

Huom: Voit saada tästä harjoituskerrasta max. 6 pistettä.

Tehtävä 1. Naivi Bayes ja SPAM (3 pistettä)

- a) (1 piste). Laske kynän ja paperin (ja laskimen) avulla todennäköisyys

$$P(\text{SPAM} \mid \text{SANA}_1 = \text{Viagra}),$$

kun $P(\text{SPAM}) = 0.5$, $P(\text{SANA}_1 = \text{Viagra} \mid \text{SPAM}) = 0.0021$, ja $P(\text{SANA}_1 = \text{Viagra} \mid \neg\text{SPAM}) = 0.0001$. (*Vihje*: Bayesin kaava).

- b) (2 pistettä). Toteuta luennolla esitetyn pseudokoodin mukaisesti roskapostisuodatin, joka lukee tiedostosta jonon sanoja ja palauttaa *Odds*-arvon. Käytä luokkaehdollisina todennäköisyyksinä itse omasta datastasi laskemiasi arvioita. Korvaa nollatodennäköisyydet jollakin pienellä positiivisella luvulla.

Muunna *Odds*-arvo roskapostin todennäköisyydeksi kaavalla

$$P(\text{SPAM} \mid \text{EVIDENSSI}) = \frac{\text{Odds}}{1 + \text{Odds}}.$$

Testaa suodatintasi viestillä

SPECIAL OFFER : VIAGRA on SALE at \$1.38 !!!

Compare the best online pharmacies to buy Viagra. Order Viagra online with huge discount. Multiple benefits include FREE shipping, Reorder discounts, Bonus pills

ja jollakin asiallisella viestillä (ham).

Ohjeita: Kun arvioit luokkaehdollisia todennäköisyyksiä, poista viesteistä tunnisteet (headerit) ja mahdolliset liitetiedostot. Saat listattua eri sanojen esiintymismäärät seuraavalla komennolla:

```
$ cat *.txt | tr '[:upper:]' '\n[:lower:]' | tr -d '<>=(){}&,;:~!/?#_+*' |
sort | grep -v '^$' | uniq -c | sort -n
```

Komento poistaa aluksi viesteistä välimerkit ja muuntaa isot kirjaimet pieniksi. Korvaamalla jälkimmäisellä rivillä olevan osan komennolla `wc -w` saat selville sanojen kokonaismäärän.

Tehtävä 2. Hahmontunnistus (3 pistettä)

Lataa kurssin kotisivulta paketti `surf-package.zip`, pura se ja suorita komento `ant example`. (Toimii laitoksen Linux-koneissa.) Tuloksena pitäisi olla luennolla esitetty kasvontunnistusesimerkki.

- a) (1 piste). Kokeile vaihtaa `example`-hakemistossa olevat tiedostot `img1.jpg` ja `img2.jpg` toisiin kuviin ja aja ohjelma uudelleen.

Kokeile vähintään viittä muuta kuvaparia, joista osassa näkyy sama rakennus, sama henkilö ja/tai sama esine. Kokeile myös kuvapareja, jotka eivät esitä samaa kohdetta. Yritä löytää esimerkkejä, joissa SURF-piirteisiin perustuva tunnistus toimii (eli samaa esittävistä kuvista löytyy useita yhteensopivia piirteitä) ja esimerkkejä, joissa se ei toimi.

- b) (1 piste). Valitse yksi (tai useampi) kuva ja muokkaa sitä

- säätämällä valoisuutta ja kontrastia,
- vaihtamalla värisävyjä,
- kallistamalla kuvaa (rotaatio),
- rajaamalla ja skaalaamalla uudelleen,
- lisäämällä kohinaa (noise) tai sumentamalla (blur).

Voit käyttää haluamaasi kuvankäsittelyohjelmaa (Picasa, iPhoto, GIMP, Photoshop, jne).

Toimiiko SURF, kun sovellat sitä alkuperäiseen ja muokattuun kuvaan?

Vaihtoehtoisesti: Ota itse kameralla kuvapareja, jotka esittävät samaa esinettä, rakennusta tai henkilöä, ja kokeile toimiiko SURF.

Pohdi millaiset muutokset kuvissa ja olosuhteissa aiheuttavat ongelmia SURF:iin perustuvalla kuvantunnistukselle ja millaiset taas eivät.

- c) (1 piste). Pohdi mitkä ovat tärkeimmät erot digitaalisten signaalien ja symbolisen datan (kuten tekstin, monivalintavastausten tai logiikkaan perustuvan tiedon) käsittelyn välillä?

Mainitse esimerkki menetelmästä, joka soveltuu hyvin vain jälkimmäisen tyyppisen aineiston käsittelyyn. Miksi se ei sovellu digitaalisten signaalien käsittelyyn?

Vastaavasti, mainitse esimerkki sovelluksesta/metodista, jossa digitaalisten signaalien käsittely onnistuu hyvin. Mihin onnistuminen perustuu?

Tehtävä 3. Yleiskuva (1 piste)

Pohdi mitkä ovat tekoälyn pääteemat (esim. pelit, koneoppiminen, ...). Miten kiinnostus niihin on vaihdellut eri vuosikymmeninä? Voit piirtää esim. kaavion, jossa esität pääteemojen välisiä suhteita (A sisältyy B:hen, A on lähellä B:tä, jne.).

Apuna tehtävässä voit käyttää esim. kurssin kotisivulla mainitun Ertelin kirjan lukua 1 (“What is Artificial Intelligence?”), Russellin ja Norvigin teosta “Artificial Intelligence: A Modern Approach”, Wikipediaa ja muita lähteitä.

Luettele myös kurssilla toistaiseksi käsiteltyjä asioita, mukaan lukien ensi viikolla käsiteltävät koneoppiminen ja robotiikka, ja yhdistä ne edellä mainittuihin pääteemoihin (elleivät ne jo ole niitä). Kommentoi kurssin sisältöä ja painotuksia suhteessa pääteemoihin.

Tehtävä 3 palautetaan erillisellä paperilla tai sähköpostitse laskuharjoitusten ohjaajalle.