

Final Project

Η τελική εργασία αποτελείται από 2 προβλήματα. Και τα 2 προβλήματα πρέπει να προσεγγιστούν επαρκώς και να υποβληθούν για να θεωρηθεί ότι η εργασία είναι ολοκληρωμένη.

Πρόβλημα 1

Για το πρόβλημα 1 να χρησιμοποιηθεί το dataset [NSL-KDD](#), το οποίο περιέχει πακέτα δεδομένων που αφορούν την επικοινωνία μεταξύ συστημάτων σε ένα δίκτυο (περισσότερες πληροφορίες [εδώ](#)). Το dataset αυτό δημιουργήθηκε για το σκοπό του εντοπισμού εισβολών (δηλ. επιθέσεων) μέσα στο δίκτυο.

Σκοπός επομένως για αυτό το πρόβλημα είναι να δοκιμαστούν αλγόριθμοι Μηχανικής Μάθησης οι οποίοι μπορούν να ανιχνεύσουν πότε μια εγγραφή στο σύνολο δεδομένων (δηλ. ένα πακέτο επικοινωνίας) χαρακτηρίζει την επικοινωνία ως φυσιολογική ή αποτελεί εισβολή στο δίκτυο.

Το αρχείο *NSL-KDDTrain.csv*, το οποίο θα χρησιμοποιηθεί για την εκπαίδευση των αλγορίθμων, είναι unlabeled, δηλαδή δεν περιέχει το είδος της επικοινωνίας (normal ή attack). Παρ'όλ'αυτά, το αρχείο *NSL-KDDTest.csv* περιέχει τα labels, το οποίο θα χρησιμοποιηθεί για την επαλήθευση των επιδόσεων των μοντέλων. Αυτό δεν σημαίνει όμως ότι μπορεί να χρησιμοποιηθεί κατά την εκπαίδευση! Είναι μόνο για να δείτε πόσο καλά τα πήγαν οι αλγόριθμοι.

Τελικό παραδοτέο είναι ο κώδικας (με σχόλια), ο οποίος, μετά από την εκπαίδευση, θα εκτυπώνει το ποσοστό ακρίβειας όσον αφορά το label που πρόβλεψε το μοντέλο σε σχέση με το πραγματικό label της εγγραφής. Επίσης, να παραδοθεί ένα έγγραφο/report όπου θα περιγράφονται οι προσεγγίσεις που δοκιμάστηκαν και τα συμπεράσματα από αυτές.

Για οποιαδήποτε άλλη πληροφορία, μπορείτε να αναρτήσετε τις απορίες και τα σχόλιά σας στο forum του μαθήματος.

Πρόβλημα 2

Για το πρόβλημα 2, να χρησιμοποιηθεί μια τροποποιημένη έκδοση του dataset **Car Fuel Consumptions and Emissions 2000-2013** (θα τη βρείτε [εδώ](#)), το οποίο περιέχει πληροφορίες/προδιαγραφές αυτοκινήτων, καθώς και τις εκπομπές καυσίμων του καθενός ανά 12.000 μίλια (στήλη *fuel_cost_12000_miles*).

Να εκπαιδευτούν αλγόριθμοι Μηχανικής Μάθησης οι οποίοι θα μπορούν να υπολογίζουν τις εκπομπές των καυσίμων ανάλογα με τα χαρακτηριστικά ενός αυτοκινήτου.

Τελικό παραδοτέο είναι ο κώδικας (με σχόλια), ο οποίος θα εκτυπώνει τις επιδόσεις του μοντέλου βάσει των μετρικών που έχουν αναφερθεί στο μάθημα. Ενδεικτικά (μιας και είναι ένα πρόβλημα παρεμβολής (regression), μπορούν να χρησιμοποιηθούν οι μετρικές MAE, MAPE, RMSE κλπ. Επίσης, να παραδοθεί ένα έγγραφο/report όπου θα περιγράφονται οι προσεγγίσεις που δοκιμάστηκαν και τα συμπεράσματα από αυτές.

Για οποιαδήποτε άλλη πληροφορία, μπορείτε να αναρτήσετε τις απορίες και τα σχόλιά σας στο forum του μαθήματος.