

# Εργασία Υπολογιστική Νοημοσύνη

## Μέρος Α'

Γρηγόρης Καπαδούκας (ΑΜ: 1072484)

2 Απριλίου 2023

### 0 Περιβάλλον Εργασίας - Σύνδεσμος GitHub με Κώδικα

Για την διεκπεραίωση αυτής της εργασίας έχω επιλέξει να χρησιμοποιήσω γλώσσα προγραμματισμού Python μαζί τις βιβλιοθήκες TensorFlow (κυρίως το API της, το Keras) για τον σχεδιασμό και την εκπαίδευση των νευρωνικών δικτύων και το Pandas με σκοπό τον χειρισμό του CSV αρχείου και της προεπεξεργασίας.

Ο κώδικας που γράφτηκε για την εργασία βρίσκεται στο repository στον παρακάτω σύνδεσμο:

<https://github.com/GregKapadoukas/University-Computational-Intelligence-Project-A>

### 1 Προεπεξεργασία και Προετοιμασία Δεδομένων

#### 1.a Κωδικοποίηση και προεπεξεργασία δεδομένων

##### 1.a.1 Διάβασμα του CSV αρχείου και μετατροπή κατηγορικών δεδομένων σε αριθμητικά

Αρχικά, με σκοπό το διάβασμα του dataset στη μορφή του CSV αρχείου χρησιμοποιώ εντολές της βιβλιοθήκης Pandas για φορτώσω τα δεδομένα

στη μορφή ενός DataFrame. Έπειτα χωρίζω το αρχικό DataFrame σε δύο, από τα οποία το πρώτο περιέχει τα δεδομένα των αισθητήρων και τα στοιχεία του ατόμου πάνω στο οποίο έγιναν οι μετρήσεις και το δεύτερο περιέχει την κλάση δραστηριότητας στην οποία ανήκει το άτομο.

Έπειτα μετατρέπω τα κατηγορικά δεδομένα των κλάσεων του δεύτερου DataFrame σε αριθμητικά δεδομένα, τα οποία όμως είναι one-hot encoded διανύσματα μεγέθους  $\mathbb{R}^{1 \times 5}$ . Άρα οι τιμές μετατρέπονται σε διανύσματα με την εξής αντιστοίχιση:

- 'sitting': [1 0 0 0 0]
- 'sitting-down': [0 1 0 0 0]
- 'standing': [0 0 1 0 0]
- 'standing-up': [0 0 0 1 0]
- 'walking': [0 0 0 0 1]

Επέλεξα την παραπάνω one-hot encoded προσέγγιση αντί για την αριθμητική 1-5 όπως προτείνεται στην εκφώνηση, επειδή σκοπεύω να έχω 5 εξόδους στο νευρωνικό δίκτυο, όπως θα εξηγήσω στο κεφάλαιο 2.

Έπειτα μετατρέπω και τα κατηγορικά δεδομένα του δεύτερου DataFrame, δηλαδή τα series 'Name' και 'Gender' σε αριθμητικά δεδομένα, με βάση την εξής αντιστοίχιση.

Για τα ονόματα:

- 'debora': 1
- 'katia': 2
- 'wallace': 3
- 'jose\_carlos': 4

Για το φύλλο:

- 'Man': 1
- 'Woman': 2

Έτσι πλέον έχω μετατρέψει όλα τα κατηγορικά δεδομένα σε αριθμητικά, το οποίο είναι αναγκαίο για να μπορέσει να γίνει η μετέπειτα προεπεξεργασία των δεδομένων και η χρήση τους για την εκπαίδευση του νευρωνικού δικτύου.

## **2   Επιλογή Αρχιτεκτονικής**