

HY-119 ΓΡΑΜΜΙΚΗ ΑΛΓΕΒΡΑ ΑΣΚΗΣΗ 8

Η άσκηση θα παραδοθεί ηλεκτρονικά στη σελίδα του μαθήματος <https://elearn.uoc.gr/>.

Η καταληκτική προθεσμία της παράδοσης είναι την Κυριακή **5/5/2019, 23:55**.

Οδηγίες Παράδοσης: Παραδώστε ένα αρχείο [αριθμός μητρώου σας]_ask8.zip που να περιέχει:

Ένα αρχείο **RootMeanSquareError.m** που υλοποιεί τη συνάρτηση RootMeanSquareError

Ένα αρχείο **main_[αριθμός μητρώου σας].m** που να υλοποιεί τα ερωτήματα

Ένα αρχείο **ask8.pdf** με σύντομη δομημένη αναφορά που να περιέχει τα γραφήματα και τις απαντήσεις στα υποερωτήματα που χρειάζεται.

Ό,τι άλλο αρχείο **.m** υλοποιήσατε ως συνάρτηση για να λειτουργεί ο κώδικας της main σας.

Άσκηση (MATLAB)

Μια αντιπροσωπεία πωλήσεως αυτοκινήτων διαθέτει καταστήματα σε 15 διαφορετικές πόλεις. Σας δίνεται το αρχείο data.mat το οποίο περιέχει τις εξής μεταβλητές:

y	Διάνυσμα με τον αριθμό πωλήσεων σε κάθε πόλη
aa	Το μέτρο του διανύσματος με τον πληθυσμό κάθε πόλης
bb	Το μέτρο του διανύσματος με το εισόδημα σε κάθε πόλη
ab	Το εσωτερικό γινόμενο των διανυσμάτων του πληθυσμού με του εισοδήματος
p_ab	Η προβολή του διανύσματος εισοδήματος b στο διάνυσμα πληθυσμού a
p_ba	Η προβολή του διανύσματος πληθυσμού a στο διάνυσμα εισοδήματος b

Φορτώστε το αρχείο με την εντολή load (Σημείωση: η εντολή ψάχνει το αρχείο στην ενεργή διεύθυνση εργασίας του Matlab, (Current Folder))

1. Υλοποιήστε τη συνάρτηση
[RMSE] = RootMeanSquareError(prediction, groundtruth)
Είσοδοι είναι δύο ίσου μεγέθους διανύσματα και έξοδος η scalar τιμή του μέσου τετραγωνικού σφάλματος μεταξύ των τιμών των δύο διανυσμάτων.
2. Υπολογίστε το διάνυσμα του πληθυσμού κάθε πόλης.
3. Υπολογίστε το διάνυσμα του εισοδήματος κάθε πόλης.
4. Σχεδιάστε όλα τα σημεία που προκύπτουν από το συνδυασμό των τιμών διανυσμάτων εισοδήματος και πωλήσεων στο δισδιάστατο χώρο. Τι παρατηρείτε για τη μεταξύ τους σχέση; (hint: scatter, plot, xlabel, ylabel, title)
5. Σχεδιάστε όλα τα σημεία που προκύπτουν από το συνδυασμό των τιμών διανυσμάτων πληθυσμού και πωλήσεων στο δισδιάστατο χώρο. Τι παρατηρείτε για τη μεταξύ τους σχέση; (hint: scatter, plot)
6. Υπολογίστε εκ νέου τις προβολές των διανυσμάτων πληθυσμού και εισοδήματος. Υπολογίστε το μέσο τετραγωνικό σφάλμα των προβολών που υπολογίσατε με τις αντίστοιχες δοθείσες τιμές χρησιμοποιώντας τη συνάρτηση του ερωτήματος 1. Τι παρατηρείται και γιατί;
7. Χρησιμοποιώντας τα διανύσματα εισοδήματος και πληθυσμού κάθε πόλης βρείτε **γραμμικούς μετασχηματισμούς** του κάθε διανύσματος ξεχωριστά ή του συνδυασμού τους και σχεδιάστε τα νέα διανύσματα ως προς το διάνυσμα των πωλήσεων.

8. Χρησιμοποιώντας τα διανύσματα εισοδήματος και πληθυσμού κάθε πόλης βρείτε **μη γραμμικούς μετασχηματισμούς** του κάθε διανύσματος ξεχωριστά ή του συνδυασμού τους και σχεδιάστε τα νέα διανύσματα ως προς το διάνυσμα των πωλήσεων. (*hint: x^2 , \cos , \sin , \log , etc*)
9. Έχοντας ως δεδομένα τα διανύσματα των ερωτημάτων 2,3,6,7 και 8, χωριστά το κάθε ένα, και χρησιμοποιώντας **ΜΟΝΟ** τις 10 πρώτες τιμές των διανυσμάτων, δηλαδή τις 10 πρώτες πόλεις, βρείτε την ευθεία που παρεμβάλλει καλύτερα τα δεδομένα σας σε σχέση με το διάνυσμα των πωλήσεων, βρίσκοντας τη λύση *ελαχίστων τετραγώνων* του συστήματος σας $Ax = b$.
Για κάθε ευθεία:
 - a. Δώστε το διάνυσμα με τις εκτιμώμενες τιμές πώλησης βάση των ευθειών και των αντίστοιχων διανυσμάτων που χρησιμοποιήσατε.
 - b. Σχεδιάστε σε ένα γράφημα (*hint: hold on/off, legend*)
 - i. όλα τα σημεία που προκύπτουν από το συνδυασμό των τιμών του διανύσματος που χρησιμοποιείτε και των δοθέντων πωλήσεων y στο δισδιάστατο χώρο
 - ii. όλα τα σημεία που προκύπτουν από το συνδυασμό των τιμών του διανύσματος που χρησιμοποιείτε και των τιμών πωλήσεων που υπολογίσατε στο δισδιάστατο χώρο
 - iii. την ευθεία που βρήκατε
 - c. Βρείτε το μέσο τετραγωνικό σφάλμα (*Root Mean Square Error*) των διανυσμάτων αυτών με τη συνάρτηση του ερωτήματος 1
 - d. Βρείτε το σφάλμα E των διανυσμάτων αυτών (τη διαφορά των προβλεπόμενων τιμών από τις δοθείσες τιμές πώλησης y (*residuals*)) και υπολογίστε το μέτρο του.

Συγκεντρώστε τα αποτελέσματα σας σε έναν πίνακα όπου στην πρώτη στήλη δηλώνετε ποιο διάνυσμα ή μετασχηματισμό του χρησιμοποιήσατε, στη δεύτερη το σφάλμα του υποερωτήματος c και στην τρίτη το σφάλμα του υποερωτήματος d.

10. Βάση των σφαλμάτων που βρήκατε δηλώστε ποιο γνώρισμα/-τα θεωρείτε ως αυτό/-ά με την καλύτερη προβλεπτική ικανότητα για τον αριθμό των πωλήσεων και από ποια ευθεία γίνεται αντίστοιχα αυτή η πρόβλεψη.
11. Βάση της ευθείας και του γνωρίσματος που επιλέγετε ως επαρκή, προβλέψτε τις τιμές των πωλήσεων στις 5 τελευταίες πόλεις. Υπολογίστε το μέσο τετραγωνικό σφάλμα με τη συνάρτηση του ερωτήματος 1 και το αντίστοιχο σφάλμα E . Τι παρατηρείτε για τις προβλεπόμενες τιμές σε σχέση με τις πραγματικές;

Συγχαρητήρια!

Αν ολοκληρώσατε όλα τα βήματα σωστά, υλοποιήσατε έναν απλό αλγόριθμο μηχανικής μάθησης, κατασκευάζοντας χαρακτηριστικά και επιλέγοντας αυτά που θεωρήσατε ως κατάλληλα. Εκπαιδεύσατε τον αλγόριθμο σας σε ένα μικρό γνωστό δείγμα, μετρήσατε την απόδοση του και μπορέσατε να τον εφαρμόσετε σε ένα άγνωστο δείγμα ώστε να προβλέψετε τις τιμές πώλησης!