

Exercițiu laborator 3

Retele Bayesiene

Un sistem de clasificare a e-mail-urilor folosește o rețea bayesiană pentru a evalua probabilitatea ca un e-mail să fie spam (S) în funcție de anumite caracteristici observate. Rețeaua include următoarele variabile:

- S : E-mail-ul poate fi spam ($S = 1$) sau non-spam ($S = 0$).
- O : E-mail-ul poate conține cuvântul "ofertă" ($O = 1$) sau nu ($O = 0$).
- L : E-mail-ul poate conține link-uri ($L = 1$) sau nu ($L = 0$).
- M : E-mail-ul poate avea o lungime mare ($M = 1$) sau nu ($M = 0$).

STRUCTURA REȚELEI BAYESIENE

- Spam (S) influențează probabilitatea ca e-mail-ul să conțină cuvântul "ofertă" (O) și să conțină link-uri (L).
- Lungimea e-mail-ului (M) este influențată atât de faptul că este spam (S) cât și de prezența link-urilor (L).

TABELE DE PROBABILITĂȚI

- $P(S = 1) = 0.4$, $P(S = 0) = 0.6$
- $P(O = 1|S = 1) = 0.7$, $P(O = 1|S = 0) = 0.1$
- $P(L = 1|S = 1) = 0.8$, $P(L = 1|S = 0) = 0.3$
- $P(M = 1|S = 1, L = 1) = 0.9$, $P(M = 1|S = 1, L = 0) = 0.5$
- $P(M = 1|S = 0, L = 1) = 0.6$, $P(M = 1|S = 0, L = 0) = 0.2$

CERINȚE

- a) Identificați independentele din rețea.
- b) Determinați cum clasifică rețeaua bayesiană e-mail-urile în funcție de atributele O , L și M .