

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ

Μεταπτυχιακό Μάθημα

ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ

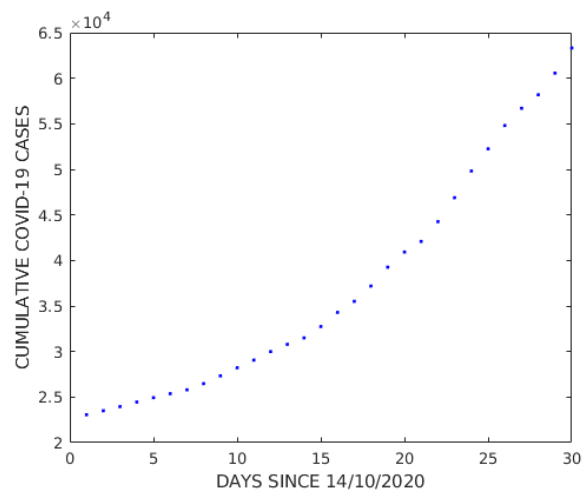
Ακαδημαϊκό Έτος 2020-21

- ΕΡΓΑΣΙΑ 1 -

Παράδοση μέχρι: 30 Νοεμβρίου 2020

Η κατασκευή πολυωνυμικών μοντέλων που προσεγγίζουν με τη μέγιστη δυνατή ακρίβεια τον αριθμό των συνολικών κρουσμάτων κατά τη διάρκεια πανδημιών αποτελούν βασικό εργαλείο στην κατανόηση και ανάλυση της δυναμικής της εξάπλωσης της νόσου.

Στο παρεχόμενο αρχείο *covid_data_30_GR.dat* δίνεται ο συσσωρευτικός αριθμός κρουσμάτων του covid-19 στην Ελλάδα για διάστημα 30 ημερών από 14 Οκτωβρίου 2020 μέχρι και 12 Νοεμβρίου 2020. Στην πρώτη στήλη των δεδομένων δίνεται ο αριθμός x_i της ημέρας ($i=1,2,\dots,30$) ξεκινώντας από την 14^η Οκτωβρίου (ημέρα 1) και στη δεύτερη στήλη δίνεται ο αντίστοιχος συνολικός αριθμός κρουσμάτων y_i που έχει καταγραφεί στη χώρα από την αρχή της πανδημίας (1 Μαρτίου 2020) μέχρι την αντίστοιχη ημέρα (π.χ. μέχρι την ημέρα 1, δηλαδή από την 1^η Μαρτίου 2020 μέχρι τις 14 Οκτωβρίου 2020 είχαν καταγραφεί 23060 κρούσματα). Τα δεδομένα απεικονίζονται στο παρακάτω γράφημα.



Γραφική αναπαράσταση των δεδομένων του αρχείου *covid_data_30_GR.dat*.

Ο στόχος της παρούσας εργασίας είναι να υπολογιστεί ένα πολυωνυμικό μοντέλο 4ου βαθμού της μορφής:

$$p(\alpha, x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \alpha_4 x^4$$

όπου $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)^T$ το άγνωστο διάνυσμα των συντελεστών του μοντέλου, το οποίο περιγράφει τα δεδομένα με τον καλύτερο δυνατό τρόπο, δηλαδή ελαχιστοποιεί τη συνάρτηση τετραγωνικού σφάλματος:

$$f(\alpha) = \sum_{i=1}^M [p(\alpha, x_i) - y_i]^2$$

όπου $M=30$ το πλήθος των διαθέσιμων μετρήσεων στα δεδομένα.

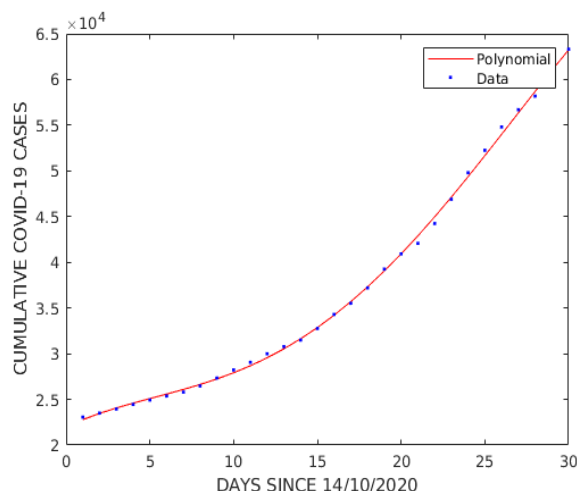
Για την επίλυση του προβλήματος θα πρέπει να χρησιμοποιηθούν και να συγκριθούν οι ακόλουθες μέθοδοι βελτιστοποίησης με παραγώγους:

1. Η μέθοδος Newton με ευθύγραμμη αναζήτηση με συνθήκες Wolfe.
2. Η μέθοδος Newton ασφαλούς περιοχής.
3. Η μέθοδος BFGS με ευθύγραμμη αναζήτηση με συνθήκες Wolfe.
4. Η μέθοδος steepest descent με ευθύγραμμη αναζήτηση με συνθήκες Wolfe.

Στο αρχείο *initial_points.dat* δίνονται 5 αρχικές τιμές του διανύσματος $\alpha=(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)^T$ των συντελεστών του μοντέλου, στις οποίες θα πρέπει να εφαρμοστεί κάθε μέθοδος και να συγκριθεί με τις υπόλοιπες ως προς το πλήθος των επαναλήψεων που απαιτεί για την εύρεση μιας λύσης με:

$$|| \nabla f(\alpha) || \leq 10^{-6}$$

Τα πλήρη αποτελέσματα της μελέτης θα πρέπει να αναλυθούν και να δοθούν σε μια ολοκληρωμένη αναφορά (report). Επιπλέον, θα πρέπει να περιγράφεται η λειτουργία του κώδικα, ο οποίος θα πρέπει να παραδοθεί μαζί με την αναφορά και θα πρέπει να παράγει τα ίδια αποτελέσματα με τα αναφερθέντα στην αναφορά. Η ποιότητα των λύσεων θα πρέπει να διαπιστώνεται από τις παρεχόμενες τιμές σφάλματος (δοσμένες σε πίνακες, γραφήματα κ.λ.π.) καθώς και με γραφική αναπαράσταση του βέλτιστου μοντέλου και των αντίστοιχων δεδομένων, όπως η παρακάτω, η οποία απεικονίζει το βέλτιστο μοντέλο.



Τέλος, θα πρέπει να χρησιμοποιηθεί το μοντέλο προκειμένου να προβλεφθούν οι τιμές των κρουσμάτων για 5 ημέρες μετά από το πέρας των δεδομένων, δηλαδή από 13 έως 17 Νοεμβρίου 2020 (θα βρείτε τις τιμές αυτές στην ιστοσελίδα <https://covid19.who.int/table> μετά από τις 18 Νοεμβρίου 2020) και να εξαχθούν συμπεράσματα ως προς την ποιότητα πρόβλεψης του μοντέλου.

Βασικές παρατηρήσεις και υποδείξεις:

- Θα πρέπει να υπολογίσετε αρχικά τις μερικές παραγώγους 1^{ης} και 2^{ης} τάξης της αντικειμενικής συνάρτησης $f(\alpha)$ ως προς τις μεταβλητές της, δηλαδή ως προς τα $\alpha_0, \alpha_1, \dots, \alpha_4$. Αυτές θα σας χρειαστούν στον υπολογισμό των gradients και του Εσσιανού πίνακα στη μέθοδο Newton.
- Θα πρέπει να χρησιμοποιήσετε κάποια μέθοδο διόρθωσης του Εσσιανού πίνακα αν αυτός δεν είναι θετικά ορισμένος.
- Για τις υλοποιήσεις στηριχτείτε στους ψευδοκώδικες που δόθηκαν στο μάθημα και χρησιμοποιήστε τις αντίστοιχες default τιμές των παραμέτρων.
- Κάθε μέθοδος θα πρέπει να εκτελεστεί το πολύ για 1000 επαναλήψεις.
- Επειδή τα x_i (ημέρες) και y_i (κρούσματα) στα δεδομένα έχουν σημαντικά διαφορετική κλίμακα, καλό είναι να αλλαχθεί η κλίμακά τους. Σας συνιστώ να διαιρέσετε τα x_i με το 10 και τα y_i με το 10000 προκειμένου να υπολογίσετε το μοντέλο. Μην παραλείψετε να επαναφέρετε τις τιμές αυτών αλλά και του μοντέλου στην αρχική κλίμακα προκειμένου να κάνετε τα γραφήματα.
- Οι υλοποιήσεις μπορούν να γίνουν σε γλώσσα προγραμματισμού της επιλογής σας ανάμεσα στις ακόλουθες: C/C++, Java, Fortran, Matlab/Octave, Python, R.

Ημερομηνία παράδοσης: **30 Νοεμβρίου 2020**

Η αξιολόγηση των εργασιών θα βασίζεται στην ποιότητα της αναφοράς, του κώδικα και στην αξιοποίηση γνώσεων από το μάθημα.