

# Εργασία στην Τεχνητή Νοημοσύνη #2

Βασίλειος Κοκκινογένης it2021042

**Ερώτηση 1:** Πως ερμηνεύετε τον πίνακα σύγχυσης;

**Απάντηση:** Γνωρίζουμε ότι:

Το επάνω αριστερό κελί είναι ο αριθμός των περιπτώσεων που είναι true negatives (το μοντέλο προβλέπει σωστά αρνητικά).

Το επάνω δεξιά κελί είναι ο αριθμός των περιπτώσεων που είναι false positives (το μοντέλο προβλέπει εσφαλμένα θετικά)

Το κάτω αριστερό κελί είναι ο αριθμός των περιπτώσεων που είναι false negatives (το μοντέλο προβλέπει λανθασμένα αρνητικά)

Το κάτω δεξιά κελί είναι ο αριθμός των περιπτώσεων που είναι true positives (το μοντέλο προβλέπει σωστά θετικά)

Εμείς στην περίπτωση μας έχουμε:

Confusion matrix for training set:

7138	160
824	509

Confusion matrix for test set:

3061	63
370	205

Δηλαδή στο training set έχουμε 7138 true negatives, 509 true positives, 160 false positives, 824 false negatives, ενώ αντίστοιχα στο test set έχουμε 3061 true negatives, 205 true positives, 63 false positives, 370 false negatives. Και στα 2 sets έχουμε ένα μικρό ποσοστό false προβλέψεων. Πιο συγκεκριμένα, έχουμε  $824 / (7138 + 824) = 10.34\%$  false negatives στο training set όπως και  $370 / (3061 + 370) = 10.78\%$  false negatives στο test set. Αντίστοιχα έχουμε  $160 / (509 + 160) = 23.91\%$  false positives στο training set όπως και  $63 / (205 + 63) = 23.50\%$  false positives στο test set. Ωστόσο, παρατηρούμε ότι σε σχέση με τα false negatives, τα false positives έχουν μεγαλύτερο ποσοστό. Συνεπώς, μια βελτίωση του μοντέλου θα μπορούσε να είναι η μείωση του ποσοστού αυτού.

**Ερώτηση 2:** Τι τροποποιήσεις ή επιπλέον πειράματα θα υλοποιούσατε ώστε να βελτιώσετε το μοντέλο σας;

**Απάντηση:** Για να βελτιώσουμε το μοντέλο μας, θα μπορούσαμε να τροποποιήσουμε την υλοποίηση του μοντέλου μας (Ερώτημα 4: Υλοποίηση μοντέλου). Αντί για `penalty=None` θα μπορούσαμε να δοκιμάσουμε να χρησιμοποιήσουμε `penalty=l1` ή κάποια άλλη μεταβλητή. Επιπλέον θα μπορούσαμε να αλλάξουμε τον τρόπο με τον οποίο προετοιμάζουμε τα δεδομένα μας (Ερώτημα 3: Προετοιμασία δεδομένων). Αντί για `train_size=0.70` (και `test_size=0.30`), θα μπορούσαμε να πειραματιστούμε με διαφορετικά train-test split ratios, π.χ. `train_size = 0.65` (και `test_size=0.35`).