

Lecture 3: Decision Trees

Project 2

- Να χρησιμοποιήσετε το dataset [Breast cancer](#) και να εφαρμόσετε έναν Decision Tree αλγόριθμο δοκιμάζοντας διάφορες τιμές των παραμέτρων split function (criterion) και max depth. Ο κώδικας να δίνει σαν έξοδο την απόδοση του μοντέλου, όπως αυτή ορίζεται από τις μετρικές Accuracy, Precision, Recall, F1, και να δημιουργεί ένα plot του δέντρου.
- Να εφαρμόσετε τον αλγόριθμο Random Forest δοκιμάζοντας διάφορες τιμές παραμέτρων split function (criterion) και number of trees (n_estimators). Ο κώδικας να δίνει σαν έξοδο την απόδοση του μοντέλου, όπως αυτή ορίζεται από τις μετρικές Accuracy, Precision, Recall, F1. Επιπλέον, να δημιουργεί 4 γραφήματα (ένα για κάθε μετρική), όπου θα φαίνεται η απόδοση του μοντέλου (άξονας y) καθώς αλλάζει το πλήθος των δέντρων σε ένα μοντέλο Random Forest (άξονας x, από 1 έως 200 δέντρα).

Ενδεικτικά αποτελέσματα από τα πειράματα να καταγραφούν στον παρακάτω πίνακα.

a/a	Algorithm	Criterion	Max depth	Accuracy	Precision	Recall	F1
1	Decision Tree	gini	3	0.937063	0.928773	0.942156	0.934165
2	Decision Tree	gini	25	0.902098	0.893002	0.903142	0.897270
3	Decision Tree	entropy	3	0.909091	0.900198	0.919684	0.905983
4	Decision Tree	entropy	25	0.916084	0.910737	0.910737	0.910737
5	Decision Tree	log_loss	3	0.909091	0.900198	0.919684	0.905983
6	Decision Tree	log_loss	25	0.916084	0.910737	0.910737	0.910737

a/a	Algorithm	Number of Estimators	Criterion	Accuracy	Precision	Recall	F1
1	Random Forest	5	gini	0.951049	0.944512	0.953392	0.948468
2	Random Forest	25	gini	0.951049	0.943200	0.957033	0.948795
3	Random Forest	5	entropy	0.979021	0.973684	0.983146	0.977915
4	Random Forest	25	entropy	0.951049	0.944512	0.953392	0.948468
5	Random Forest	5	log_loss	0.979021	0.973684	0.983146	0.977915
6	Random Forest	25	log_loss	0.951049	0.944512	0.953392	0.948468