

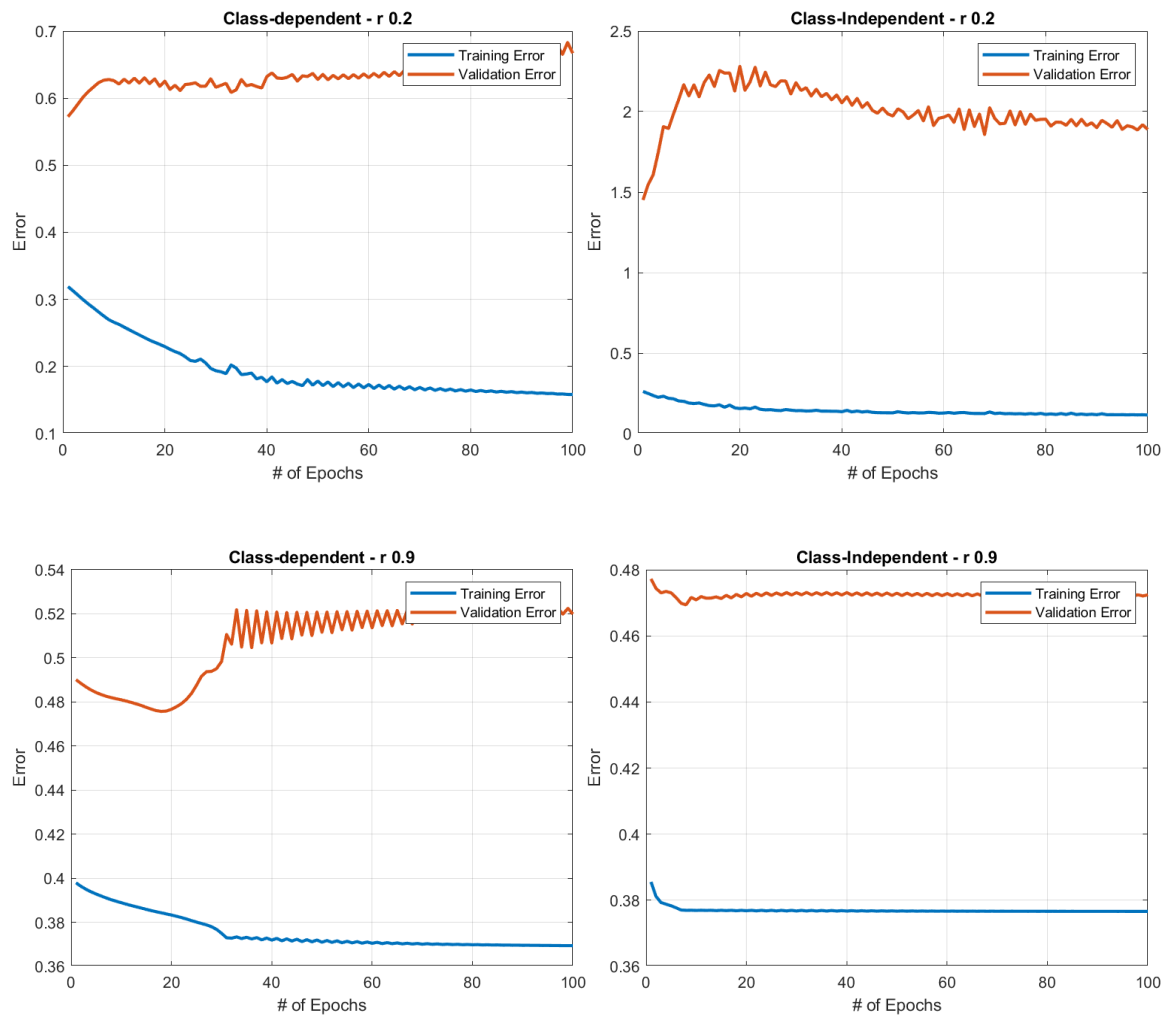
Υπολογιστική Νοημοσύνη
Εργασία 4
Classification

Ηλιάνα Κόγια
(AEM: 10090)
ilianakogia@ece.auth.gr

1 Εφαρμογή σε απλό dataset

Χωρίζουμε το dataset σε train - check - test με αντίστοιχα ποσοστά 60 - 20 -20.
Τα class-independent μοντέλα εκπαιδεύονται κατευθείαν μετά την αρχικοποίησή τους, ενώ για τα class-dependent εισάγουμε χειροκίνητα τις συναρτήσεις συμμετοχής, τους κανονές κτλ. με βάση τα αποτελέσματα της subclust.

Παρακάτω παρατίθενται τα αποτελέσματα για τα 4 μοντέλα: learning curves



Class-dependent - r: 0.9

OA	70.492
PA_1	100
PA_2	14.286
UA_1	68.966
UA_2	100
k	0.17937

Class-Independent - r: 0.9

OA	68.852
PA_1	97.5
PA_2	14.286
UA_1	68.421
UA_2	75
k	0.14591

Class-dependent - r: 0.2

OA	59.016
PA_1	80
PA_2	19.048
UA_1	65.306
UA_2	33.333
k	-0.010603

Class-Independent - r: 0.2

OA	62.295
PA_1	80
PA_2	28.571
UA_1	68.085
UA_2	42.857
k	0.093083

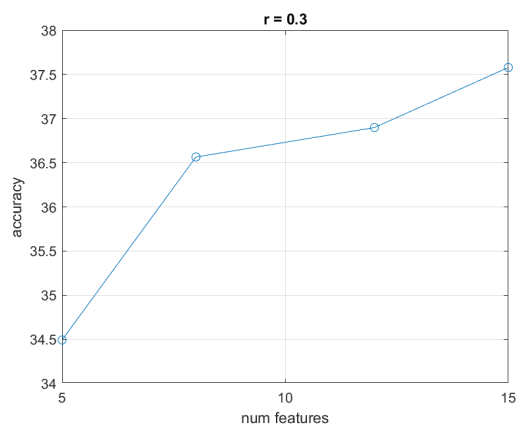
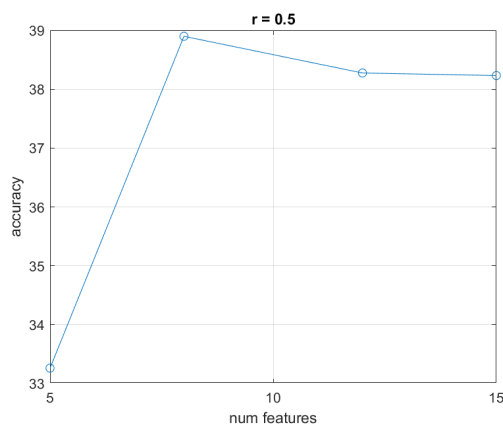
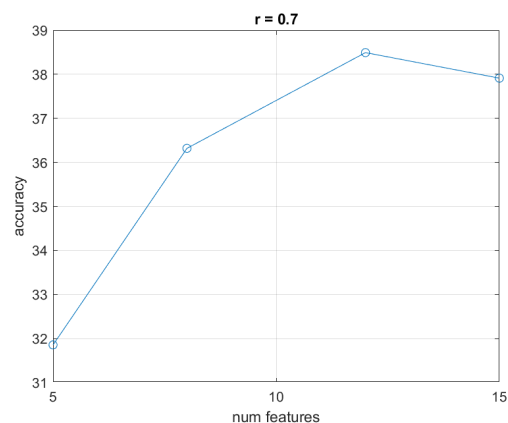
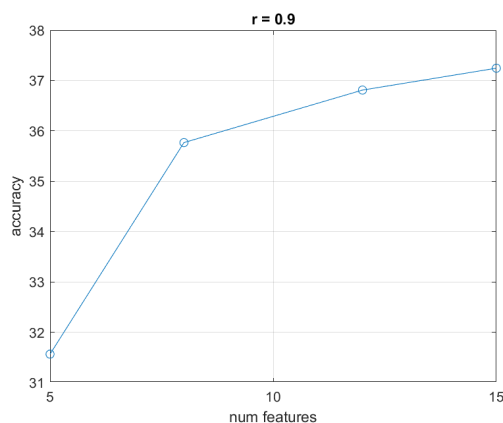
Τα καλύτερα μοντέλα είναι εκείνα για ακτίνα $r = 0.9$ και το βέλτιστο από αυτά το class dependent με $r = 0.9$.

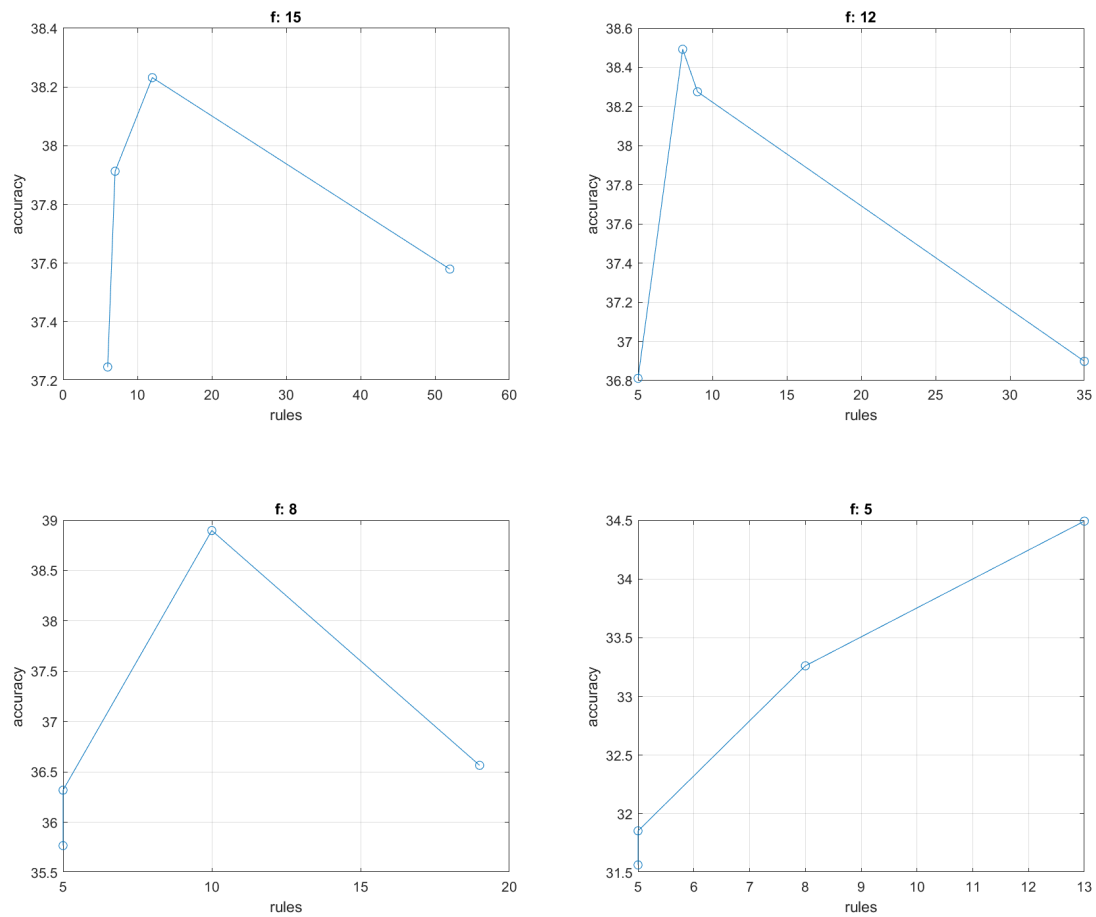
Για $r = 0.2$ δεν έχουν μεγάλη απόκλιση τα 2 μοντέλα.

2 Εφαρμογή σε **dataset** με υψηλή διαστασιμότητα

Χρησιμοποιούμε grid search για την επιλογή του βέλτιστου αριθμού features και την ακτίνα των clusters / αριθμό των κανόνων. Χωρίζουμε το αρχικό dataset σε 60-20-20 όπως πριν και χρησιμοποιώντας το training set υλοποιούμε 5-fold cross-validation. Ελέγχουμε τα μοντέλα για αριθμό features = 5, 8, 12, 15 και ακτίνα = 0.3, 0.5, 0.7, 0.9.

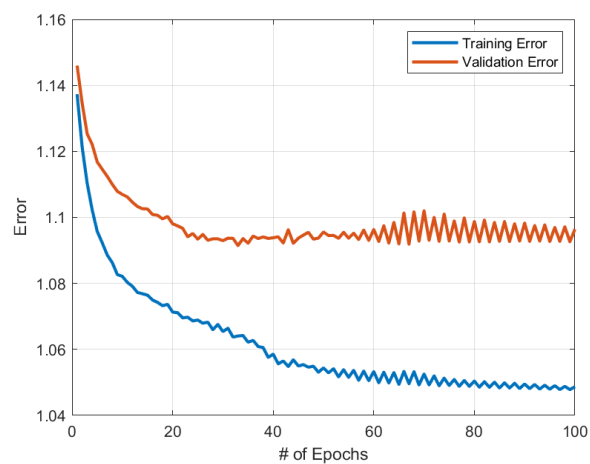
Κριτήριο επιλογής είναι ο μέσος όρος του OA κάθε μοντέλου μετά από τα 5 fold.





Από τα παραπάνω διαγράμματα βλέπουμε δεν υπάρχει κάποιος κανόνας όσον αφορά το accuracy και τον αριθμό των κανόνων ή των feature.

Το καλύτερο βρέθηκε ότι είναι για 8 feature και ακτίνα 0.5



Έχουμε παρόμοια στατιστικά για το OA και στο test set: 41.2 % γεγονός που δείχνει ότι κάνει καλό generalization