Μέθοδοι Μηχανικής Μάθησης στη Χρηματοοικονομική

Οικονομικό Πανεπιστήμιο Αθηνών

Τμήμα Λογιστικής και Χρηματοοικονομικής

2022-2023

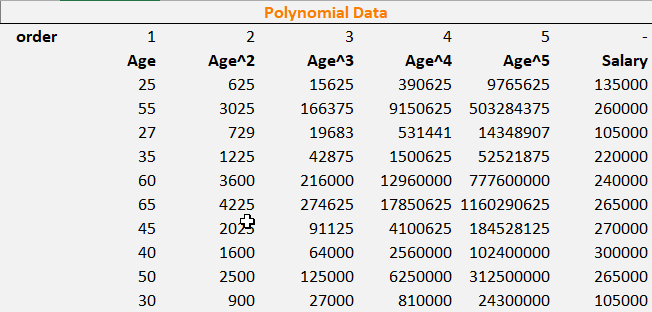
Εργασία 1

|  |  |
| --- | --- |
| Ονοματεπώνυμο Φοιτητή |  |
| Αριθμός Μητρώου |  |
| Ημερομηνία Παράδοσης |  |

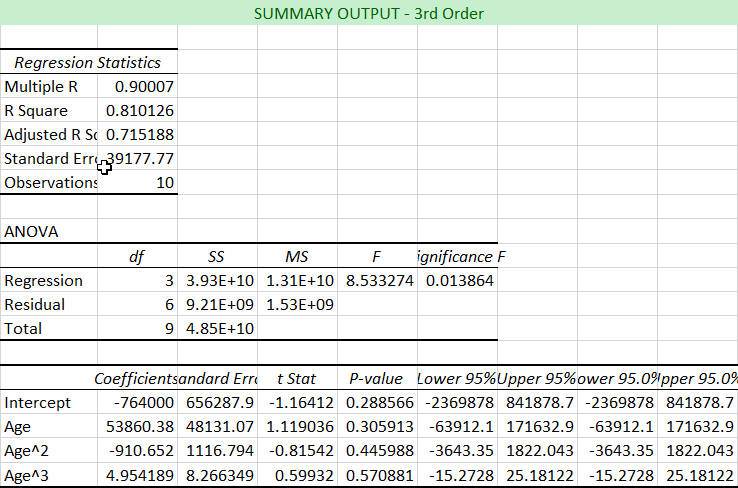
# Άσκηση 1

Για το σκοπό της άσκησης, δημιουργήθηκε ένα Excel Workbook αρχείο με 5 worksheets, Data, Coefficients, Training Analysis, Validation Analysis και Test Analysis. Στο πρώτο βρίσκονται τρια datasets με Age και Salary δεδομένα. Το training dataset περιέχει τα δεδομένα της εκπαίδευσης, το validation dataset περιέχει μερικά δεδομένα από το training dataset αλλά και «ξένα», ενώ το test dataset περιέχει κυρίως «ξένα» δεδομένα.

Στο δεύτερο υλοποιήθηκαν τα βήματα για τον υπολογισμό των συντελεστών της πολυωνυμικής συνάρτησης. Πιο συγκεκριμένα, αναπτύχθηκε το Age σε πολυωνυμικούς όρους (Εικ. 1) και στη συνέχεια εκτελέστηκε γραμμική παλινδρόμιση για κάθε βαθμό (2, 3, 4 και 5) με το εργαλείο Analysis ToolPak (Εικ 2.).



Εικόνα 1 Πολυωνυμικοί όροι του Age



Εικόνα 2 Ενδεικτικά αποτελέσματα Regression του εργαλείου Analysis ToolPak για πολυώνυμο 3ου βαθμού.

Αξίζει να σημειωθεί πως το πολυώνυμο 5ου βαθμού παρουσίασε την μεγαλύτερη τιμή R Square (Εικ 3.) πράγμα το οποίο σημαίνει πως οι ανεξάρτητες μεταβλητές της συνάρτησης (Age, Age2,...) εκφράζουν σε μεγάλο βαθμό τη διακύμανση της εξαρτημένης μεταβλητης (Salary). Σε αυτό το σημείο πρέπει να δοθεί προσοχή αφού οι τιμές του R Square έχουν υπολογισθεί με βάση το σύνολο δεδομένων εκπαίδευσης και όχι επικύρωσης και τα «φανταχτερά» αποτελέσματα ίσως κρύβουν φαινόμενο overfitting.

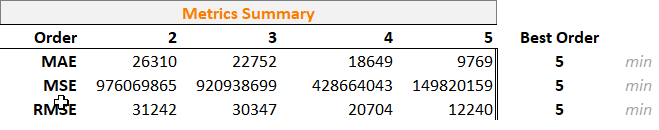


Εικόνα 3 Συγκριτικά αποτελέσματα R Square

Αφού υπολογίσθηκαν οι συντελεστές για κάθε πολυώνυμο, δημιουργήθηκε το 3ο υπολογιστικό φύλλο, Training Analysis, στο οποίο υπολογίσθηκαν οι εκτιμήσεις του Salary για κάθε πολυωνυμική συνάρτηση που υπολογήσθηκε στο προηγούμενο βήμα, χρησιμοποιώντας τα δεδομένα training. Επιπλέον, υπολογίσθηκαν το μέσο πραγματικό σφάλμα (Mean Error), το μέσο απόλυτο σφάλμα (Mean Absolute Error), το μέσο τετραγωνικό σφάλμα (Mean Squared Error), καθώς και τη ρίζα μέσου τετραγωνικού σφάλματος (Root Mean Square Error). Επιπλέον, δημιουργήθηκαν βοηθητικά γραφήματα για την οπτικοποιήση των εκτιμήσεων.

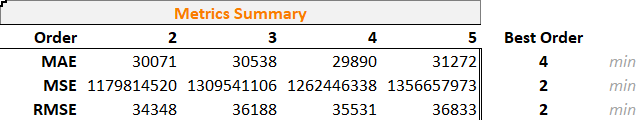
|  |  |
| --- | --- |
| Εικόνα 4 Εκτιμήσεις και σφάλματα για το πολυώνυμο 4ου βαθμού | Εικόνα 5 Γράφημα πολυωνύμου και πραγματικών τιμών Salary |

Συγκεντρωτικά, το πολυώνυμο 5ου βαθμού φαίνεται να αποδίδει καλύτερα στο σύνολο εκπαίδευσης με RMSE ίσο με 12,240$, αρκετά μικρότερο σφάλμα σε σχέση με το 31,242$ του 2ου βαθμού πολυωνύμου. Την δεύτερη θέση παίρνει το πολυώνυμο 4ου βαθμού με 20,704$.



Εικόνα 6 Συγκεντρωτικά αποτελέσματα μετρικών για κάθε πολυώνυμο επί training δεδομένων.

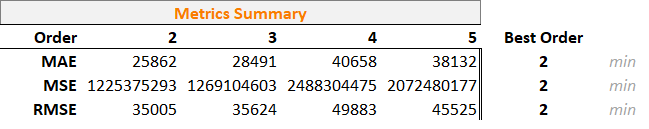
Το ίδιο πρότυπο ακολουθήθηκε και στο 4ο υπολογιστικό φύλλο, Validation Analysis. Χρησιμοποιήθηκαν οι συντελεστές του φύλλου Coefficients και το validation dataset, ώστε να υπολογισθούν οι εκτιμήσεις του Salary. Αξίζει να επικεντρωθούμε πάλι στον συγκεντρωτικό πίνακα (Εικ. 7).



Εικόνα 7 Συγκεντρωτικά αποτελέσματα μετρικών για κάθε πολυώνυμο επί validation δεδομένων.

Ενώ στο training dataset το πολυώνυμο 5ου βαθμού απέδιδε καλύτερα, στην περίπτωση του validation dataset το πολυώνυμο 2ου βαθμού παίρνει την πρωτιά. Αυτή τη φορά, το RMSE για το 2ου βαθμου παραμένει περίπου σταθερό σε σχέση με το training analysis, ίσο με 34,348$, ενώ το 5ου βαθμού αυξάνει το σφάλμα του στο χειρότερο μεταξύ των πολυονύμων, ίσο με 36,833$. Από τα παραπάνω μπορούμε να ισχυριστούμε πως το πολυώνυμο 5ου βαθμού εμφανίζει φαινόμενο overfitting, αφού αποδίδει καλά στα δεδομένα εκπαίδευσης αλλά δεν γενικεύει σε άγνωστα δεδομένα. Σε αντίθεση το 2ου βαθμού, έχει μια σταθερά «μέτρια[[1]](#footnote-1)» απόδοση σε σχέση με τα υπόλοιπα και στα δύο σύνολα δεδομένων. Τα 3ου και 4ου εμφανίζουν παρόμοια «μέτρια» απόδοση αλλά με αρνητική χροιά.

Αν και η εκφώνιση ζητάει ανάλυση σε training και validation datasets, φάνηκε χρήσιμο να γίνει μία ακόμα ανάλυση με το διαθέσιμο test dataset το οποίο περιέχει ελάχιστα (2/10) κοινά δεδομένα με το training.



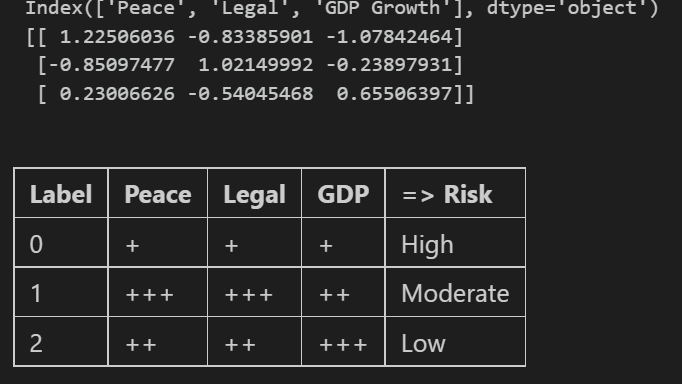
Εικόνα 8 Συγκεντρωτικά αποτελέσματα μετρικών για κάθε πολυώνυμο επί test δεδομένων.

Στα «αγνώστα» δεδομένα παρατηρούμε πως το 2ου βαθμού πολυώνυμο παραμένει στη σταθερά «μέτρια» απόδοση, αλλά βέλτιστη μεταξύ των υπολοίπων, ενώ παράλληλα, ικανοποιητικά είναι και τα αποτελέσματα του πολυωνύμο 3ου βαθμού. Το 4ου καταλαμβάνει την χαμηλότερη θέση απόδοσης, βάσει RMSE.

Ολοκληρώνοντας, το πολυώνυμο 2ου βαθμού φαίνεται να αποδίδει καλύτερα από όλα, ενώ ακολουθεί το 3ου με παρόμοια απόδοση. Ίσως ένα μεγαλύτερο σύνολο δεδομένων να διευρύνει τη διάκρισή τους. Το 4ου βαθμού δεν είναι άξιο αναφοράς, ενώ το 5ου, παρόλο που πήγε να «ξεγελάσει» με το ιδιαίτερα χαμηλό RMSE στο training dataset, κατέληξε να απογοητεύσει το τέλος στα άγνωστα δεδομένα.

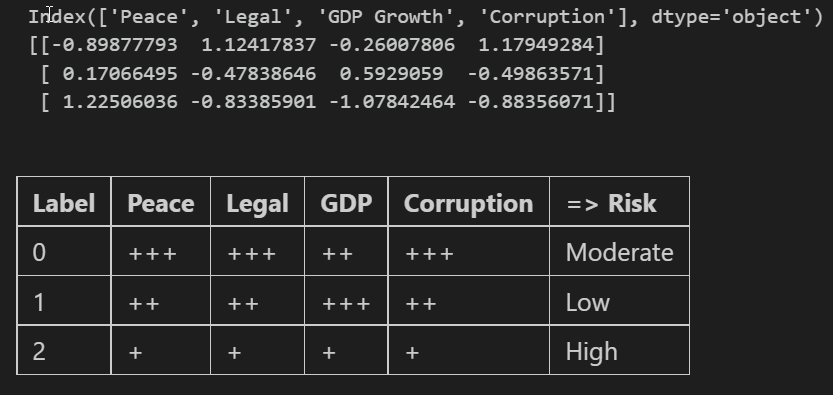
# Άσκηση 2

Στην άσκηση αυτή χρησιμοποιήθηκε το notebook *4.1 K-means\_elbowmethod.ipynb* για την παραγωγή των συστάδων με k-means, k=3 και 3 χαρακτηριστικά και το notebook *4.1 K-means\_elbowmethod\_4.ipynb* για την παραγωγή των συστάδων με k-means, k=3 και 4 χαρακτηριστικά. Για την διευκόλυνσή μας, μελετήθηκαν τα κεντροειδή και ανατέθηκαν ετικέτες, Low, Moderate και High Risk. Πιο συγκεκριμένα, στην πρώτη περίπτωση, k=3, παρήχθηκαν τα κεντροειδή της εικόνας 9 και πραγματοποιήθηκε η αντίστοιχη ανάθεση ετικετών. Να σημειωθεί πως για την ανάθεση του ρίσκου δίνουμε μεγαλύτερη βαρύτητα στο GDP Growth από ότι στα Peace και Legal, εξού και η ανάθεση ετικέτας *Moderate* στην συστάδα *1* με «πολύ καλό» Peace και Legal και «καλό» GDP και *Low* στην συστάδα *2* με «καλό» Peace και Legal και «πολύ καλό» GDP.



Εικόνα 9 Κεντροειδή συστάδων με 3 χαρακτηριστικά και η ανάθεση των ετικετών.

Όμοίως, τα κεντροειδή και η ανάθεση στην περίπτωση των τεσσάρων χαρακτηριστικών φαίνονται στην εικόνα 10. Παρατητούμε πως με την εισαγωγή του τέταρτου χαρακτηριστικού Corruption έχει δημιουργηθεί διαφορετική αντιστοίχιση Label – Risk. Επιπλέον, διατηρούμε την παραδοχή της ισχύος του GDP έναντι των υπολοίπων χαρακτηριστικών. Σε διαφορετική περίπτωση, τα δεδομένα που είχαν ετικέτα *Low* (αντ. *Moderate*) στην προηγούμενη περίπτωση τριων χαρακτηριστικών, θα είχαν σε αυτήν την περίπτωση ετικέτα *Moderate* (αντ. *Low*) και αυτό γιατί δεν αλλάζει η γενική –θα δουμε στη συνέχεια γιατί γενική– συνέπεια των συστάδων λόγω μιας αυθέρτης ανάθεσης ετικετών.

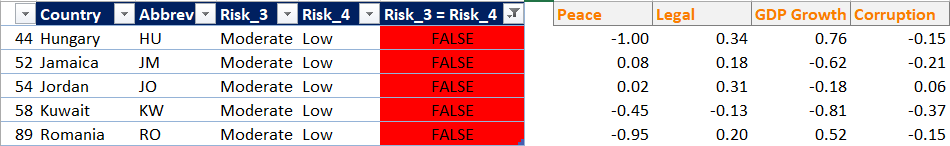


Εικόνα 10 Κεντροειδή συστάδων με 4 χαρακτηριστικά και η ανάθεση των ετικετών.

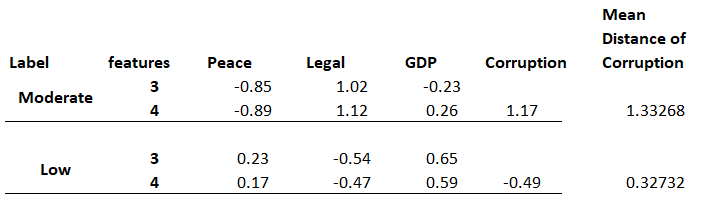
Αφού δημιουργήθηκαν τα κατάλληλα αρχεία .csv, χρησιμοποιήθηκε το εργαλείο Excel για γρήγορη ανάλυση/σύγκριση των ετικέτων των δύο περιπτώσεων. Δημιουργήθηκε πίνακας με τα στατικά χαρακτηριστικά και τις ετικέτες κάθε περίπτωσης, και φιλτράραμε τις εγγραφές κρατώντας μόνο τις ασυνέπειες. Βρέθηκαν 5 «ασυνεπείς» χώρες, οι οποίες στην πρώτη περίπτωση είχαν ετικέτα *Moderate* και στη δεύτερη περίπτωση *Low.*

Σε αυτό το σημείο αξίζει να αναφέρουμε πως οι χώρες υψηλού ρίσκου δεν διαφέρουν μεταξύ των περιπτώσεων. Όμως, μπορούμε να δούμε ότι υπάρχουν χώρες που φαινομενικά ενέχουν υψηλό ρίσκο, π.χ. η Κολομβία, αλλά έχουν χαρακτηριστεί ως χώρες χαμηλού ρίσκου. Με αυτό θα ασχοληθούμε στην επόμενη άσκηση και θα επικεντροθούμε τώρα στις υπόλοιπες ασυνέπειες.

Θα λάβουμε πιο καθαρή εικόνα αν μελετήσουμε τα κεντροειδή των ετικέτων αυτών στις δύο περιπτώσεις. Στην Εικόνα 12 παρατηρούμε πως οι συντεταγμένες των κοινών διαστάσεων (Peace, Legal, GDP) διαφέρουν μεταξύ τους μεταξύ των περιπτώσεων, αφού η εισαγωγή της τέταρτης διάστασης μετακίνησε τα κεντροειδή. Αν σκεφτούμε πως τα δεδομένα μας είναι κανονικοποιήμενα, η τάξη μεγέθους των διαφορών μπορεί να θεωρηθεί μικρή, επομένως είχαμε «μικρή» μετακίνηση στις 3 κοινές διαστάσεις. Όμως, δημιουργείται μια «μεγάλη» μετακίνηση στην τέταρτη διάσταση η οποία έχει καθοριστικό ρόλο στην ταξινόμηση των δεδομένων. Με απλά λόγια, παρόλο που η απόσταση του κεντροειδους της ετικέτας *Moderate* στα 4 χαρακτηστικά από τα υπό μελέτη δεδομένα είναι μικρή στις διαστάσεις Peace, Legal, GDP, η μεγάλη τιμή της συνεντραγμένης Corruption «κάνει» τα δεδομένα να ταξινομηθούν στην συστάδα Low με «μέτρια» απόσταση από Peace, Legal, GDP και «μικρή» Corruption. Πιο συγκεκριμενα, η μέση απόσταση του Corruption των δεδομένων από των συστάδων είναι 1,33 και 0,32 αντίστοιχα. Επισημαίνουμε πως επικεντρωνόμαστε στο Corruption αφού αυτή η διάσταςη προκάλεσε την «αναταραχή».



Εικόνα 11 "Ασυνεπείς" χώρες



Εικόνα 12 Συντεταγμένες κεντροειδών συστάδων Moderate και Low στις περιπτώσεις χαρακτηριστικών.

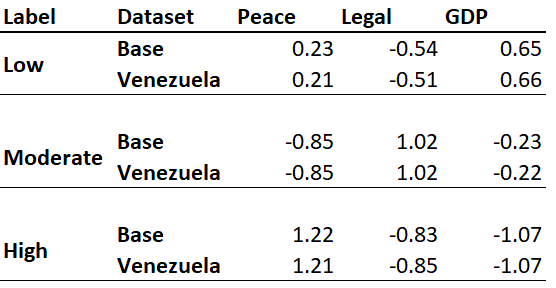
# Άσκηση 3

Στην άσκηση αυτή χρησιμοποίηθηκε πάλι το notebook *4.1 K-means\_elbowmethod.ipynb* ως βάση, ενώ το *4.1 K-means\_elbowmethod\_ven.ipynb* χρησιμοποιήθηκε για τον υπολογισμό των συστάδων συμπεριλαμβάνοντας και την Βενεζουέλα. Μεταφέροντας τα αποτελέσματα στο Excel και κάνοντας ξανά την ίδια ανάλυση ασυνέπειας, παρατηρούμε πως δεν υπάρχουν ασυνέπειες πέρα από αυτή της Βενεζουέλας αφού δεν υπάρχει στα δεδομένα βάσης. Όλες οι υπόλοιπες χώρες είναι συνέπεις ως προς τις ετικέτες ρίσκου μεταξύ των δύο συσταδοποιήσεων!



Εικόνα 13 "Ασυνέπεια" της Βενεζουέλας.

Μελετώντας τις συντεταγμένες των κεντροειδών (Εικ. 14) παρατηρούμε πως οι διαφορές είναι «μικρές». Βέβαια, αξίζει να σημειωθεί πως αυτός δεν είναι ο κανόνας αλλά έχει σημασία και η τάξη μεγέθους των δεδομένων. Στην περίπτωσή μας, η Βενεζουέλα δεν προκάλεσε κάποια τρανταχτή αλλαγή στις συστάδες.



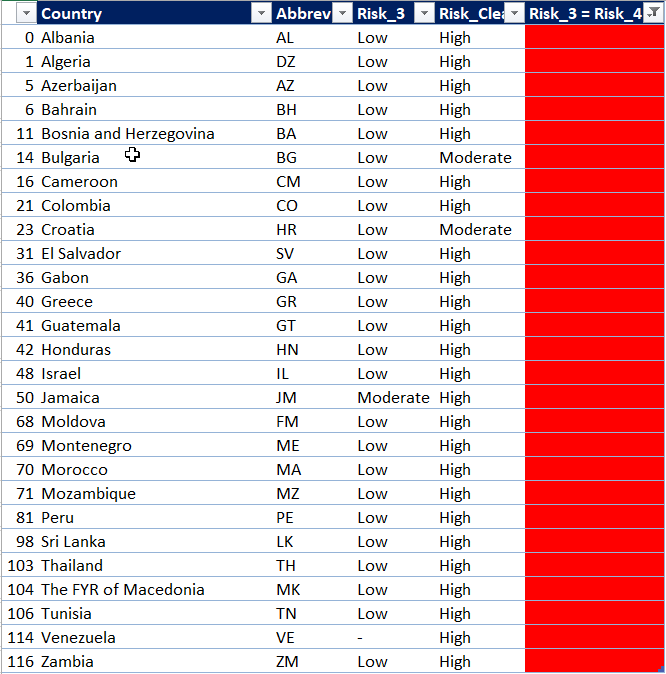
Εικόνα 14 Σύγκριση των συντεταγμένων των κεντροειδών.

Γιατί όμως γίνεται αυτό; Γενικά, τα ακραία σημεία επηρεάζουν τα κεντροειδή, αφού είναι αποτέλεσμα μέσου όρου και κατά συνέπεια επηρεάζουν τα αποτελέσματα του K-means. Αν μελετήσουμε τα διαγράμματα διασποράς (Εικ. 15) παρατηρούμε πως η Βενεζουέλα δεν είναι τόσο «ακραία» όσο κάποιες άλλες χώρες, για παράδειγμα το Ιραν (IR), Υεμένη (YE), η Αργεντινή (AR), η Νικαράγουα (NI) και η Ζιμπάμπουε (ZW). Όποτε, μπορούμε να υποθέσουμε πως τα κεντροειδή επηρεάζονται περισσότερο από αυτές τις χώρες, παρά τη Βενεζουέλα.

|  |  |  |
| --- | --- | --- |
|  |  |  |

Εικόνα 15 Διαγράμματα διασπορά συστάδων με τη Βενεζουέλα.

Για να ενισχύσουμε την υπόθεσή μας, μπορούμε να εξετάσουμε τις συστάδες που θα δημιουργήσει ο K-means αφού αφαιρέσουμε από τα δεδομένα τις προαναφερθέντες χώρες *(4.1 K-means\_elbowmethod.ipynb).* Ακολουθώντας τα ίδια βήματα, καταλήγουμε στην ανάλυση ασυνέπειας στο Excel και με έκλπηξη παρατηρούμε το τεράστιο πλήθος ασυνεπειών (Εικ. 16). Φαίνεται πως η αφαίρεση των ακραιών τιμών, άλλαξε αρκετές ετικέτες ρίσκου και συγκεκριμένα τις υποβάθμισε, από *Low* και *Moderate Risk* σε *High Risk.* Επιπλέον, μελετόντας τις χώρες βλέπουμε πως η συγκεκριμένη συσταδοποίηση φαίνεται να είναι πιο σωστή αφού χώρες που στην πραγματικότητα ενέχουν ρίσκο είχαν ταξινομηθεί σε συστάδα χαμηλού ρίσκου.



Εικόνα 16 Ανάλυση ασυνέπειας ετικετών ρίσκου σε συστοδοποίηση βάσης και συσταδοποίηση χωρίς ακραίες τιμές.

Άσκηση 4

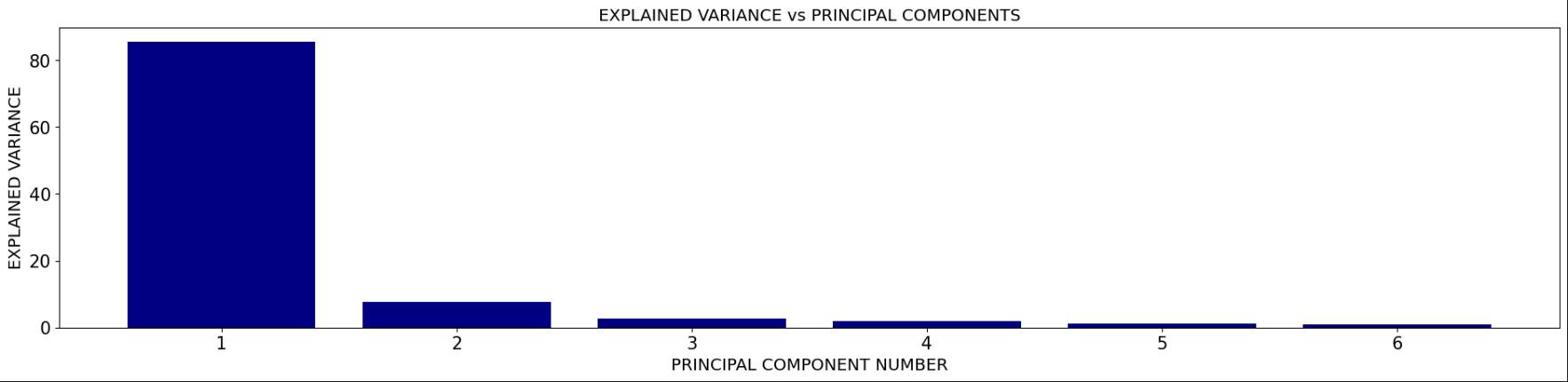
Για την άσκηση αυτή δημιουργήθηκε το notebook *PCA.ipynb,* το οποίο έχει ως βάση το αντίστοιχο notebook των σημειώσεων.

## (α)

Ακολουθώντας τα βασικά βήματα του notebook, εκτελέσαμε PCA με *n\_components = 6,* όσο δηλαδή και το πλήθος των ανεξάρτητων μεταβλητών της ανάλυσής μας, αφού πρώτα κανονικοποιήθηκαν. Σημειώνεται πως το *US RET* αποτελεί την εξαρτημένη μεταβλητή, οπότε δεν έχει νόημα να συμπεριληφθεί στην ανάλυση PCA. Από το PCA δημιουργήθηκαν τα εξής loadings.

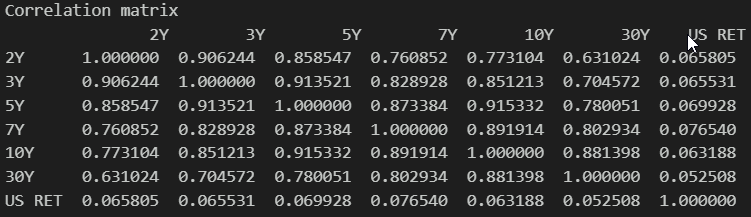
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **PC1** | **PC2** | **PC3** | **PC4** | **PC5** | **PC6** |
| **2Y** | -0.392 | -0.572 | 0.335 | 0.496 | -0.384 | -0.106 |
| **3Y** | -0.414 | -0.375 | 0.06 | -0.192 | 0.802 | -0.053 |
| **5Y** | -0.425 | -0.111 | -0.084 | -0.557 | -0.359 | 0.598 |
| **7Y** | -0.41 | 0.172 | -0.776 | 0.44 | 0.037 | 0.063 |
| **10Y** | -0.423 | 0.258 | -0.004 | -0.375 | -0.237 | -0.746 |
| **30Y** | -0.381 | 0.649 | 0.524 | 0.264 | 0.148 | 0.258 |

Ο πρώτος PC factor φαίνεται να έχει τις ίδιες σταθμίσεις για όλα τα maturities, ενώ ο δεύτερος δίνει έχει μια αύξουσα στάθμιση σε σχέση με τα χρόνια ωρίμανσης. Το ποσοστό διακύμανσης που εξηγεί κάθε PCA φαίνεται στο παρακάτω γράφημα.



Εικόνα 17 Ποσοστό διακύμανσης ανά PCA factor.

Αξιοσημείωτο είναι το γεγονός ότι ο πρώτος PCA factor εξηγεί το 86% της διακύμανσης πράγμα το οποίο υποδειγκνύει πως υπάρχει μεγάλη συσχέτιση μεταξύ των χαρακτηριστικών. Πράγματι, ο πίνακας συσχέτισης (Εικ 18) δείχνει πως τα interest rates όλων των maturities έχουν, ανά δυο, υψηλή συσχέτιση.



Εικόνα 18 Μήτρα συσχέτισης διαφορετικών λήξεων και απόδοσης δείκτη.

Με απλά λόγια, ένα χαρακτηριστικό, πχ 2Υ, εξηγεί πολύ πληροφορία για το υπόλοιπα χαρακτηριστικα, και αυτό είναι λογικό γιατί μια μεταβολή του επιτοκίου ενός 2Y, τις περισσότερες φορές θα μεταβάλει το επιτόκιο του 3Υ σε παρόμοιο, ίσως λιγότερο, βαθμό και ίδια κατεύθυνση. Έτσι, ο PCA «συμπιέζει» αυτή την πληροφορία σε λιγότερα χαρακτηριστικά-διαστάσεις, με αποτέλεσμα η πρώτη του διάσταση να περιγράφει το 86% της διακύμνασης των δεδομένων.

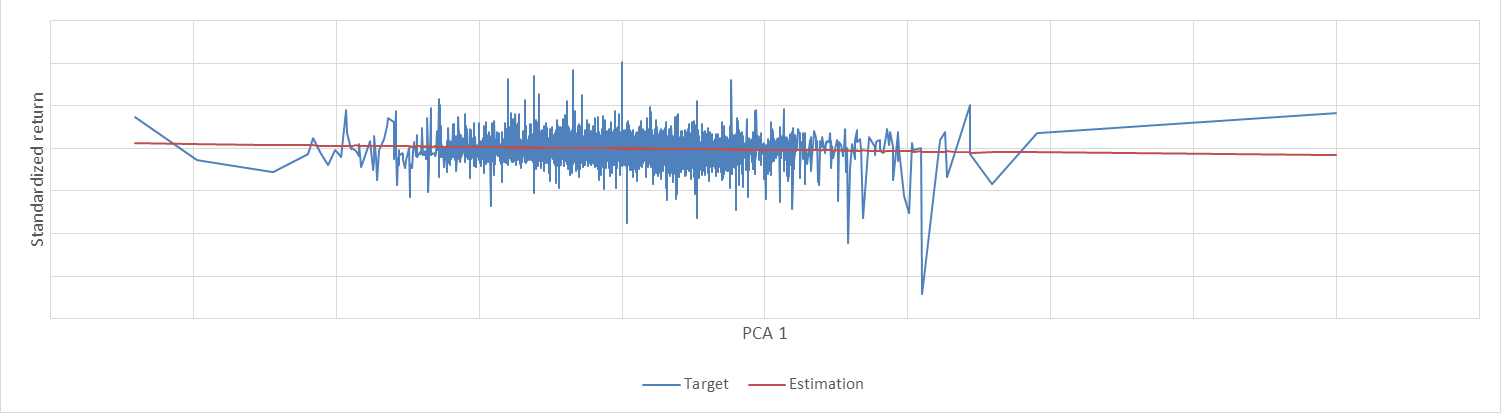
## (β)

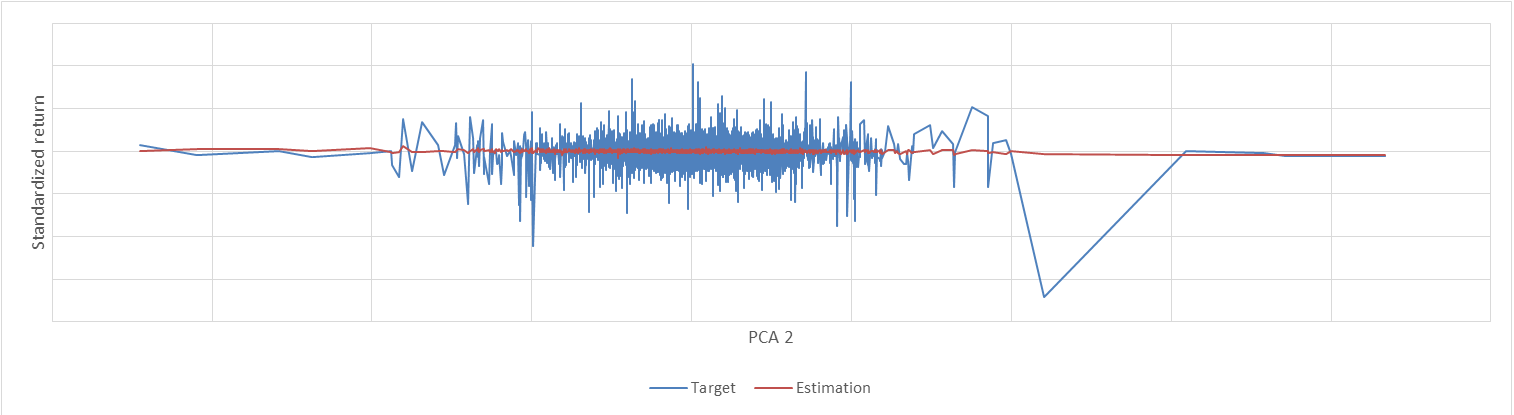
Συνήθως το πλήθος των χαρακτηριστικών είτε «ασυμπίεστων» δεδομένων, είτε παραγόντων PCA εξαρτάται από το είδος της ανάλυσης. Η ανάλυσή μας περιλαμβάνει Γραμμική Παλινδρόμηση η οποία επηρεάζεται εξαιρετικά από συσχετισμένα χαρακτηριστικά. Στην δικιά μας περίπτωση η μείωση των χαρακτηριστικών είναι μονόδρομος. Επιλέγοντας τον πρώτο παράγοντα PCA χάνουμε 14% της διακύμανσης των δεδομένων, διόλου άσχημα. Συμπεριλαμβάνοντας όμως και τον δεύτερο, ο οποίος προσφέρει 8%, χάνουμε τελικά μόνο το 6%. Φαίνεται πως η επιλογή των 2 παραγόντων PCA αρκεί για την ανάλυσή μας. Για το σκοπό αυτό, παρήχθηκε το αρχείο *PCAfactors\_2\_factors.csv* το οποίο περιέχει τα μετασχηματισμένα δεδομένα βάσει των 2 παραγόντων PCA και τα κανονικοποιημένα δεδομένα απόδοσης του δείκτη.

## (γ)

Έχοντας εισάξει τα δεδομένα στο υπολογιστκό φύλλο, εκτελέσαμε Γραμμική Παλινδρόμηση με το εργαλείο Analysis ToolPak. Ύστερα υπολογίσθηκαν οι εκτιμήσεις, τα σφάλματα, τα τετραγωνικά σφάλματα και τα απόλυτα ποσοστιαία σφάλματα αφού θα μας βοηθήσουν να κατανοήσουμε τον βαθμό απόκλισης στα μετασχηματισμένα δεδομένα.

Τα αποτελέσματα δεν είναι ενθαρρυντικά με R Square = 0.005 και MAPE = 164% όπως φαίνεται στο φύλλο *LR 2 factors.* Μπορούμε να κάνουμε διάφορες υποθέσεις για το αποτέλεσμα αυτό. Θα βοηθούσε αν εξετάζαμε τα γραφήματα διασποράς της μετασχηματισμένης απόδοσης ως προς τους παράγοντες PCA.





Εικόνα 19 Πραγματικές και εκτιμημένες κανονικοποιημένες αποδόσεις ως προς του παράγοντες PCA.

Παρατηρούμε πως δεν υπάρχει ξεκάθαρη γραμμική συσχέτιση μεταξύ των PCA παραγόντων και της απόδοσης. Συγκεκριμένα λαμβάνουμε Pearson Correlation (*PCA 1, Target*) = -0.071 και (*PCA 2, Target*) = -0.009 με αποτέλεσμα να δυσκολεύει την πρόβλεη του γραμμικού εκτιμητή. Κρατώντας μόνο τον πρώτο παράγοντα *(n\_components=1)* πάλι θα είχαμε το ίδιο αποτέλεσμα αφού η γραμμική συσχέτιση παραμένει ίδια.

## (δ)

Η απάντηση στο ερώτημα αυτό έχει γίνει πλέον ξεκάθαρη. Αν κοιτάξουμε πάλι την μήτρα συσχέτισης (Εικ. 18) και επικεντρωθούμε στην τελευταία γραμμή θα παρατηρήσουμε ιδιαίτερα μικρές, κοντά στο μηδέν, τιμές συσχέυτησης μεταξύ της απόδοσης του δείκτη και των διαφορειτκών προϊόντων. Ήδη υποψιαζόμαστε ότι η γραμμική παλινδρόμηση δεν θα παράξει ικανοποιητικά αποτελέσματα. Αν λάβουμε υπόψη την ισχυρή πολυσυγγραμμικότητα των ανεξάρτητων μεταβλητών, οι ελπίδες μας για αποτέλεσμα καταστρέφονται.

Πράγματι, στο φύλλο *LR all vars* εκτελέσθηκε Γραμμική Παλινδρόμηση χρησιμοποιώντας όλα τα χαρακτηριστικά και οδηγηθήκαμε σε χείροτερα αποτελέσματα από ότι στην προηγούμενη περίπτωση.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **LR with** | **R Square** | **MSE** | **RMSE** | **MAPE** |
| **2 PCA factors** | 0.00514 | 0.994865 | 0.99743 | 164% |
| **All Variables** | 0.00643 | 0.99357 | 0.99678 | 170% |

Υπάρχουν εναλλακτικές μέθοδοι για να αντιμετοπιστεί το πρόβλημα της πολυγραμμικότητας αλλά δεν μπορούν να λύσουν το πρόβλημα της «φύσης» των δεδομένων. Τα επιτόκια ομολόγων διαφόρων λήξεων δεν αποτελούν επαρκείς παράγοντες ώστε να περιγράψουν-προβλέψουν έναν τόσο περίπλοκο δείκτη ο οποίος επηρεάζεται από ιδιαίτερα πολλούς και πολύπλοκους κοικωνικοπολιτικούς και οικονομικούς παράγοντες.

1. Προσεγγιστικά μπορούμε να θεωρήσουμε ότι με «μέσο σφάλμα» (RMSE) ίσο με 34,348$ σε δεδομένα με δειγματικό μέσο 178,000$, άρα μέση ποσοστιαία απόκλιση 19.2%, έχουμε μια «μέτρια» απόδοση. [↑](#footnote-ref-1)