

1. Asenna päätöspuuohjelmisto C4.5, joka löytyy [täältä](#). Asennuksen pitäisi onnistua suoraviivaisesti ainakin laitoksen Linux-ympäristössä (lue ohjeet paketissa olevasta ReadMe-tiedostosta) – älä välitä virheilmoituksista...
Tutustu C4.5:n käyttöön (esim. komento `man Doc/c4.5.1` saattaa auttaa).
2. Kokeile päätöspuun oppimista hakemistosta `Data` löytyvällä `vote`-datasetillä. Laske opitun puun testivirhe (vivulla `-u`). Tutkaile C4.5:n tulostamaa karsimatonta ja karsittua puuta ja niihin liittyviä statistiikkoja. Halutessasi voit toistaa samat kokeet myös muilla hakemiston `Data` dataseiteillä.
3. Palataan viime laskuharjoitusten tehtävän 1 [dataan](#), ja sovelletaan siihen harjanneregressiota. Käytä ristiinvalidointiin perustuvaa mallinvalintaa hyvän regularisointiparametrin δ arvon löytämiseksi. Opi ennustaja, ja esitä arvio sen tarkkuudesta.
4. Palataan [viiksi-dataan](#). Etsi (luennoilla esiteltyä koneoppimis-, mallinvalinta- ja testauskalustoa käyttäen) ongelmaan neliövirheen suhteen hyvä ennustaja, ja esitä arvio sen tarkkuudesta.
Huom: `Data` ei ole satunnaisessa järjestyksessä, joten se kannattaa ennen opetus- ja testiosaan jakamista satunnaispermutoida (kts. funktio [randperm](#)).
5. (Johdatusta ohjaamattoman oppimisen osioon.) Etäisyysfunktiot ovat keskeisessä asemassa monissa ohjaamattoman oppimisen menetelmissä. Etäisyysfunktion määrittelyä vaikeuttaa se, että käsiteltävien piirrevektorien komponentit ovat usein eri tyyppisiä. Tällaisissa tilanteissa etäisyysfunktio piirrevektoreille määritellään usein summana piirrekohtaisista etäisyysfunktioista.

Miten määrittelisit piirrekohtaisen etäisyysfunktion seuraavissa tapauksissa? Etäisyysfunktion arvon tulisi siis mitata sitä, kuinka kaukana kaksi kyseisen piirteen arvoa ovat toisistaan.

- (a) Piirre kuvaa k :n vaihtoehdon osajoukkoa, ts. piirteen arvona on jokin 2^k erilaisesta k :n vaihtoehdon osajoukosta.

Tällainen piirre voisi kuvata esimerkiksi vastausta kysymykseen: “Mikä seuraavista asioista ovat sinulle tärkeitä: X Y Z”. Tässä siis $k = 3$, piirteen mahdollisia arvoja ovat tyhjä joukko, X, Y, Z, XY, XZ, YZ, ja XYZ, ja etäisyysfunktiolla pitäisi voida mitata kuinka lähellä esim. osajoukko X on osajoukkoa XY.

- (b) Piirre kuvaa k :n vaihtoehdon järjestystä, ts. piirteen arvona on järjestetty k -alkioinen lista.

Tällainen piirre voisi kuvata esimerkiksi vastausta kysymykseen: "Aseta tärkeysjärjestykseen seuraavat asiat: X Y Z". Tässä siis $k = 3$, piirteen mahdollisia arvoja ovat XYZ, XZY, YXZ, YZX, ZXY, ja ZYX, ja etäisyysfunktioilla pitäisi voida mitata, kuinka lähellä esim. järjestys XYZ on järjestystä YZX.