

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ · ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΕΞΟΥΞΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ ΜΑΘΗΣΗΣ

ΕΡΓΑΣΤΗΡΙΑΚΗ ΑΣΚΗΣΗ · 2023–2024

ΠΕΡΙΕΧΟΜΕΝΑ

1	ΕΡΩΤΗΜΑ 1	2
1.1	ΑΝΑΛΥΣΗ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ	2
1.2	ΓΡΑΦΙΚΕΣ ΠΑΡΑΣΤΑΣΕΙΣ	2

1 ΕΡΩΤΗΜΑ 1

Για την εισαγωγή και τη προεπεξεργασία του `.csv` αρχείου, θα χρησιμοποιήσουμε τη βιβλιοθήκη `pandas` της Python.

1.1 ΑΝΑΛΥΣΗ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ

Καταρχάς, χρησιμοποιώντας την συνάρτηση `head()` μπορούμε να δούμε τις πρώτες εγγραφές από το πρώτο αρχείο του συνόλου δεδομένων `S006.csv`:

	timestamp	back_x	back_y	back_z	thigh_x	thigh_y	thigh_z	label
0	2019-01-12 00:00:00.000	-0.760242	0.299570	0.468570	-5.092732	-0.298644	0.709439	6
1	2019-01-12 00:00:00.010	-0.530138	0.281880	0.319987	0.900547	0.286944	0.340309	6
2	2019-01-12 00:00:00.020	-1.170922	0.186353	-0.167010	-0.035442	-0.078423	-0.515212	6
3	2019-01-12 00:00:00.030	-0.648772	0.016579	-0.054284	-1.554248	-0.950978	-0.221140	6
4	2019-01-12 00:00:00.040	-0.355071	-0.051831	-0.113419	-0.547471	0.140903	-0.653782	6

Σε συνδυασμό με την `info()`, παρατηρούμε πώς για κάθε χρονική στιγμή δίνονται οι τιμές των αισθητήρων, αποθηκευμένες ως `float24`, στις τρεις διαστάσεις (x, y, z) για τις περιοχές της πλάτης και του μηρού, καθώς και ένα `int64` `label`.

Για να ελέγξουμε την ακεραιότητα και να εντοπίσουμε τυχούσες συνέπειες, μέσω της συνάρτησης `concat()` ενώνουμε όλα τα 22 αρχεία σε ένα ενιαίο `dataframe`.

	sum
timestamp	0
back _x	0
back _y	0
back _z	0
thigh _x	0
thigh _y	0
thigh _z	0
label	0
index	5740689
Unnamed: 0	6323682

Παρατηρούμε πως έχουν εμφανιστεί `NaN` τιμές στις στήλες `index` και `Unnamed: 0`, οι οποίες δεν ήταν παρούσες στο αρχικό `S006.csv`. Ελέγχοντας όλα τα αρχεία, η στήλη `index` εμφανίζεται στο αρχείο και η στήλη `Unnamed: 0` στο αρχείο.

Χρησιμοποιώντας τη συνάρτηση `describe()` μπορούμε να υπολογίσουμε βασικές στατιστικές μετρικές για τα δεδομένα μας. Παίρνοντας ως παράδειγμα το `S006.csv`:

	back _x	back _y	back _z	thigh _x	thigh _y	thigh _z	label
count	408709	408709	408709	408709	408709	408709	408709
mean	-0.802201	-0.000687	-0.274718	-0.370317	0.143471	0.617527	10.190187
std	0.238347	0.189062	0.441805	0.506666	0.213864	0.536430	20.328336
min	-3.542889	-3.016498	-1.024363	-6.844045	-5.757406	-4.884791	1.000000
25%	-0.983647	0.001063	-0.702338	-0.952840	0.022534	0.144114	6.000000
50%	-0.937195	0.033240	-0.277446	-0.277711	0.086248	0.924066	7.000000
75%	-0.654541	0.074822	0.064811	0.068999	0.246292	1.001372	7.000000
max	0.952109	2.569339	1.628023	3.898547	4.602909	5.391660	130.000000

1.2 ΓΡΑΦΙΚΕΣ ΠΑΡΑΣΤΑΣΕΙΣ