και αναλογα το 2 θα μετατραπει σε 010 . Ο στοχος θα αποθηκευτει ως ενας πολυδιαστατος πινακας με διαστασεις 306 x 3

A picture containing line chart

Description automatically generated

1.3

Στα διαγραμματα [1.4.1] και [1.5.1] μπορουμε να διακρινουμε την σχεση Age – Axillary nodes , Παρατηρουμε ότι οσο μεγαλυτερος ο αριθμος Axillary nodes τοσες περισσοτερες οι πιθανοτητες να επιβιωσει ο ασθενης . Ιδιατερα εντονα το βλεπουμε στις ηλικιες 40 και ανω

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

1.5

1.4

Στο τελευταιο διαγραμμα [1.6.1] Scatter Plοt απεικονιζονται όλα τα δεδομενα του Dataset

Chart, scatter chart

Description automatically generated

1.6

**Step 2 (optional)**

Ελεγχο το dataset gia πιθανες null τιμες

A picture containing text, monitor, screen, black

Description automatically generated

2.1

Από τον παραπανω πινακα [2.1.2] μπορουμε να διευκρινησουμε αν υπαρχει καποια «κενη» σειρα στον πινακα του dataset . Αφου μας εμφανιζει σε ολες τις γραμμες False σημαινει ότι ο πινακας μας δεν εχει καποιο κενο κελή και μπορουμε να συνεχισουμε την διαδικασια χωρις καποια περεταιρω τροποιηση

**Step 3**

Χωριζουμε το dataset σε 2 κομματια . Το ένα κομματι θα είναι για να εκπαιδευσουμε το δικτυο ενώ το δευτερο κομματι θα χρησιμοποιειται για να δοκιμασουμε το performance του δικτυου . Το αρχικο split που εφαρμοζουμε είναι 70 – 30 % .

Επειτα χρησιμοποιουμε την συναρτηση MinMaxScaler() από την βιβλιοθηκη scikit-learn η οποια αφου την εφαρμοσουμε αλλαζει το ευρως τιμων σε μηδεν και ένα [ 0 , 1 ]

**Step 4**

Για την δημιουργια του Νευρωνικου δικτυου χ

|  |  |
| --- | --- |
| **Epoch** | **Error** |
| **5** | 19.32565459106957 |
| **10** | 0.09763118756962419 |
| **15** | 1.1309042186301525e-05 |
| The goal of learning is reached | |

**Chart, line chart

Description automatically generated**

**Step 5**

**Step 7**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **70 -30** | **60-40** | **50 – 50** | **80 -20** | **90 -10** |
| **Epochs** | 15 | 15 | 18 | 21 | 26 |
| **Accuracy** | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |
| **Error** | 1.13090421863  01525e-05 | 3.1433613687417  77e-05 | 9.8643916436138  25e-06 | 4.1134081526478  51e-05 | 1.01949090119  16465e-05 |