Πανεπιστήμιο Ιωαννίνων – Πολυτεχνική Σχολή
Τμήμα Μηχανικών Η/Υ και Πληροφορικής

Μεταπτυχιακό Μάθημα
(Δ3) ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ
Ακαδημαϊκό Έτος 2023-2024

[**Εργασία 1**]

1 Εισαγωγή

Ένας επενδυτής θέλει να επενδύσει σε επενδυτικά προϊόντα των οποίων οι αποδόσεις σχετίζονται με 4 εθνικούς εμπορικούς κλάδους των ΗΠΑ:

(Κ1) Πωλήσεις υπολογιστών-ηλεκτρονικών ειδών.

(Κ2) Πωλήσεις αμυντικού εξοπλισμού.

(Κ3) Πωλήσεις κινητήρων οχημάτων και εξαρτημάτων.

(Κ4) Πωλήσεις μετάλλων.

Τα συγκεκριμένα επενδυτικά προϊόντα προσφέρουν μηνιαίως τόση απόδοση όση και η ποσοστιαία μεταβολή των εξυπηρετηθέντων παραγγελιών (σε \$ εκατ.), συνολικά σε εθνικό επίπεδο ΗΠΑ, μεταξύ της στιγμής της επένδυσης (μήνας έναρξης) και της στιγμής αποχώρησης (μήνας λήξης) από το επενδυτικό προϊόν. Για παράδειγμα, αν επενδυθεί \$1 στον κλάδο των υπολογιστών-ηλεκτρονικών ειδών τον μήνα t όπου ο συγκεκριμένος κλάδος είχε πωλήσεις \$17950 εκατ. και ο επενδυτής αποχωρήσει τον μήνα $t+k$ όπου ο κλάδος κατέγραψε πωλήσεις \$21658 εκατ., τότε θα του επιστραφεί:

$$\frac{\$21658}{\$17950} \approx \$1.20657$$

δηλαδή η απόδοση της επένδυσής του είναι περίπου στο 20.66% για το χρονικό διάστημα των k μηνών. Αντίθετα, αν αποχωρήσει σε κάποια άλλη χρονική στιγμή $t+m$ όπου ο κλάδος κατέγραψε πωλήσεις \$16108 εκατ., τότε θα του επιστραφεί:

$$\frac{\$16108}{\$17950} \approx \$0.89738$$

δηλαδή η απόδοση της επένδυσής του σε αυτή την περίπτωση είναι περίπου στο -10.26% για τους m μήνες.

Στα παρεχόμενα αρχεία δεδομένων με ονόματα:

(Α1) `data_ComputersElectronicProducts.csv`

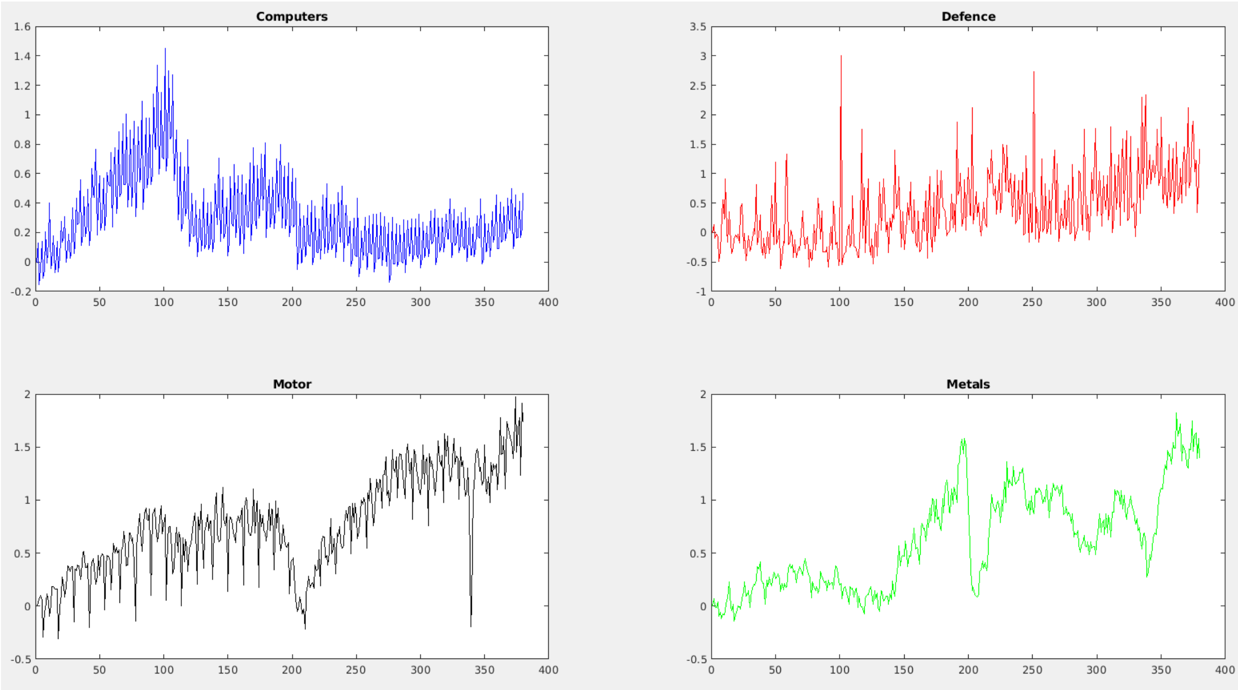
(Α2) `data_DefenseCapitalGoods.csv`

(Α3) `data_MotorVehiclesParts.csv`

(Α4) `data_PrimaryMetals.csv`

δίνονται, αντίστοιχα για καθέναν από τους 4 εμπορικούς κλάδους, οι μηνιαίες τιμές (σε \$ εκατ.) των εξυπηρετηθέντων παραγγελιών στις ΗΠΑ για το διάστημα μεταξύ Φεβρουαρίου 1992 και Σεπτεμβρίου 2023, δηλαδή 380 μηνιαίες τιμές για κάθε κλάδο.

Ο στόχος της παρούσας εργασίας είναι να βοηθήσει τον επενδυτή να αποφασίσει τι ποσοστό του κεφαλαίου του θα πρέπει να επενδύσει σε καθέναν από τους 4 εμπορικούς κλάδους Κ1-Κ4 προκειμένου να μεγιστοποιήσει την αναμενόμενη σχετική απόδοση της επένδυσής του,



Σχήμα 1: Τιμές σχετικής απόδοσης ως προς την τιμή βάσης (1ος μήνα).

ελαχιστοποιώντας ταυτόχρονα το ρίσκο που εκφράζεται ως η αλληλοσυσχέτιση των τιμών απόδοσης των 4 κλάδων. Για τον σκοπό αυτό θα γίνει κατάλληλη βελτιστοποίηση που θα χρησιμοποιήσει όλα τα παρεχόμενα ιστορικά δεδομένα. Θα υποθέσουμε ότι η χρονική στιγμή έναρξης της επένδυσης γίνεται αμέσως μετά την πάροδο των 380 διαθέσιμων μηνών.

2 Αναλυτική περιγραφή

Οι τιμές που δίνονται σε καθένα από τα αρχεία A1-A4 αφορούν στις μηνιαίες πωλήσεις των εμπορικών κλάδων K1-K4 σε $ εκατ. Αν θεωρηθεί ως αρχική στιγμή του χρονικού ορίζοντα ο Φεβρουάριος 1992 (1η καταχωρημένη τιμή σε κάθε αρχείο) τότε μπορούμε να μετατρέψουμε όλες τις δοθείσες τιμές σε αντίστοιχες τιμές σχετικής απόδοσης ως προς την τιμή βάσης του 1ου μήνα ως εξής:

$$R_j^{(i)} = \frac{\text{τιμή μήνα } j \text{ κλάδου } Ki}{\text{τιμή μήνα } 1 \text{ κλάδου } Ki} - 1,$$

για κάθε $j = 1, 2, \ldots, 380$, και $i = 1, 2, 3, 4$. Στο Σχήμα 1 δίνονται οι ληφθείσες τιμές σχετικής απόδοσης $R_j^{(i)}$ και στις 4 περιπτώσεις. Ας συμβολίσουμε, αντίστοιχα, με $R^{(i)}$ το διάνυσμα των σχετικών αποδόσεων:

$$R^{(i)} = \left[R_1^{(i)}, R_2^{(i)}, \ldots, R_{380}^{(i)}\right],$$

και με $\bar{R}^{(i)}$ τη μέση τιμή της σχετικής απόδοσης:

$$\bar{R}^{(i)} = \text{mean}\left(R^{(i)}\right) = \frac{1}{380}\sum_{j=1}^{380} R_j^{(i)},$$

για κάθε $i = 1, 2, 3, 4$. Επίσης, έστω ο πίνακας συνδιασπορών (covariance matrix) των τιμών σχετικής απόδοσης:

$$M = \begin{pmatrix} \text{cov}\left(R^{(1)}, R^{(1)}\right) & \text{cov}\left(R^{(1)}, R^{(2)}\right) & \text{cov}\left(R^{(1)}, R^{(3)}\right) & \text{cov}\left(R^{(1)}, R^{(4)}\right) \\ \text{cov}\left(R^{(2)}, R^{(1)}\right) & \text{cov}\left(R^{(2)}, R^{(2)}\right) & \text{cov}\left(R^{(2)}, R^{(3)}\right) & \text{cov}\left(R^{(2)}, R^{(4)}\right) \\ \text{cov}\left(R^{(3)}, R^{(1)}\right) & \text{cov}\left(R^{(3)}, R^{(2)}\right) & \text{cov}\left(R^{(3)}, R^{(3)}\right) & \text{cov}\left(R^{(3)}, R^{(4)}\right) \\ \text{cov}\left(R^{(4)}, R^{(1)}\right) & \text{cov}\left(R^{(4)}, R^{(2)}\right) & \text{cov}\left(R^{(4)}, R^{(3)}\right) & \text{cov}\left(R^{(4)}, R^{(4)}\right) \end{pmatrix}$$

όπου $\mathrm{cov}\left(R^{(i)}, R^{(j)}\right)$ η συνδιασπορά των $R^{(i)}$, $R^{(j)}$, $i, j = 1, 2, 3, 4$.

Αν συμβολίσουμε με $w_i \in [0.0, 1.0]$, $i = 1, 2, 3, 4$, το ποσοστό του κεφαλαίου που επενδύεται στους αντίστοιχους εμπορικούς κλάδους K1-K4, τότε ο στόχος του προβλήματος είναι να μεγιστοποιηθεί η αναμενόμενη απόδοση και ταυτόχρονα να ελαχιστοποιηθεί το ρίσκο. Δηλαδή πρέπει να επιλυθεί το πρόβλημα μεγιστοποίησης:

$$\max_{w}\ f(w) = \underbrace{w^T \bar{R}}_{\text{απόδοση}} - \lambda\ \underbrace{w^T M\, w}_{\text{ρίσκο}},$$

όπου $w = (w_1, w_2, w_3, w_4)$ είναι τα ποσοστά συμμετοχής ανά εμπορικό κλάδο, $\lambda > 0$ είναι μια παράμετρος που καθορίζει τη σημαντικότητα του ρίσκου και:

$$\bar{R} = \left[\bar{R}^{(1)}, \bar{R}^{(2)}, \bar{R}^{(3)}, \bar{R}^{(4)}\right],$$

είναι το διάνυσμα των μέσων αποδόσεων. Το παραπάνω πρόβλημα μπορεί να γραφτεί ως πρόβλημα ελαχιστοποίησης και σε πιο αναλυτική μορφή ως εξής:

$$\min_{w}\ F(w) = -\left(\sum_{i=1}^{4} w_i\, \bar{R}^{(i)} - \lambda \sum_{j=1}^{4}\sum_{k=1}^{4} w_j\, w_k\, \mathrm{cov}\left(R^{(j)}, R^{(k)}\right)\right).$$

Προφανώς, τα ποσοστά w_i συμμετοχής στους κλάδους K1-K4 θα πρέπει να υπακούν στον περιορισμό:

$$\sum_{i=1}^{4} w_i = 1.$$

Επειδή η ενσωμάτωση τέτοιων περιορισμών προς το παρόν δεν είναι εφικτή, μπορεί κανείς να θεωρήσει ως μεταβλητές του προβλήματος τις:

$$x_1, x_2, x_3, x_4 \in [0.0, 1.0],$$

από τις οποίες θα προκύπτουν τα ποσοστά ως εξής:

$$w_i = \frac{x_i}{\sum\limits_{j=1}^{4} x_j}, \quad i = 1, 2, 3, 4.$$

Επομένως, η αντικειμενική συνάρτηση $F(w)$ γίνεται πλέον συνάρτηση των x_1, x_2, x_3, x_4, οι οποίες είναι οι μεταβλητές ως προς τις οποίες θα κάνουμε τη βελτιστοποίηση εντός του χώρου αναζήτησης $X \triangleq [0.0, 1.0]^4$. Για τις ανάγκες της παρούσας εργασίας θα θεωρήσουμε ότι το $\lambda = 1.5$ στην $F(w)$.

3 Εφαρμογή

Η επίλυση του προβλήματος βελτιστοποίησης θα πρέπει να γίνει με τους ακόλουθους αλγορίθμους βελτιστοποίησης:

(1) BFGS με ευθύγραμμη αναζήτηση με συνθήκες Wolfe.

(2) Dogleg BFGS.

(3) Μέθοδος συζυγών κλίσεων Polak-Ribiere.

Οι αλγόριθμοι θα πρέπει να υλοποιηθούν πλήρως από τον φοιτητή. Ωστόσο, για τον υπολογισμό των απαιτούμενων ποσοτήτων που αφορούν στα δεδομένα (σχετικές αποδόσεις, μέσες τιμές, συνδιασπορές) μπορούν να χρησιμοποιηθούν έτοιμες ρουτίνες (π.χ. από Matlab, Python, R ή άλλες). Θα πρέπει να δοθεί ιδιαίτερη προσοχή στον σωστό υπολογισμό των παραγώγων της αντικειμενικής συνάρτησης. Κάθε αλγόριθμος θα πρέπει να εκτελεστεί από 10 ανεξάρτητα και τυχαία επιλεγμένα αρχικά σημεία, με το ίδιο υπολογιστικό κόστος (πλήθος συναρτησιακών υπολογισμών) για όλα. Στα αποτελέσματα θα πρέπει να καταγράφεται η ευρεθείσα λύση (ποσοστά συμμετοχής κεφαλαίου), η τιμή της αναμενόμενης απόδοσης, του ρίσκου, καθώς και η συνολική τιμή της αντικειμενικής συνάρτησης.

Τα κριτήρια βαθμολόγησης της εργασίας είναι:

- Η κατάλληλη εφαρμογή και παραμετροποίηση των αλγορίθμων.
- Η ποιότητα της αναφοράς και της επιστημονικής ανάλυσης.
- Η ποιότητα του κώδικα.
- Η αξιοποίηση γνώσεων από το μάθημα.
- Η πρωτοβουλία και αυτονομία του φοιτητή.

Τα παραδοτέα της εργασίας είναι:

(1) **Αναλυτική αναφορά**: εκτενές κείμενο στο οποίο θα επεξηγείται επακριβώς η μεθοδολογία, όλες οι λεπτομέρειες εφαρμογής της και τα αποτελέσματα σε πλήρη παρουσίαση και σχολιασμό.

(2) **Κώδικας**: ο σχετικός κώδικας και οδηγίες εφαρμογής του. Θα πρέπει να αποθηκεύσετε όλες τις λεπτομέρειες που είναι αναγκαίες για πλήρη αναπαραγωγή των πειραμάτων σας (π.χ. αρχικά σημεία, συνθήκη τερματισμού κ.λ.π.).

Όλα τα παραπάνω θα πρέπει να αποσταλούν σε ένα αρχείο ZIP μέχρι την παρακάτω καταληκτική ημερομηνία. Οι υλοποιήσεις μπορούν να γίνουν σε γλώσσα προγραμματισμού της επιλογής σας ανάμεσα στις ακόλουθες: C/C++, Java, Fortran, Matlab/Octave, Python, R.

Ημερομηνία παράδοσης: 22 Δεκεμβρίου 2023