

Automated Data Scientist

Νησιώτη Ελένη

AEM : 7737

Επιβλέπων: Χατζηδημητρίου Κυριάκος

Δεκέμβριος, 2016

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

HMMY

Σκοπός διπλωματικής εργασίας

Το 75% ενός πειράματος μηχανικής μάθησης αφιερώνεται στην προετοιμασία της εφαρμογής του αλγορίθμου και το 15% στα βήματα που την ακολουθούν. Το μεγαλύτερο μέρος της έρευνας επικεντρώνεται στο ενδιάμεσο 10% ...

— Rich Caruana
Automl 2016 @ ICML

Απαρχές	Unica, MarketSwitch, KXEN
Πεδία Εφαρμογής	Προεπεξεργασία, Ρύθμιση αλγορίθμου, Αξιολόγηση και κατανόηση μοντέλου
Σύγχρονα Εργαλεία	AutoWeka, Microsoft Azure, caret, HPOlib

Ένας αυτοματοποιημένος αναλυτής δεδομένων για προβλήματα δυαδικής ταξινόμησης με εμπειρία παλαιότερων πειραμάτων και κατανοητή έξοδο.

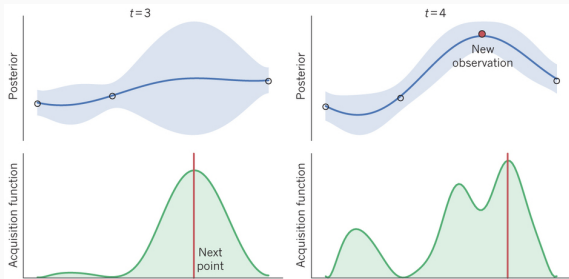
R

Automl

Τεχνολογία Λογισμικού

Μεθοδολογία

Σκοπός	Βελτιστοποίηση υπερπαραμέτρων αλγορίθμου
Τεχνική	Bayesian βελτιστοποίηση
Ιδέα	Αντικατάσταση της χρονοβόρας διαδικασίας βελτιστοποίησης με πρόβλεψη
Υλοποίηση	Εισαγωγή μετα-μάθησης για παραγωγή χαρακτηριστικών χρήσιμων σε ένα νευρωνικό πρόβλεψης των βέλτιστων υπερπαραμέτρων (HPPNN).



TPE

Speararmint

SMAC

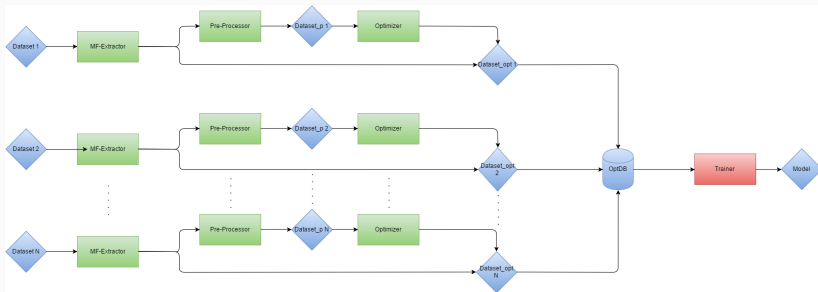
Σχήμα 1: Μία γενιά Bayesian Βελτιστοποίησης με Γκαουσιανές διαδικασίες

Σκοπός	Δημιουργία μετα-γνώσης από πειράματα μηχανικής μάθησης
Τρόπος	Εξαγωγή μετα-χαρακτηριστικών των σετ δεδομένων, τα οποία περιέχουν ουσιώδη πληροφορία
Εφαρμογές	Αρχικοποίηση αλγορίθμων βελτιστοποίησης

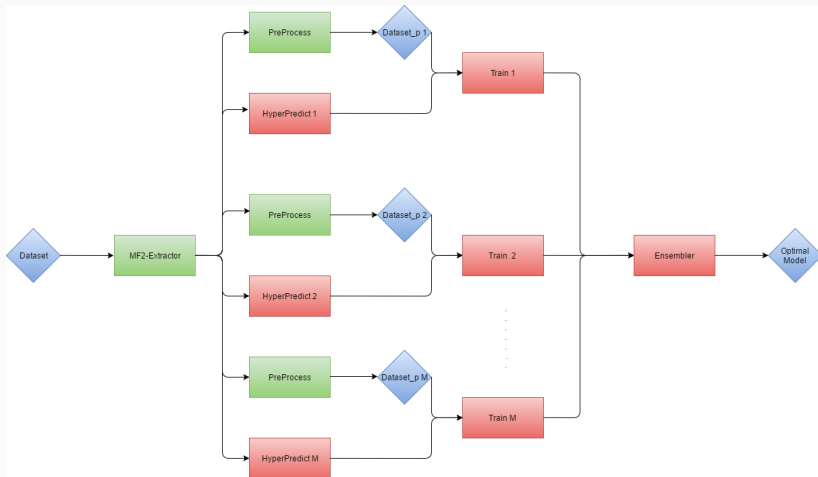
ENSEMBLE FORWARD SELECTION

[]

Αρχιτεκτονική Συστήματος



Σχήμα 2: Θα αλλαχθεί, το βάζω για αναφορά



Σχήμα 3: Θα αλλαχθεί, το βάζω για αναφορά

Πειραματικά Αποτελέσματα

Σετ δεδομένων 50 σετ δυαδικής ταξινόμησης από το UCI Repository.

Off-line Βελτιστοποίηση των αλγορίθμων SVM, NN, Tree, Bayes της caret με χρήση της βιβλιοθήκης HPOlib με hold-out validation, εξαγωγή μετα-χαρακτηριστικών και εκπαίδευση νευρωνικού με 10-fold CV.
[Χρόνος και hardware]

[διαγράμματα]

[διαγράμματα]

Η επιστήμη του automl επιδιώκει ένα συμβιβασμό μεταξύ αυτοματοποίησης και κατανοησιμότητας (intuition). Το λογισμικό που σχεδιάσαμε συνδυάζει τη λογική (reasoning) ενός αναλυτή δεδομένων με τεχνικές αυτοματοποίησης. Η αρχιτεκτονική εξασφαλίζει την εύκολη ενσωμάτωση state of the art τεχνικών και πρόσβαση (interface) στη χρήσιμη πληροφορία.

[??? Ανοιχτά θέματα ???]

[μακάρι]

ΕΡΩΤΗΣΕΙΣ

**[να επαναλάβω τα συμπεράσματα? ή να βάλω
κάτι που παραπέμπει στην ιδέα?]**

Βιβλιογραφία(να μπει??)



Rich Caruana. “Research Opportunities in Automl”. In: *Automl Workshop, ICML 2015* (2015).