

Υπολογιστική Φυσική Στερεάς Κατάστασης

Problem 6

Haka Kevin

Σε αυτή την άσκηση θα υπολογίσουμε μια στατιστική ποσότητα πολύ χρήσιμη στην ανάλυση δεδομένων, καθώς εκφράζει την γραμμική συσχέτιση αναμεσα σε δυο μεταβλητές (τον συντελεστή Pearson). Αυτό που πρόκειται να κάνουμε είναι να βρούμε την συσχέτιση ανά δυο ανάμεσα στις τιμές κλεισίματος 20 μετοχών του χρηματιστηρίου Αθηνών το έτος 2004 (253 τιμές ανά μετοχή) και στην συνέχεια να φτιάξουμε τον πίνακα συσχέτισης.

Τα βήματα που ακολουθούμε είναι τα παρακάτω:

1. Εισάγουμε τις τιμές από κάποιο αρχείο.
2. Λογαριθμούμε όλες τις τιμές.
3. Υπολογίζουμε την επιστροφή, δηλαδή, από κάθε επόμενη λογαριθμιμένη τιμή αφαιρούμε την προηγούμενη λογαριθμιμένη τιμή. (Απ' αυτό θα χάσουμε ένα δεδομένο ανά μετοχή καθώς η πρώτη τιμή της μετοχής δεν έχει προηγούμενο.)
4. Στη συνέχεια υπολογίζουμε τον συντελεστή Pearson από την παρακάτω σχέση. Όπου x και y οι τιμές επιστροφής του προηγούμενου βήματος για δυο μετοχές. Το κάνουμε αυτό για όλους τους πιθανούς συνδυασμούς μετοχών.

$$p = \frac{\sum_{i=1}^n (x_i y_i) - n * \bar{x} * \bar{y}}{\sqrt{(\sum_{i=1}^n (x_i^2) - n * \bar{x}^2)(\sum_{i=1}^n (y_i^2) - n * \bar{y}^2)}}$$

5. Τέλος για τον υπολογισμό του πίνακα αποστάσεων εφαρμόζουμε τον παρακάτω μετασχηματισμό.

$$d = \sqrt{2 * (1 - p)}$$

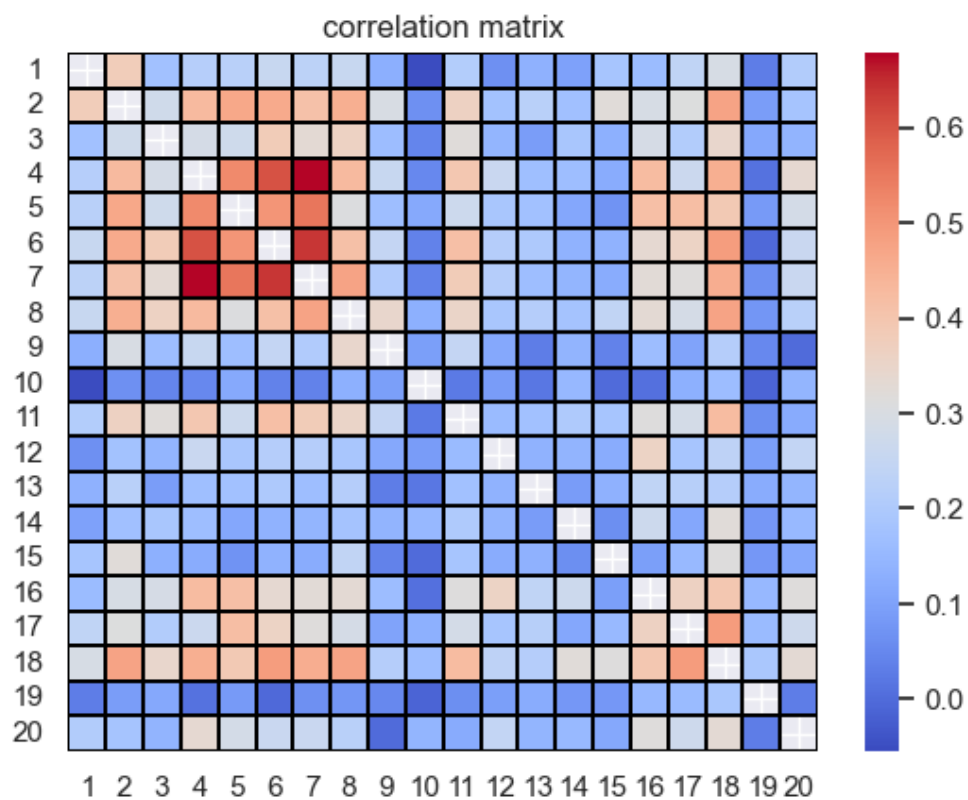


Figure 1 Στο σχήμα αυτό βλέπουμε τον πίνακα συσχέτιση.

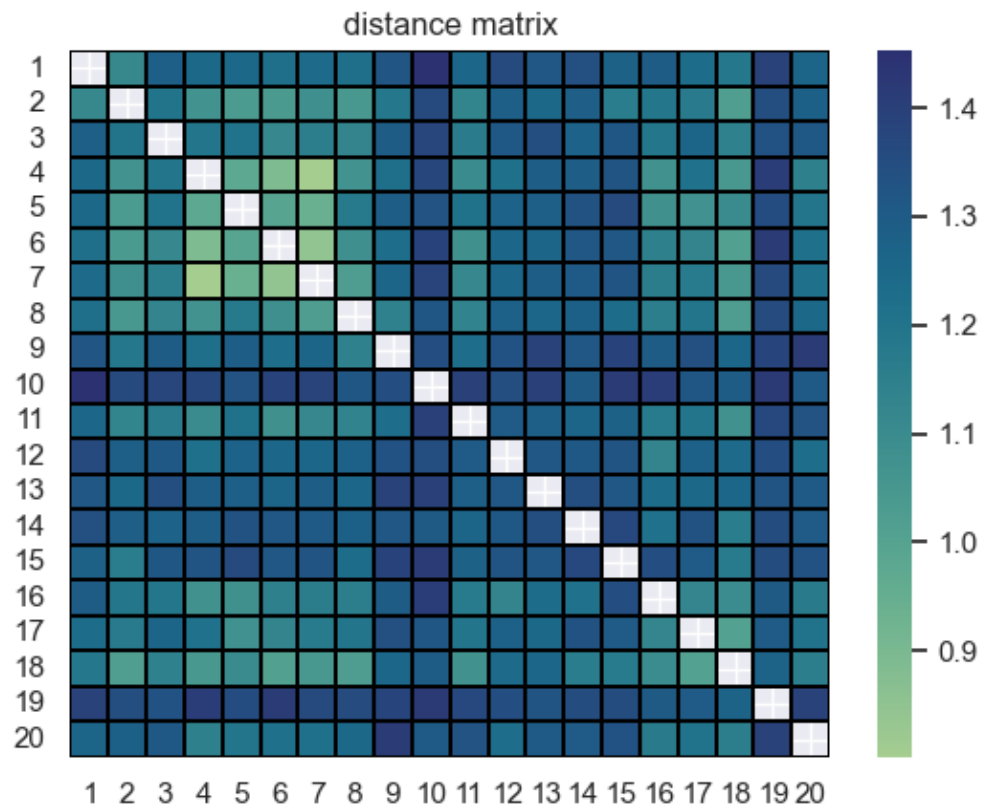


Figure 2 Στο σχήμα αυτό βλέπουμε τον πίνακα απόστασης.