

Johdatus tekoälyyn, harjoitus 4

- $$P(\text{"algoritmi"}|\text{spam}) = P(\text{"algoritmi"})P(\text{spam}|\text{"algoritmi"})/P(\text{spam}) \sim (12/15000)*0/(5000/10000) = 0/0.5 = 0$$

$$P(\text{"sika halpa"}|\text{spam}) = P(\text{"sika halpa"})P(\text{spam}|\text{"sika halpa"})/P(\text{spam}) \sim (8/15000)*1/(1/2) \sim 0.0010666$$

$$P(\text{"tarjous"}|\text{spam}) \sim ((22/15000)*(21/22))/(1/2) \sim 0.0028$$

$$P(\text{"päivän"}|\text{spam}) \sim ((17/15000)*(9/17))/(1/2) \sim 0.0012$$

$$P(\text{"algoritmi"}|\text{ham}) \sim ((12/15000)*(1))/(1/2) \sim 0.0016$$

$$P(\text{"sika halpa"}|\text{ham}) \sim ((8/15000)*(0))/(1/2) = 0$$

$$P(\text{"tarjous"}|\text{ham}) \sim ((22/15000)*(1/22))/(1/2) \sim 0.000133$$

$$P(\text{"päivän"}|\text{ham}) \sim ((17/15000)*(8/17))/(1/2) \sim 0.001066$$

$$P(\text{spam}|\text{"päivän"}) = P(\text{spam})P(\text{"päivän"}|\text{spam})/P(\text{"päivän"}) \sim ((1/2)*(0.0012))/(1/2) \sim 0.0012$$

$$P(\text{spam}|\text{"päivän"}, \text{"tarjous"}, \text{"sika halpa"}, \text{"algoritmi"}) = P(\text{spam})P(\text{"päivän"}, \text{"tarjous"}, \text{"sika halpa"}, \text{"algoritmi"}|\text{spam})/P(\text{"päivän"}, \text{"tarjous"}, \text{"sika halpa"}, \text{"algoritmi"}) \sim ((1/2)*(0.0010666 + 0.0028 + 0.0012))/(17/15000 + 22/15000 + 8/15000 + 12/15000) \sim 0.644059$$

- Kun opetusaskeleiden määrä nousee niin jatkuvasti hidastuen luokittelijan tarkkuus nousee. 10 iteraatiolla vaikeimmat erottaa ovat 1 ja 2. Niiden muoto onkin hyvin samanlainen, mutta kun iteraatiot kymmenkertaistetaan, niin päästään jo alle 5% virhemarginaaliin. Tästä seuraava aleneminen tapahtuukin vasta 100000 iteraation paikkeilla. Vaikeat erottaa toisistaan havaitsin olevan 2 ja 3. Niiden erottaminen onkin kiinni vain tietystä bittien osaluueesta, joten tämä hankaluus on järkeenkäypä. Kun askeleiden määrää kymmenenkertaistaa niin virhemarginaali jää hyvin pieneksi tässäkin tapauksessa, esim 10000 iteraatiolla alle prosentin tuntumalle. Noin sadalla askeleella myös 9 ja 8 ovat vaikeita erottaa toisistaan.

Hyvin erimuotoiset numerot, esim 1 ja 9 ovat jo 10 iteraatiollakin hyvin erotettavissa, noin 5% virhemarginaalilla.

- Parhaiten mieleen ovat jääneet viime aikoina käsitellyt luokittimet ja koneoppiminen yleensä. Näistä päästäänkin minua eniten kiinnostavaan aiheeseen: koneoppimiseen. Vähemmän kiinnostavaa taas on ollut GOFAI, sillä se ei jotenkin konseptina jaksa innostaa minua. Sitä ei kuitenkaan juurikaan käsiteltykään kurssilla. Kaipaisin kurssille hieman laajempaa johdatusta koneoppimiseen muunmuassa hyvin matemaattista perustaa lineaarisille luokittimille, sekä lyhyitä sanasia konvergenssista: esim voitaisiin esitellä lineaarinen regressio ja miten sen parametrit voidaan säätää vaikka gradient descent algoritmilla. Nämä tosin ehkä esitellään machine learning kurssilla. En oikeastaan kaipaa kurssille yhtään mitään vähemmän. Kurssin jokaisesta aihepiiristä on hyvä tietää jotain, ja siksi ei sen sisältöä kannata juuri karsia. Arvosana: 5
 - En ole ikävä kyllä päässyt itse laskuharjoitustilanteisiin, sillä olen ollut koko kurssin ajan Saksassa. Olen kuitenkin pitänyt laskuharjoitustehtävistä, mutta jotkut simulointitehtävät ovat olleet turhan pitkiä kokonaan auki täytettäväksi ja niiden tekemiseen on vaikeampi löytää motivaatiota. Tärkeys arvosana: 5
 - Luentokalvot ja yhteenveto täydentävät toisiaan. Kuitenkin aina silloin tällöin olen turvautunut ylimääräiseen materiaaliin kun haluan syvällisemmin ymmärtää joitain asioita, esim siitä miten Perceptronia voidaan soveltaa multiclass luokitteluun. Myös naivi bayes

luokittimen käsittelyssä turvauduin ulkopuoliseen materiaaliin, sillä pidin joitain yksityiskohtia hieman vaikeatajuisina yhteenvedossa. Arvosana: 4

d) Toki

e) Kurssi on erittäin mielenkiintoinen, oikeastaan tähän mennessä ottamistani kursseista ylivoimaisesti mielenkiintoisin. Ikävä kyllä yhden viikon laskuharjoitukset jäivät minulta välistä työkiireiden takia, ja osasta en ehtinyt tehdä aivan kaikkea. Olisin toivonut pystyvän satsaamaan kurssiin vieläkin enemmän, mutta koen jo nyt saaneeni kurssista paljon irti, vaikka useat aiheet olivatkin entuudestaan tuttuja, kuten reitinhaku ja luokittimien perusidea. Kiitos paljon kurssista sekä ohjaajille että luennoijalle!