

2012ko maiatza. 1. deialdia

1. (Puntu 1) Futboleko munduko txapelketako finalaren atsedenaldira A eta B taldeen arteko partidari buruzko apustua egin nahi da. Horretarako, bi parametrotan oinarritzen den sailkapen paradigma bat eraiki nahi da. Parametroak honakoak dira: atsedenaldira partidaren emaitza zeinen aldekoa den eta txapelketa osoan zehar jasotako golak. A eta B futbol taldeek txapelketa irabazteko a priori probabilitate berbera dute.

Txapelketaren irabazlea A taldea bada, %80-ko probabilitatez B taldeak baino gol gutxiago jaso du txapelketan zehar. Irabazlea B taldea bada, %50-eko probabilitatez A taldeak baino gol gutxiago jaso du.

Atsedenaldi garaian A taldea B taldeari irabazten ari zaio, eta B taldeak A taldeak baino gol gutxiago jaso du txapelketan zehar. Futbol partidaren emaitzari buruzko iragarpena egin ezazu, zein apustu egitea komeni zaigun erabaki ahal izateko. Erabilitako paradigma justifika ezazu.

- | Kasua | X_1 | X_2 | C |
|----------------|-------|-------|---|
| $\mathbf{x}_1$ | 1 | 1 | 1 |
| $\mathbf{x}_2$ | 1 | 6 | 2 |
| $\mathbf{x}_3$ | 4 | 3 | 2 |
| $\mathbf{x}_4$ | 5 | 1 | 1 |
| $\mathbf{x}_5$ | 5 | 3 | 2 |
| $\mathbf{x}_6$ | 7 | 6 | 2 |
- Datubasea
- | | $\mathbf{x}_1$ | $\mathbf{x}_2$ | $\mathbf{x}_3$ | $\mathbf{x}_4$ | $\mathbf{x}_5$ | $\mathbf{x}_6$ |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| $\mathbf{x}_1$ | 0 | 5 | 3.6 | 4 | 4.47 | 7.81 |
| $\mathbf{x}_2$ | 5 | 0 | 4.24 | 6.4 | 5 | 6 |
| $\mathbf{x}_3$ | 3.6 | 4.24 | 0 | 2.24 | 1 | 4.24 |
| $\mathbf{x}_4$ | 4 | 6.4 | 2.24 | 0 | 2 | 5.38 |
| $\mathbf{x}_5$ | 4.47 | 5 | 1 | 2 | 0 | 3.6 |
| $\mathbf{x}_6$ | 7.81 | 6 | 4.24 | 5.38 | 3.6 | 0 |
- Distantzia Euklidestararak

- 1

- 2.3 (0.75 puntu) Bootstrap bidezko estimazio-metodoa. 3-NN sailkatzailearen ebaluazioa egin ezazu. Suposa ezazu datubasetik kasuak birjarritz entrenamendurako honako bi laginak sortu direla zoriz:

$$D_1 = \{\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_5\}$$

$$D_2 = \{\mathbf{x}_1, \mathbf{x}_1, \mathbf{x}_1, \mathbf{x}_4, \mathbf{x}_6, \mathbf{x}_6\}$$

Zein da sailkatzailearen asmatze-tasaren estimazioa?

- 2.4 (0.5 puntu) Sailkatzaileen konbinaketa: Bagging. Suposa ezazu 3-NN sailkatzailea bagging metodoaren bidez konbinatuz multisailkatzaile bat osatu dugula. Datubasetik kasuak birjarritz honako bi laginak sortu dira zoriz:

$$D_1 = \{\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_5\}$$

$$D_2 = \{\mathbf{x}_1, \mathbf{x}_1, \mathbf{x}_1, \mathbf{x}_4, \mathbf{x}_6, \mathbf{x}_6\}$$

Eman ezazu $\mathbf{x}_7 = (2, 3)$ kasu berriaren C klase-aldagaiaren iragarpena.

3. (0.75 puntu) Beheko taulan hiru indibiduen hamar aldagai bitarrak neurtu dira.

3.1 Azaldu nola neurtzen diren bi indibiduen arteko antzekotasun koefizienteak.

3.2 Kalkula ezazu hiru indibiduen arteko *Sokal-Michener*-en antzekotasuna.

3.3 Kalkula ezazu hiru indibiduen arteko *Jaccard*-en antzekotasuna.

Indibiduoak	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$	$\mathbf{x}_7$	$\mathbf{x}_8$	$\mathbf{x}_9$	$\mathbf{x}_{10}$
1	1	1	0	1	1	0	1	0	1	0
2	1	1	1	0	0	0	1	1	1	1
3	0	0	1	1	1	1	1	1	1	0

4. (0.75 puntu) Bigarren ariketako taulako kasuak erabiliz, kalkula ezazu Forgy-ren algoritmoaren sailkapena hasierako partiketa ondokoa izanik:

$$K_1 = \{1, 4, 6\} \quad K_2 = \{2, 3, 5\}$$