

ΠΑΡΟΥΣΙΑΣΗ ΕΡΓΑΣΙΑΣ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΤΥΠΩΝ

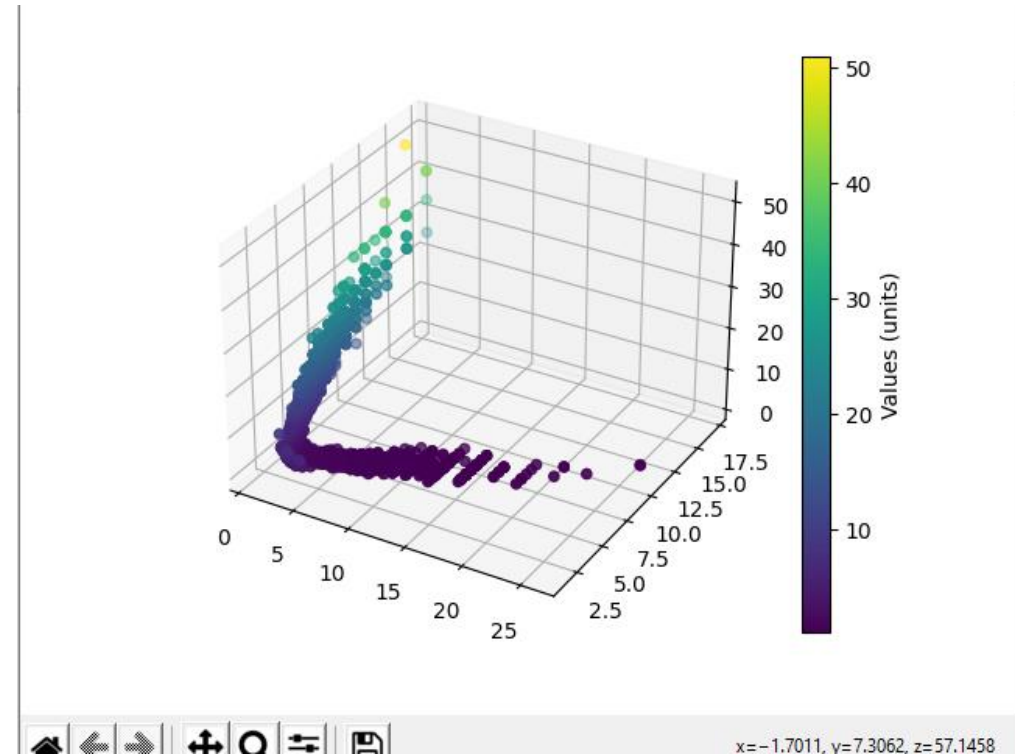
Κορώνης Ευάγγελος(Π17050)

Η παρούσα εργασία έχει υλοποιηθεί σε γλώσσα python. Απαραίτητες βιβλιοθήκες για την λειτουργία των προγραμμάτων είναι οι εξής:

- numpy
 - sqlite3
 - sklearn
 - pandas
 - matplotlib
-
- Επίσης μέσα στον ίδιο φάκελο με τα προγράμματα θα πρέπει να βρίσκεται και το αρχείο `database.sqlite` .

Ερώτημα 1

- Το πρόγραμμα που υλοποιεί το ερώτημα 1 είναι το **lms.py**
- Σε αυτό το ερώτημα χρησιμοποιούμε τον αλγόριθμο του Ελάχιστου Μέσου Τετραγωνικού Σφάλματος (**Least Mean Squares**), για να ταξινομήσουμε τα δεδομένα σε τρεις κλάσεις {H,D,A}.
- Το διάνυσμα χαρακτηριστικών που χρησιμοποιείται είναι οι τρεις αποδόσεις που δίνει η κάθε στοιχηματική εταιρεία για κάθε αγώνα. Παρακάτω φαίνεται η αναπαράσταση τους στον τρισδιάστατο χώρο.



Ερώτημα 1

- Δημιουργούμε τέσσερις ταξινομητές, έναν για κάθε εταιρεία. Επίσης χρησιμοποιούμε και την μέθοδο 10-fold cross validation. Άρα η διαδικασία του training και του testing επαναλαμβάνεται 10 φορές.
- Επιπλέον χρησιμοποιούμε την μέθοδο **one vs all**, δηλαδή υπολογίζουμε για κάθε κλάση την συνάρτηση διάκρισης μεταξύ του εαυτού της και των άλλων δύο κλάσεων.
- Χρησιμοποιούμε τον υλοποιημένο από την **padasip** βιβλιοθήκη αλγόριθμο **FilterNLMS**.
- Τα αποτελέσματα από την διαδικασία του training και του testing φαίνονται στην διπλανή εικόνα.
- Όπως βλέπουμε το ποσοστό των σωστών προβλέψεων είναι περίπου το ίδιο για όλες τις στοιχηματικές εταιρίες, με μεγαλύτερη ακρίβεια να έχει η B365.

```
===== RESTART: C:\Users\user\Desktop\pattern recognition\lms.py =====
```

```
1. Computing for B365 odds and results...
```

```
Mean training score: 52.08234931959502 %
```

```
Successfull predictions mean rate 52.12465183513732 %
```

```
2. Computing for BW odds and results...
```

```
Mean training score: 52.063494444824315 %
```

```
Successfull predictions mean rate 52.079992841874095 %
```

```
3. Computing for IW odds and results...
```

```
Mean training score: 51.713045194395114 %
```

```
Successfull predictions mean rate 51.69626998223801 %
```

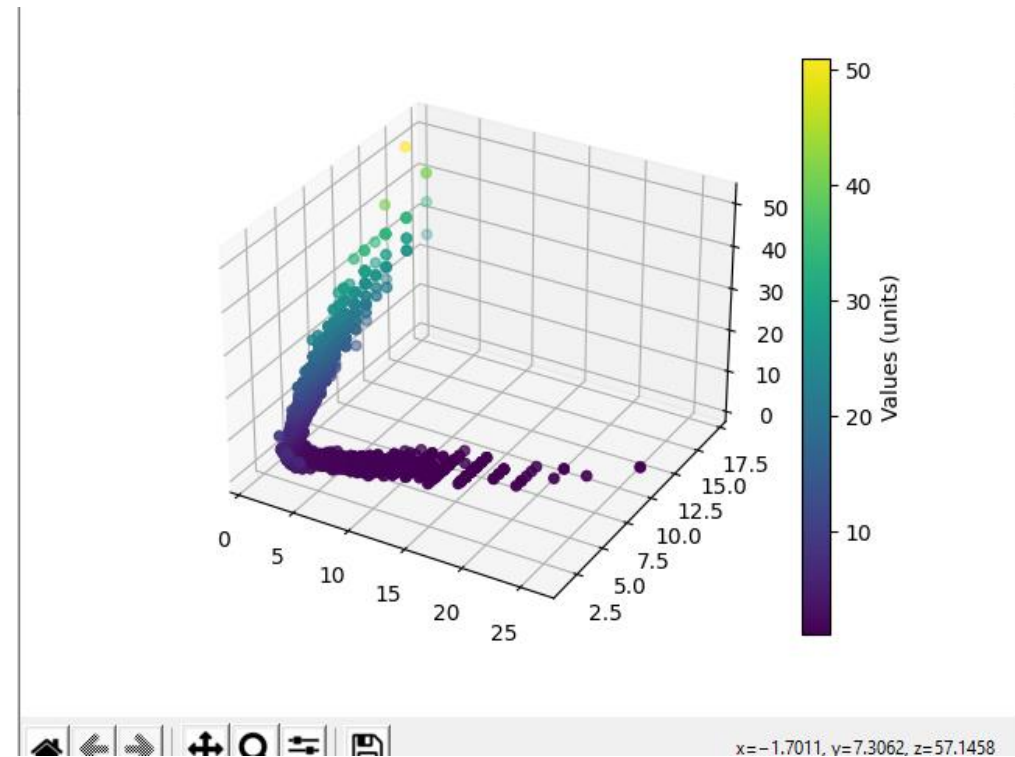
```
4. Computing for LB odds and results...
```

```
Mean training score: 51.96991460781252 %
```

```
Successfull predictions mean rate 51.99530004245884 %
```

Ερώτημα 2

- Το πρόγραμμα που υλοποιεί το ερώτημα 2 είναι το **least_squares.py**
- Σε αυτό το ερώτημα χρησιμοποιήσουμε τον αλγόριθμο του Ελάχιστου Τετραγωνικού Σφάλματος (**Least Squares**), για να ταξινομήσουμε τα δεδομένα σε τρεις κατηγορίες {H,D,A}.
- Το διάνυσμα χαρακτηριστικών που χρησιμοποιείται είναι οι τρεις αποδόσεις που δίνει η κάθε στοιχηματική εταιρεία για κάθε αγώνα. Παρακάτω φαίνεται η αναπαράσταση τους στον τρισδιάστατο χώρο.



Ερώτημα 2

- Δημιουργούμε τέσσερις ταξινομητές, έναν για κάθε εταιρεία. Επίσης χρησιμοποιούμε και την μέθοδο 10-fold cross validation. Άρα η διαδικασία του training και του testing επαναλαμβάνεται 10 φορές.
- Κάθε κλάση συμβολίζεται ως ένα διάνυσμα τριών στοιχείων: [1,0,0]= νίκη γηπεδούχου, [0,1,0] =ισοπαλία,[0,0,1]=νίκη φιλοξενούμενου.
- Χρησιμοποιούμε τον υλοποιημένο από την **sklearn** βιβλιοθήκη αλγόριθμο **LinearRegression**.
- Τα αποτελέσματα από την διαδικασία του training και του testing φαίνονται στην διπλανή εικόνα.
- Όπως βλέπουμε το ποσοστό των σωστών προβλέψεων είναι περίπου το ίδιο για όλες τις στοιχηματικές εταιρίες, με μεγαλύτερη ακρίβεια να έχει η B365.

```
==== RESTART: C:\Users\user\Desktop\pattern recognition\least_square
1. Computing for B365 odds and results...
Train score: 0.0859527176036176
Mean squared error: 0.20
Coefficient of determination: 0.08
Successfull predictions mean rate 52.73549655850542 %

2. Computing for BW odds and results...
Train score: 0.08583876111291774
Mean squared error: 0.20
Coefficient of determination: 0.08
Successfull predictions mean rate 52.68682061085029 %

3. Computing for IW odds and results...
Train score: 0.08625270573250253
Mean squared error: 0.19
Coefficient of determination: 0.08
Successfull predictions mean rate 52.45559502664298 %

4. Computing for LB odds and results...
Train score: 0.08342637456858894
Mean squared error: 0.20
Coefficient of determination: 0.08
Successfull predictions mean rate 52.4784894875061 %

|
```

Ερώτημα 3

- Το πρόγραμμα που υλοποιεί το ερώτημα 3 είναι το **mlp.py**
- Σε αυτό το ερώτημα χρησιμοποιήσουμε ένα πολυστρωματικό νευρωνικό δίκτυο (**Multilayer Perceptron**), για να ταξινομήσουμε τα δεδομένα σε τρεις κατηγορίες **{H,D,A}**.
- Το διάνυσμα χαρακτηριστικών που χρησιμοποιείται είναι οι αποδόσεις και των τεσσάρων στοιχηματικών εταιρειών για κάθε αγώνα καθώς και 8 χαρακτηριστικά για κάθε ομάδα.
- Επίσης χρησιμοποιούμε και την μέθοδο 10-fold cross validation. Άρα η διαδικασία του training και του testing επαναλαμβάνεται 10 φορές.

Κάθε κλάση συμβολίζεται ως ένα διάνυσμα τριών στοιχείων: $[1,0,0]$ = νίκη γηπεδούχου, $[0,1,0]$ =ισοπαλία, $[0,0,1]$ =νίκη φιλοξενούμενου.

Ερώτημα 3

- Χρησιμοποιούμε τον υλοποιημένο από την **sklearn** βιβλιοθήκη αλγόριθμο **MLPClassifier**.
- Τα αποτελέσματα από την διαδικασία του training και του testing φαίνονται στην διπλανή εικόνα.
- Στην εικόνα φαίνονται και τα αποτελέσματα του ταξινομητή για κάθε επανάληψη του 10-fold cross validation.

```
===== KESIAKI: C:\Users\user\Desktop\patu
aking Dataset...
training and testing multilayer perceptron...
)Successfull predictions rate 0.282559711581'
)Successfull predictions rate 0.301937809824:
)Successfull predictions rate 0.240648940964:
)Successfull predictions rate 0.279855790896:
)Successfull predictions rate 0.305092383956'
)Successfull predictions rate 0.296529968454:
)Successfull predictions rate 0.344749887336:
)Successfull predictions rate 0.370437133844:
)Successfull predictions rate 0.379900856241:
)Successfull predictions rate 0.31064021641:
an training score: 0.3099185666988791
ccessfull predictions mean rate 31.123526995:
```

> |