

**Tehtävä 1. Naivi-Bayes ja SPAM (2 pistettä).**

Naivi-Bayes-mallia käytetään roskapostin suodattamiseen. Mallin todennäköisyysparametrien oppimista varten on kerätty valmiiksi luokiteltuja sähköpostiviestejä. Seuraavassa on joidenkin sanojen esiintymismäärät kummassakin luokassa (spam ja ham):

sana	spam	ham
algoritmi	0	12
sikahalpa	8	0
tarjous	21	1
päivän	9	8
<i>yht.</i>	5 000	10 000

Arvioi ehdolliset todennäköisyydet  $P(\text{SANA}_i = s \mid \text{spam})$  ja  $P(\text{SANA}_i = s \mid \text{ham})$ , missä  $s \in \{\text{algoritmi, sikahalpa, tarjous, päivän}\}$ .

Laske näitä hyödyntäen seuraavat todennäköisyydet:

- i. (1 piste)  $P(\text{spam} \mid \text{päivän})$
- ii. (1 piste)  $P(\text{spam} \mid \text{päivän, tarjous, sikahalpa, algoritmi})$

Käytä prioritodennäköisyytenä arvoa  $P(\text{spam}) = 0.5$ . (*Vihje:* Pienten todennäköisyyksien kohdalla on syytä käyttää jotakin alarajaa, esim. 0.0001.)

**Tehtävä 2. Numeroiden luokittelu neuroverkolla (2 pistettä).**

Toteuta valmiiseen (Java-)ohjelmarunkoon, tai kokonaan alusta, perseptronialgoritmi numeroiden tunnistamiseen. Tiedosto `'mnist-x.data'` sisältää yhteensä 6000 kuvaa, yksi kuva jokaisella tiedoston rivillä, ja jokainen kuva on  $28 \times 28$  pikseliä (eli jokaisella rivillä siis  $28 \times 28 = 784$  numeroa). Jokainen pikseli on joko musta (-1) tai valkoinen (1). Tiedosto `'mnist-y.data'` sisältää näitä kuvia vastaavat luokat (0-9). Tiedostossa `'Perceptron.java'` löytyy datan lukemiseen ja kuvien esittämiseen tarkoitetut apufunktiot. Voit ajaa esimerkkikoodin komennolla

```
java Perceptron 10
```

missä 10 on opetusaskelten määrä.

Netbeans-ohje:

Talleta tiedostot `mnist-x.data` ja `mnist-y.data` projektin päähakemistoon, kuten vastaavat datatiedostot aiemmissa tehtävissä. Anna `main`-metodissa muuttujalle `steps` arvoksi haluamasi opetusaskelten määrä, oletusarvona on 100.

- a) Näytä ruudulla ensimmäiset 100 merkkiä ja printtaa vastaavat 100 luokkaa. Nämä pitäisi vastata toisiaan. Verifioi tällä tavoin että datan lukeminen on onnistunut.
- b) (1 piste). Muokkaa perseptronialgoritmia (`train()`-metodi) siten, että se oppii erottamaan numeroa '3' esittävät kuvat numeroa '5' esittävistä kuvista. (Aseta luokkamuuttujan arvoksi 1, kun kuva esittää kolmosta, -1, jos esittää vitosta, ja ohita kaikki muut esimerkit.) Minkä luokitteluvirheen saat aikaiseksi (metodi `test()`)?
- c) (1 piste). Kokeile eri numeroparien erottelua (muitakin kuin 3 vs 5). Mitkä numerot on helpoin erottaa toisistaan, mitkä vaikein?

### Tehtävä 3. Palaute (1 piste).

Tämä tehtävä palautetaan joko laskuharjoitusten ohjaajalle tulostettuna (tai käsin kirjoitettuna) tai sähköpostitse luennoijalle (`teemu.roos at cs.helsinki.fi`).

- a) Aiheet: Mitkä tekoälyn osa-alueet ovat jääneet kurssilta parhaiten mieleen? Mikä aihe on ollut kiinnostavin? Mikä vähiten kiinnostava? Mitä kaipaisit kurssille lisää ja mitä vähemmän? (Viimeisillä viikoilla on tulossa robotiikkaa ja hahmontunnistusta.) Asteikolla 1–5, miten kiinnostava kurssin aihe sinusta on?
- b) Laskuharjoitukset: Mitä olet pitänyt laskuharjoituksista? Asteikolla 1–5, miten tärkeinä olet pitänyt laskuharjoituksia oppimisen kannalta?
- c) Materiaali: Ovatko luennot, luentokalvot, kurssimoniste (ks. kurssin verkkosivu) ja muu lisämateriaali sisältäneet sopivan määrän tietoa? Onko joku kohta jäänyt liian vähälle käsittelylle tai onko jostain aiheesta liikaakin materiaalia? Asteikolla 1–5, miten hyvin kurssin oppimateriaalit ovat tukeneet opiskelua?
- d) Lupaatko täyttää kurssipalautelomakkeen kurssin päätteeksi? Vihje: Lupaat.
- e) Vapaa sana. Saa olla useampi.