

Komitety klasyfikatorów (Ensemble methods)

Zadania:

1. Wykonaj implementację metody 'bagging' opartą na drzewach klasyfikacyjnych. Skorzystaj z gotowej implementacji drzew w R (biblioteka `rpart`). Porównaj działanie metody bagging i pojedynczego drzewa na wybranym zbiorze danych z repozytorium UCI, dokonując podziału danych na część treningową i testową.
2. Porównaj jakość predykcji (wyznaczając krzywe ROC i parametr AUC) dla różnych metod:

- pojedynczego drzewa,
- metody bagging,
- boosting,
- lasu losowego,
- gradient boosting

na wybranym zbiorze danych z repozytorium UCI, dokonując podziału danych na część treningową i testową.

3. Porównanie stabilności pojedynczego drzewa i komitetu klasyfikatorów.
 - Wybierz dowolny zbiór danych dotyczących klasyfikacji lub regresji.
 - Podziel dane na zbiór treningowy i testowy.
 - Porównujemy dwa modele: drzewo (funkcja `rpart`), las losowy (funkcja `randomForest`) oraz funkcyjny algorytm gradientowy (funkcja `glmboost` w pakiecie `mboost`).
 - Powtórz następujący krok $B = 100$ razy: dopasuj model klasyfikacji/regresji na podstawie zaburzonego zbioru treningowego, który powstaje poprzez usunięcie 10% obserwacji z oryginalnego zbioru treningowego. Dokonaj oceny modelu obliczając na zbiorze testowym błąd klasyfikacji lub błąd średnio-kwadratowy (dla regresji).
 - Na podstawie $B = 100$ symulacji oszacuj wariancje błędów dla drzewa oraz lasów losowych.