

Μέθοδοι Μηχανικής Μάθησης στη Χρηματοοικονομική

Εργασία 3

Καταληκτική ημερομηνία: Κυριακή 28 Μαΐου. Καλείστε να παραδώσετε ένα pdf αρχείο με τις απαντήσεις σας σε όλες τις ερωτήσεις. Στις απαντήσεις σας παρακαλείστε να έχετε βασικά screenshots από το excel/python output. Το pdf αρχείο θα πρέπει να συνοδεύεται και από ένα zip αρχείο με τους excel/python κώδικές σας.

Ενότητα 5: Supervised Learning: Linear Regression

1. Επιθυμείτε να βρείτε ένα μοντέλο προβλεπτικής ικανότητας για την πρόβλεψη των αποδόσεων του Χρηματιστηριακού Δείκτη των Η.Π.Α. (S&P500). Οι μεταβλητές που έχει δείξει η ακαδημαϊκή θεωρία ότι είναι στατιστικά σημαντικές μεταβλητές στην πρόβλεψη αποδόσεων μετοχικών δεικτών είναι στο excel *PredictorData2022.xlsx*. Εκτιμήστε το παρακάτω προβλεπτικό μοντέλο με regularization (RIDGE και LASSO),

$$R_{t+1} = \beta_0 + \beta_1 DY_t + \beta_2 EP_t + \beta_3 (b/m)_t + \beta_4 tbl_t + \beta_5 dfr_t + \beta_6 lty_t + \beta_7 ts_t + \beta_8 ntis_t + \beta_9 infl_t + \beta_{10} ltr_t + \beta_{11} svar_t + \varepsilon_{t+1}$$

όπου R είναι οι αποδόσεις του Χρηματιστηριακού Δείκτη των Η.Π.Α. (S&P500) και 11 μεταβλητές (predictor variables) ορίζονται στο αρχείο excel. Αναλύστε το μοντέλο που θα επιλέγατε.

Ενότητα 6: Supervised Learning: Logit Model

2. Βασιζόμενοι στην Εφαρμογή LendingClub, επιπλέον των τεσσάρων χαρακτηριστικών που έχουμε χρησιμοποιήσει για classification καλών δανείων και δανείων αθέτησης, καλείστε να προσθέσετε δύο επιπλέον μεταβλητές της επιλογής σας και εκτιμήστε ξανά ένα logit model. Θα κάνατε χρήση του regularization; Το excel με τις μεταβλητές είναι *lending_clubFull_Data_Set.xlsx* και η περιγραφή των μεταβλητών βρίσκεται στο αρχείο *lendingclub_datadictionary.xlsx*. Σχολιάστε την αποδοτικότητα του μοντέλου που εσείς θεωρείτε καλύτερο.