

Lab 2. Tehnici de învățare supervizată

Studiați utilizarea următoarelor 4 tehnici de clasificare:

- **bazate pe criteriul Bayes:** clasificator probabilistic, de regulă binar (două clase), ce se bazează pe o ipoteză de independență a atributelor de intrare; fiecărei clase i se asociază o probabilitate, ieșirea clasificatorului este clasa cea mai probabilă (optimizare în funcție de datele de antrenare).
- **bazate pe funcții:** problema clasificării este modelată cu ajutorul unei reprezentări funcționale a datelor de intrare, reprezentare ce este optimizată folosind datele de antrenare. (Una din metodele bazate pe funcții este *Support Vector Machines* – optimizează un hiperplan de separare a datelor din spațiul de caracteristici).
- **bazate pe arbori:** reprezintă informația sub formă arborescentă, unde clasele sunt reprezentate de frunzele arborelui, nodurile corespund atributelor datelor iar ramurile reprezintă relaționarea valorilor atributelor pentru a ajunge la un anumit nod.
- **meta-metode:** reunesc mai mulți clasificatori existenți; de regulă învață în mod iterativ un set de clasificatori “slabi” și îi adaugă progresiv la clasificatorul global; pe măsură ce sunt adăugați, datele sunt reponderate, datelor clasificate eronat li se crește ponderea în timp ce datelor clasificate corect le scade ponderea; astfel încât următorul clasificator slab se va focaliza pe datele clasificate greșit.

Sarcină : Alegeți spre implementare (în limbajul Python) câte o metodă din fiecare categorie de mai sus. Explicați conceptul și pașii de bază ai fiecărui algoritm, utilizând pe măsură posibilităților același set de date.

Notă ! Evaluați rezultatele obținute (*precision*, *recall*, *f-score*, *confusion_matrix*).

Comentați rezultatatele obținute în urma evaluării.

*from sklearn.metrics import classification_report,confusion_matrix, accuracy_score