

ΜΕΡΟΣ Α'

Σε πρώτη φάση θα πρέπει να αναλύσεις τα δεδομένα που περιέχονται στο αρχείο `weather_data.csv`, και να εφαρμόσεις βασικές μεθόδους προεπεξεργασίας δεδομένων.

Συγκεκριμένα:

- θα πρέπει να βρεις τις μέγιστες και ελάχιστες τιμές των αριθμητικών στηλών καθώς και των μέσο όρο
- θα πρέπει να βρεις τις μοναδικές τιμές των κατηγορικών δεδομένων καθώς επίσης και την συχνότητα εμφάνισης της κάθε μιας
- θα πρέπει να μετατρέψεις τα κατηγορικά δεδομένα σε μορφή που θα μπορεί να επεξεργαστεί ένας αλγόριθμος (αριθμητική μορφή)
- θα πρέπει να μετατρέψεις τις χρονικές σημάνσεις σε μορφή που θα μπορεί να επεξεργαστεί ένας αλγόριθμος χάνοντας όσο το δυνατό λιγότερη πληροφορία
- θα πρέπει να εφαρμόσεις μεθόδους κλιμακώσεις των δεδομένων

ΜΕΡΟΣ Β'

Έπειτα θα πρέπει να φτιάξεις ένα νευρωνικό δίκτυο, που θα προβλέπει την τιμή του πεδίου 'value' ανά 'location_id' και ανά τύπο μέτρησης. Θα πρέπει να μπορείς να δίνεις ως παραμέτρους στο δίκτυο:

- των αριθμών των νευρώνων (neurons) στα στρώματα (layers)
- τη συνάρτηση ενεργοποίησης των στρωμάτων (activation function)
- τη συνάρτηση μέτρησης του λάθους (error/cost function)
- τη μέθοδο εκπαίδευσης του δικτύου, δλδ. τη μέθοδο βελτιστοποίησης των βαρών του δικτύου (optimization method)

Συγκεκριμένα:

- για τις συναρτήσεις ενεργοποίησης θα πρέπει να μπορείς να επιλέξεις τουλάχιστον μεταξύ των sigmoid και relu
- για τις συναρτήσεις μέτρησης του λάθους θα πρέπει να μπορείς να επιλέξεις τουλάχιστον μεταξύ των mean square error, mean absolute error.
- για τις μεθόδους εκπαίδευσης θα πρέπει να μπορείς να επιλέξεις τουλάχιστον μεταξύ των stochastic gradient descent και του αλγόριθμου Adam.

ΜΕΡΟΣ Γ'

Επιπλέον καλό θα ήταν:

- να συλλέγεις τα αποτελέσματα της αξιολόγησης του δικτύου με βάση τις παραμέτρους που χρησιμοποιήθηκαν (αριθμός νευρώνων, συνάρτηση ενεργοποίησης κτλ.)
- να αξιολογείς την χρησιμότητα του κάθε εισερχόμενου δεδομένου (στήλες) ως προς το τελικό αποτέλεσμα, χρησιμοποιώντας τα Shapley values; για αρχή μπορείς να διαβάσεις αυτά:

<https://towardsdatascience.com/one-feature-attribution-method-to-supposedly-rule-them-all-shapley-valuesf3e04534983d>, <https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>.

Θα ήταν προτιμότερο να υλοποιήσεις την άσκηση στη Python και να χρησιμοποιήσεις το Tensor Flow 2.x. Σημείωσε ότι το σημαντικό δεν είναι να πετύχεις καλά αποτελέσματα αλλά θα πρέπει να φτιάξεις ένα εύχρηστο και εύκολο να κατανοηθεί κώδικα που θα μπορούσε οποιοσδήποτε να χρησιμοποιήσει, και αλλάζοντας τις παραμέτρους του δικτύου να μπορεί να παράγει νέα αποτελέσματα, καθώς επίσης να παρουσιάσεις με κατανοητό τρόπο τα ευρήματά σου του πρώτου μέρους, και οποιαδήποτε αποτελέσματα των αλγορίθμων που θα δοκιμάσεις.