

4. Regresia logistică

Reprezentarea funcției ipoteză
Marginea de decizie
Funcția logistică de cost
Algoritm de minimizare al gradientului

În această lucrare ne vom concentra asupra problemei de **clasificare binară**, în care y poate lua doar două valori, 0 sau 1. De exemplu, în cazul în care dorim să construim un clasificator ce ne poate spune dacă un email este spam sau nu, atunci $x^{(i)}$ ar fi un set de caracteristici ce descrie email-ul, iar y ar putea fi 1 dacă email-ul este spam și 0 în caz contrar. Altfel spus, $y \in \{0, 1\}$. De obicei clasa 0 se numește clasă negativă ("−"), iar 1 clasa pozitivă ("+"). Valoarea $y^{(i)}$ este denumită *eticheta* (eng. label) exemplului de antrenare $x^{(i)}$.

4.1 Cerințe

Cerința din această lucrare de laborator este de a construi un model de clasificare binară ce va estima, pe bazele rezultatelor a două teste, șansele unui elev de a fi admis la facultate.

4.1.1 Baza de date de antrenare

Pentru a implementa algoritmul, descărcați datele din arhiva "lab4Data.zip".

În acest exercițiu vom presupune că un liceu are o bază de date reprezentând 40 de elevi ce sunt admiși la facultate și 40 ce nu sunt admiși. Fiecare exemplu de antrenare $(x^{(i)}, y^{(i)})$ conține scorul unui elev la două teste standard și o etichetă ce indică dacă elevul a fost admis sau nu la facultate.

Structura bazei de date de antrenare are următoarea formă:

- (a) fișierul "lab4x.dat" conține caracteristicile x (rezultatele la cele două examene); prima coloană conține rezultatele pentru examenul 1, iar a doua coloană rezultatele pentru examenul 2;
- (b) fișierul "lab4y.dat" conține etichetele y , unde $y = 1$ reprezintă admiterea elevului la facultate; elevul nu a fost admis atunci când $y = 0$.