



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

**ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ**

**Αναφορά Εργαστηριακής Άσκησης
Μέρος Β΄
Υπολογιστική Νοημοσύνη**

Κωνσταντίνος Τσάκωνας
Α.Μ.: 1059666

Ακαδημαϊκό έτος 2020-21
Εαρινό Εξάμηνο

Περιεχόμενα

Repository Κώδικα

Για την ανάπτυξη της άσκησης χρησιμοποιήθηκε η γλώσσα Python με τις βιβλιοθήκες Tensorflow, numpy, pandas, dear και matplotlib. Παρακάτω υπάρχει το repository του κώδικα στο github

[github link](#)

Σχεδιασμός ΓΑ

1. Τα άτομα του αρχικού πληθυσμού θα αναπαραστηθούν ως δυαδικές συμβολοσειρές. Ο λόγος που θα ακολουθηθεί αυτή η κωδικοποίηση προέρχεται από το σκεπτικό ότι, αυτό που θέλουμε να κάνουμε είναι να μειώσουμε το είσοδος από τα 784 pixels, δηλαδή να μηδενισούμε πολλές από αυτές. Δημιουργώντας λοιπόν άτομα του πληθυσμού ως πίνακες 784 pixels που περιέχουν δυαδικά στοιχεία, μπορούμε με ένα πολλαπλασιασμό element-wise να κρατήσουμε τις εισόδους που θέλουμε.
2. Ο αρχικός πληθυσμός θα είναι N τυχαίοι πίνακες 28x28 και οι τιμές που θα περιέχουν θα είναι 0 και 1.
3. Η συνάρτηση καταλληλότητας που επιλέχθηκε είναι οι εξής:
 - κάθε φορά που θα γίνεται έλεγχος στο νευρωνικό σύμφωνα με τις εισόδους που προκύπτουν από κάθε άτομο του πληθυσμού θα κάνουμε ταξινόμηση των πρώτων 10.000 εικόνων και θα συγκρίνουμε την ταξινόμηση αυτή με βάση τα labels για να δούμε πόσο ακριβής είναι. Το αποτέλεσμα που θα προκύπτει από αυτό θα είναι μια τιμή μεταξύ του διαστήματος $[0, 1]$ και στόχος του γενετικού αλγορίθμου θα είναι να πλησιάσει όσο το δυνατό πιο κοντά στο 1.
 - Επίσης θα επιβάλεται μία ποινή σε άτομα του πληθυσμού που έχουν μεγάλο αριθμό εισόδων, συγκεκριμένα άτομα που έχουν περισσότερες από 392 εισόδους θα αφαιρείται μία τιμή η οποία θα είναι ανάλογη με το ποσοστό που δεν ταξινομήθηκε σωστά επί το πλήθος των παραπάνω εισόδων που έχει σε σχέση με αυτές που έχουμε θέσει ως επιθυμητό άνω όριο.

Ο λόγος που επιλέχθηκε η μεγιστοποίηση του ποσοστού ταξινόμησης έχει να κάνει με το γεγονός ότι κατά την δημιουργία και την εκπαίδευση του μοντέλου στην προηγούμενη εργαστηριακή άσκηση σε πάρα πολύ

μικρές τιμές του loss η ταξινόμηση δεν ήταν πάντα σωστή και δεν συμβάδιζε με το accuracy του μοντέλου. Ακόμα ο λόγος που εφαρμόζουμε ποινή σε άτομα που έχουν περισσότερες από 392 εισόδους είναι γιατί θα έχουμε μία μείωση 50% αλλά για λιγότερες εισόδους από αυτές θα χάνουμε χαρακτηριστικά που χρειαζόμαστε αφού το σχήμα των ψηφιών στις εικόνες συνεχώς μεταβάλλεται, με αποτέλεσμα να μην έχουμε καλή ταξινόμηση.

4. Γενετικοί Τελεστές:

(α') Επιλογή:

- Ρουλέτα βάσει κόστους: Με αυτή την μέθοδο από N άτομα επιλέγουμε με βάση μία πιθανότητα p_i τα άτομα τα οποία θα γίνει η διασταύρωση. Η πιθανότητα υπολογίζεται ως εξής:

$$p_i = \frac{f_i}{\sum_{i=1}^N f_i}$$

όπου,

f_i : το fitness του i -οστού ατόμου,

N : το μέγεθος του πληθυσμού.

Κατα αυτό τον τρόπο τα άτομα με το μεγαλύτερο fitness έχουν μεγαλύτερη πιθανότητα να επιλεγούν, δηλαδή βαρύνουν περισσότερο.

- Ρουλέτα βάσει κατάταξης: Με την μέθοδο αυτή τα άτομα κατατάσσονται με σύμφωνα με την καταλληλότητα σε αύξουσα σειρά. Στο άτομα με τη μικρότερη καταλληλότητα ανατίθεται κατάταξη 1, στο αμέσως επόμενο 2 και η διαδικασία συνεχίζεται μέχρι το καταλληλότερο άτομα να έχει κατάταξη N , όπου N το μέγεθος του πληθυσμού. Το μέγεθος που πιάνει το άτομα πάνω στη ρουλέτα σε τοις εκατό προκύπτει από:

$$\frac{r_i}{\sum_{i=1}^N r_i} \times 100$$

όπου,

r_i : η κατάταξη του i -οστού ατόμου,

N : το μέγεθος του πληθυσμού.

Σύμφωνα με το παραπάνω τα άτομα βαρύνουν με τον ίδιο τρόπο ανεξαρτήτως της μεγάλης απόκλισης που μπορεί να έχει η συνάρτηση καταλληλότητας τους.

- Τουρνουά: Με την μέθοδο αυτή δημιουργούμε ένα τουρνουά μεγέθους K . Σε κάθε τουρνουά διαλέγουμε τυχαία K άτομα, από αυτά επιλέγεται αυτό με το καλύτερο fitness για να προχωρήσει στην επόμενη γενιά. Τα άτομα που παίζουν στην επόμενη γενιά δεν αφαιρούνται από τον πληθυσμό με αποτέλεσμα να μπορούν να επιλεγούν ξανά.

Από αυτές αυτή που επιλέχθηκε για τον αλγόριθμο μας σύμφωνα με το πρόβλημα που έχουμε να επιλύσουμε είναι η μέθοδος του Τουρνουά. Ο λόγος που επιλέχθηκε βασίζεται στο γεγονός ότι τα άτομα που θα περάσουν στην επόμενη γενιά και θα διασταυρωθούν θα έχουν μεγαλύτερες διαφοροποιήσεις μεταξύ τους, διότι επιλέγονται K άτομα τυχαία. Στις υπόλοιπες μεθόδους οι επιλογή γίνεται με βάση κάποια πιθανότητα από τον αρχικό πληθυσμό με αποτέλεσμα τα άτομα με καλύτερη καταλληλότητα να είναι πιο πιθανό να επιλεγούν. Άρα με την χρήση του τουρνουά και την ποικιλομορφία που προσφέρει στη διασταύρωση μπορεί να προκύψει κάποιο άτομο με πολύ καλύτερο fitness από το καλύτερο άτομο της προηγούμενης γενιάς με αποτέλεσμα ο αλγόριθμος να συγκλίνει πιο γρήγορα.

5. Διασταύρωση:

-