

Harjoitus 3

1. Lataa tiedosto managers_all_details.csv ja muodosta viestinvälitystä koskeville muuttujille 11-15 korrelaatiomatriisi. Etsi jokaiselle muuttujalle itseisarvoltaan suurin korrelaatio ja anna sille tulkinta.
2. Muodosta regressiomalli, jossa ennustat muiden antamien kommenttien lukumäärää yhden käyttäjän lähettämien viestien lukumäärän avulla. Tutki saamasi mallin parametrien tilastollista merkitsevyyttä.
3. Muodosta viestinvälitysmuuttujille keskiarvot miesten ja naisten joukossa. Tarkastele edelleen ko. keskiarvovektorien erotuksena voisiko jokin muuttuja erottaa miehet ja naiset toisistaan.
4. Valitse kaksi sellaista viestinvälitysmuuttujaa, joissa on ”eroa” miesten ja naisten välillä. Lisää viestinvälitysmuuttujien perään luokitus C siten, että miehillä on arvo 0 ja naisilla 1. Muodosta logistinen regressiomalli ja ennusta sillä sukupuolta, kun viestimuuttujat tunnetaan.
5. Muodosta managers_all_details.csv tiedostosta keskiarvovektori, jossa on mukana pelkästään Thomas analyysin muuttujat. Oletetaan, että tämä määrää tulevan johtajan ominaisuusprofiilin. Lataa no_mgrs_all_details.csv ja ota siitäkin mukaan vain Thomas muuttujat. Laske johtajien joukon keskiarvovektorin ja kaikkien työntekijöiden Euklidiset etäisyydet ja valitse johtajaksi se, jossa on pienin etäisyys.