

Laborator 2: Aplicații

1. Scrieți o funcție care calculează distanța euclidiană între două mulțimi de date reținute în matricele \mathbf{x} și \mathbf{y} (se presupune că matricea \mathbf{x} este $n \times m$, iar matricea \mathbf{y} este $p \times m$ dimensională).

Temă: Programați vectorial funcția de mai sus. (Indicație: Se poate folosi relația

$$\|a - b\|^2 = \|a\|^2 - 2a^T b + \|b\|^2, \text{ unde } a, b \text{ sunt vectori, adaptată pentru matrice }).$$

2. Matrice de confuzie

Matricele de confuzie reprezintă o modalitate utilă de prezentare a rezultatelor unui model de clasificare. Acestea furnizează informații detaliate legate de performanța modelului pentru fiecare clasă. Într-o matrice de confuzie C , liniile reprezintă clasele reale și coloanele reprezintă clasele prezise; elementul C_{ij} este numărul de exemple din clasa i care au fost clasificate în clasa j . Este evident faptul că matricea ideală este cea care conține elemente nenule numai pe diagonala principală.

Scrieți o funcție programată vectorial care returnează matricea de confuzie a mulțimii de date reținute prin perechea (\mathbf{X}, \mathbf{T}) , unde \mathbf{X} este o matrice care conține clasele prezise ale datelor, iar \mathbf{T} conține clasele reale (țintă). Se presupune că matricele \mathbf{X} și \mathbf{T} sunt introduse astfel:

- Dacă datele pot fi clasificate în c clase atunci se poate considera că fiecare linie a matricelor este un vector unitar, conținând 1 pe poziția corespunzătoare clasei din care face parte data de intrare și 0 în rest.
- Dacă datele pot fi clasificate în două clase atunci \mathbf{X} și \mathbf{T} pot fi considerați vectori coloană binari (deci valorile țintă sunt 0 și 1 corespunzătoare primei și respectiv celei de a doua clase).

De asemenea, returnați ca parametru de ieșire și rata de clasificare.

Bibliografie:

- I. T. Nabney, *Netlab. Algorithms for Pattern Recognition.*, Ed. Springer, 2002