

Feladat: becslés egy egyszerű hierarchikus modellben

Elek Péter

1. feladat. Tegyük fel, hogy m darab egységből (pl. iskolából, kórházból) van egyenként k számú megfigyelésünk (pl. az egységek teljesítményére vonatkozóan), és célunk a várható értékek becslése minden egység szintjén.

Formálisan, legyen y_{ij} az i . egység j . megfigyelése ($i = 1, \dots, m; j = 1, \dots, k$) és tegyük fel, hogy $y_{ij} \sim N(\mu_i, \sigma^2)$, és becsljük a μ_i paramétereket. Célunk az összesített átlagos négyzetes eltérés minimalizálása, azaz olyan $\hat{\mu}_i$ becslések keresése, amelyekre a következő kifejezés minimális:

$$MSE = \frac{\sum_{i=1}^m E[(\hat{\mu}_i - \mu_i)^2]}{m}.$$

- Az egységek szintjén vett részátlagok (azaz $\bar{y}_i = \frac{\sum_{j=1}^k y_{ij}}{k}$) torzítatlanul becsljük az egységszintű várható értékeket (μ_i -ket). Határozzuk meg ezen becslések összesített átlagos négyzetes eltérését!
- Definiáljunk egy másik becslőfüggvényt úgy, hogy az a részátlagoknak a főátlag (azaz $\bar{y} = \frac{\sum_{i=1}^m \sum_{j=1}^k y_{ij}}{k \cdot m}$) felé való behúzásával áll elő:

$$\hat{\mu}_i = \lambda * \bar{y} + (1 - \lambda) * \bar{y}_i,$$

ahol $0 < \lambda < 1$ a korrekciós tényező. Határozzuk meg ezen becslés torzítását, varianciáját és összesített átlagos négyzetes eltérését!

- Tegyük fel, hogy m "nagy" (formálisan $m \rightarrow \infty$) és

$$\eta^2 = \frac{\sum_{i=1}^m (\mu_i - \mu)^2}{m}$$

a rész várható értékek szórásnégyzete, ahol μ a teljes várható érték:

$$\mu = \frac{\sum_{i=1}^m \mu_i}{m}.$$

Mely λ esetén lesz MSE legkisebb ebben az esetben? Értelmezzük az eredményt!