

1. *Nopanheittoa.* Tarkastellaan satunnaiskoetta, jossa heitetään vihreää ja punaista 6-sivuista arpakuutiota eli noppaa. Olkoon satunnaismuuttuja X vihreän nopan antama silmäluku, ja satunnaismuuttuja Y punaisen nopan antama silmäluku.
 - (a) Laske todennäköisyys, että $X = 3$ ja $Y = 4$.
 - (b) Laske todennäköisyys, että $X + Y \geq 8$.
 - (c) Laske todennäköisyys, että $X + Y \geq 8$ ehdolla $Y \geq 4$.
 - (d) Laske X :n ehdollinen jakauma ehdolla $X + Y = 5$. Piirrä jakaumasta kuva.
2. *Nopanheitto jatkuu.* Tarkastellaan ennustusongelmaa, jossa vihreän ja punaisen nopan antamien silmälukujen tulon perusteella ennustetaan vihreän nopan silmälukua, ja tavoitteena on osua täsmälleen oikeaan silmälukuun mahdollisimman usein.
 - (a) Muotoile ongelma formaalisti ennustusongelmaksi – siis identifioi sopivat \mathcal{X} , \mathcal{Y} , \mathcal{Y}' , ja tappiofunktio.
 - (b) Mikä on jakauma D (tässä tapauksessa täysin tunnettu, koska ennustusongelman taustalla oleva datan generoiva ilmiö tunnetaan)? Esitä D esim. taulukkona. Tutkaile taulukosta myös jakaumia D_X , $D_{Y|X}$, ja $D_{X|Y}$.
 - (c) Laske paras mahdollinen ennustaja f^* .
 - (d) Mikä on f^* :n odotusarvoinen tappio?
3. *Nopanheitto jatkuu edelleen.* Muunnetaan edellisen tehtävän ennustusongelmaa siten, että tavoitteena on osua vihreän nopan oikean silmäluvun sijasta mahdollisimman lähelle oikeaa silmälukua (neliövirheen mielessä). Miten vastaukset edellisen tehtävän kohtiin (a)-(d) muuttuvat?
4.
 - (a) Lataa tiedoston [lh2-data.txt](#) sisältämä datamatriisi. Datamatriisissa on otos eräästä joukon $\mathbb{R} \times \mathbb{R}$ jakaumasta. Tulkitaan otos regressio-ongelman opetusdataksi: ensimmäinen \mathbb{R} tulkitaan syötejoukoksi \mathcal{X} , ja toinen \mathbb{R} "oikeiden vastausten" joukoksi \mathcal{Y} .
 - (b) Piirrä koko datasta kuva, joka havainnollistaa datan x - ja y -komponenttien välistä riippuvuutta.
 - (c) Piirrä kuva, joka havainnollistaa otoksen x -komponentin jakaumaa.
 - (d) Piirrä kuva, joka havainnollistaa otoksen y -komponentin jakaumaa ehdolla $x = -1$. (Vihje: Koska ehto $x = -1$ tuskin toteutuu usein täsmällisesti, ota huomioon otoksesta myös sellaiset (x, y) -parit, joiden x -komponentti on "lähellä" arvoa -1 .)

- (e) Jos olisit koneoppimismenetelmä ja saisit tämän opetusdatan, mitä tuottamasi ennustaja ennustaisi syötteellä -1 ? Kuten regressiossa yleensä, valitaan $\mathcal{Y}' = \mathcal{Y}$ ja tappiofunktioiksi L_2 -tappio.
 - (f) Piirrä kuva, joka havainnollistaa otoksen y -komponentin jakaumaa ehdolla $x = 0$.
 - (g) Jos olisit koneoppimismenetelmä ja saisit tämän opetusdatan, mitä tuottamasi ennustaja ennustaisi syötteellä 0 ?
5. *Datan keruusta.* Jotta ohjattua oppimista voitaisiin soveltaa, tarvitaan (paljon) opetusdataa. Opetusdatan hankkiminen (pienin kustannuksin) ei kuitenkaan aina ole suoraviivaista, vaan saattaa edellyttää nokkeluutta.
- (a) Tutustu Luis von Ahnin [ESP-peliin](#) (katso myös [wikipedia-sivu](#)), jossa ihmiset houkutellessaan kertomaan ilmaiseksi mitä heille näytetyissä kuvissa näkyy.
 - (b) ESP-peliä pelaavien ihmisten voidaan tulkita olevan ennustajia, jotka ennustavat, mitä toinen ihminen kuvan nähdessään ajattelee. Filosofoi, olisiko pelin avulla kerätystä datasta koneopitulla ennustajalla mahdollisuuksia pärjätä pelissä.
 - (c) Pohdi ESP-pelin kaltaisten strategioiden rajoituksia – kuinka luotettavaa kerätty data on, minkä tyyppisiin tehtäviin menetelmä saattaisi soveltua, mihin taas ei?