

Άσκηση #1

Αλγόριθμος γραμμικής παλινδρόμησης στο Weka

Weka Machine Learning

Η Weka διαθέτει έναν μεγάλο αριθμό αλγόριθμων παλινδρόμησης που είναι διαθέσιμα στην πλατφόρμα.

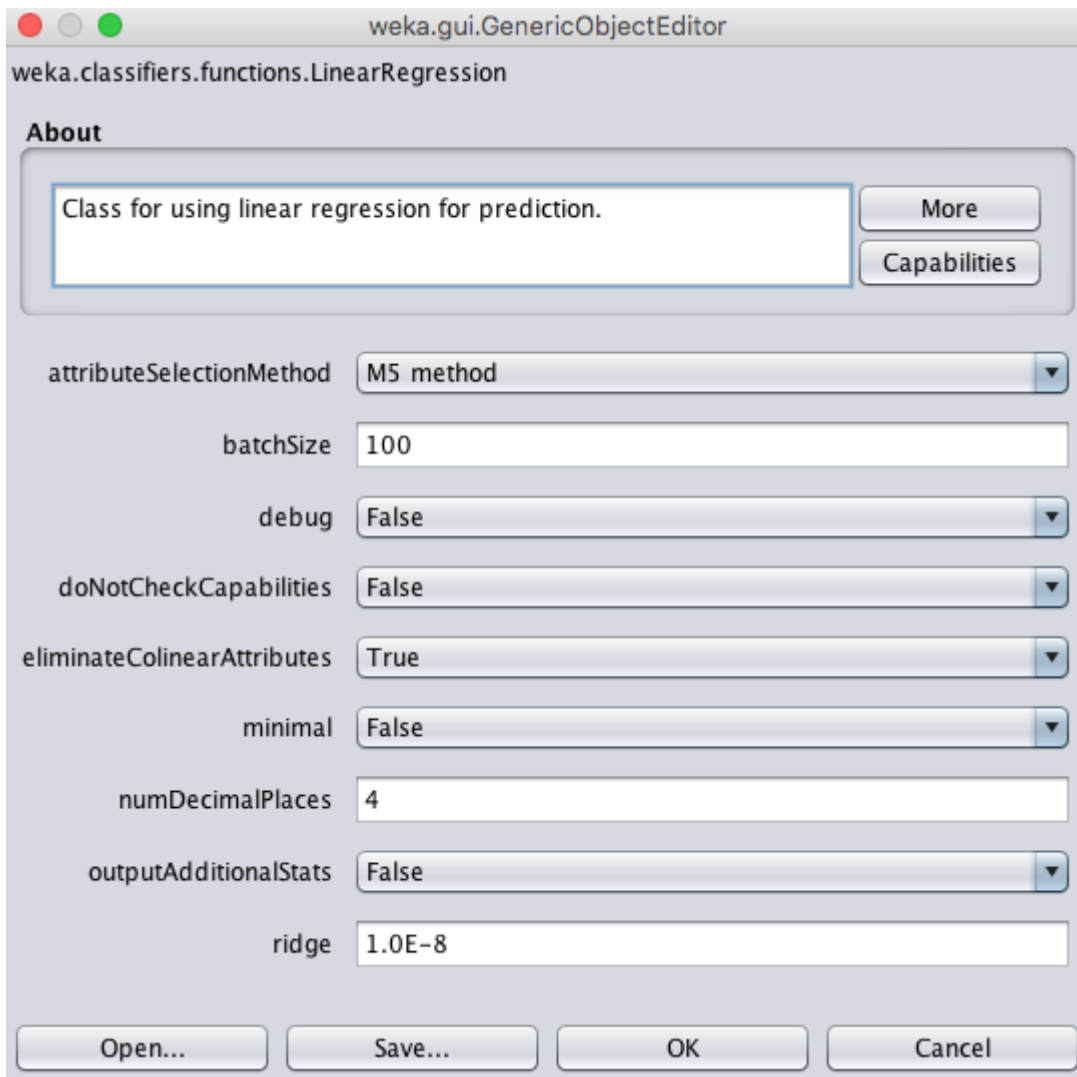
Ο μεγάλος αριθμός αλγορίθμων μηχανικής μάθησης που υποστηρίζονται από το Weka είναι ένα από τα μεγαλύτερα οφέλη από τη χρήση της πλατφόρμας.

Γραμμική παλινδρόμηση

Η γραμμική παλινδρόμηση λειτουργεί με την εκτίμηση των συντελεστών μιας ευθείας ή υπερπλησίας που ταιριάζει καλύτερα στα δεδομένα εκπαίδευσης. Είναι ένας πολύ απλός αλγόριθμος παλινδρόμησης, γρήγορος στην εκπαίδευση και μπορεί να έχει εξαιρετική απόδοση αν η μεταβλητή εξόδου για τα δεδομένα είναι ένας γραμμικός συνδυασμός των εισόδων.

Στο WEKA επιλέξτε τον αλγόριθμο γραμμικής παλινδρόμησης:

1. “Φορτώστε” το αρχείο housing.arff πατώντας στο Explorer και έπειτα το “open File” στο tab “preprocess”
2. Κάντε κλικ στο κουμπί “Classify” έπειτα στο "Choose" και επιλέξτε "**LinearRegression**" στην ομάδα "functions".
3. Κάντε κλικ στο όνομα του αλγορίθμου για να ελέγξετε τη διαμόρφωση του αλγορίθμου.



Weka Διαμόρφωση γραμμικής παλινδρόμησης

Η απόδοση της γραμμικής παλινδρόμησης μπορεί να μειωθεί εάν τα δεδομένα εκπαίδευσης έχουν χαρακτηριστικά εισόδου που είναι γραμμικά εξαρτημένα. Το Weka μπορεί να ανιχνεύσει και να αφαιρέσει αυτόματα γραμμικά εξαρτημένα στοιχεία της εισόδου αυτόματα, θέτοντας το `removeColinearAttributes` στην `True`, η οποία είναι και η προεπιλογή.

Επιπλέον, τα χαρακτηριστικά που δεν σχετίζονται με τη μεταβλητή εξόδου μπορούν επίσης να επηρεάσουν αρνητικά την απόδοση. Το Weka μπορεί να εκτελέσει αυτόματα την επιλογή χαρακτηριστικών για να επιλέξει μόνο εκείνα τα σχετικά χαρακτηριστικά και αυτό γίνεται ορίζοντας το `attributeSelectionMethod`. Αυτό είναι ενεργοποιημένο από προεπιλογή και μπορεί να απενεργοποιηθεί αν το θελήσουμε.

Τέλος, η εφαρμογή Weka χρησιμοποιεί μια τεχνική κανονικοποίησης των κορυφογραμμών για να μειώσει την πολυπλοκότητα του μοντέλου. Αυτό επιτυγχάνεται με την ελαχιστοποίηση του τετραγώνου των συντελεστών, που θα αποτρέψει να γίνει ένας συγκεκριμένος συντελεστής πολύ μεγάλος (ένα σημάδι πολυπλοκότητας στα μοντέλα παλινδρόμησης).

1. Κάντε κλικ στο "OK" για να κλείσετε τη διαμόρφωση του αλγορίθμου.
2. Κάντε κλικ στο κουμπί "Start" για να εκτελέσετε τον αλγόριθμο στο σύνολο δεδομένων για το σπίτι της Βοστώνης που σας έχει δοθεί.

Μπορείτε να δείτε ότι με την προεπιλεγμένη διαμόρφωση η γραμμική παλινδρόμηση επιτυγχάνει RMSE 4.9.

The screenshot shows the Weka Explorer application window. The 'Classify' tab is selected. The 'Classifier' section shows 'LinearRegression' with parameters '-S 0 -R 1.0E-8 -num-decimal-places 4'. Under 'Test options', 'Cross-validation' is selected with 'Folds' set to 10. The 'Classifier output' pane displays the regression equation and performance metrics.

Classifier output

```
-17.376 * NOX +  
3.8016 * RM +  
-1.4927 * DIS +  
0.2996 * RAD +  
-0.0118 * TAX +  
-0.9465 * PTRATIO +  
0.0093 * B +  
-0.5226 * LSTAT +  
36.3411
```

Time taken to build model: 0.17 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient	0.8451
Mean absolute error	3.3933
Root mean squared error	4.9145
Relative absolute error	50.8946 %
Root relative squared error	53.3085 %
Total Number of Instances	506

Status

OK Log x 0

Αποτελέσματα Weka για γραμμική παλινδρόμηση

Άσκηση #2 Clustering

Στο WEKA επιλέξτε τον αλγόριθμο SimpleKMeans:

1. “Φορτώστε” το αρχείο cadets.arff πατώντας στο Explorer και έπειτα το “open File” στο tab “preprocess”
2. Κάντε κλικ στο κουμπί “Cluster” έπειτα στο "Choose" και επιλέξτε "**SimpleKMeans**"

