

Sisältö

- Yleistä
 - Miten tehostaa tenttien arviointia?
 - Mitä on koneoppinen?
- Alustavat tulokset
 - Mitä kysyttiin ja miten data valmistellaan?
 - Mitä tarkoittaa, että vektorit ovat lineaarisesti riippuvia?
 - Miten voidaan tutkia, onko matriisi kääntyvä?
 - Todista luonnollista kieltä käyttäen, että nolla-avaruus on aliavaruus.
 - Miten voidaan tutkia, muodostaako vektorijoukko kannan jonkin toisen vektorijoukon virittämälle avaruudelle?
 - Mitkä olivat yleisimmät vastauksissa esiintyneet sanat?
 - Miten voidaan saada parempia tuloksia?
- Lähteet

Miten tehostaa tenttien arviointia?

- Sähköisten tenttien yleistyessä arviointiprosessi edelleen suurimmaksi osaksi manuaalinen
- STACK-järjestelmä mahdollistanut numeeristen tehtävien automatisoimisen
- Käsitteenmuodostus- ja todistustehtävät edelleen käsin arvioitavia

Laske $2 + 2$.

Vastaus:

Vastauksesi tulkittiin muodossa:

4

✓ Vastaus on oikein!

Oikea vastaus on 4, joka voidaan syöttää muodossa 4.

Mitä koneoppiminen on?

- Tarkoitus saada tietokone suorittamaan toimintoja itsenäisesti
- Yleisin sovellus on datan luokittelu esim. hakukoneet ja roskapostisuodattimet [2]
- Koneeseen syötetään opetusdata, josta luodaan luokittelijalle malli
- Luokittelija testataan erillisellä testidatalla mallin laadun tarkistamiseksi
- Käytettyjen luokkien ollessa jatkuvia arvoja puhutaan regressiomallista [3]

Mitä algoritmeja työssä käytetään?

- k-Nearest Neighbor (k-NN) = datalle määritetään koordinaatti moniulotteisessa koordinaatistossa ja se luokitellaan luokkaan, joka esiintyy k:n lähimmän naapurin keskuudessa [4]
- Deep Learning = ihmisen aivojen toimintaa jäljittelevä monikerroksinen neuroverkko [5]
- Decision tree = jakaa dataa osiin puudiagrammien mukaisesti kyllä/ei kysymysten perusteella [6]
- Random Forest = kokoaa suuren määrän satunnaisesti luotuja Decision tree –luokittelijoita määrää datalle näiden antamista luokista yleisimmän [6]

Mitä kysyttiin ja miten data valmistellaan?

- ”Mitä tarkoittaa, että vektorit ovat lineaarisesti riippuvia?”
- ”Miten voidaan tutkia, onko matriisi kääntyvä?”
- ”Todista luonnollisella kielellä nolla-avaruus aliavaruudeksi.”
- ”Miten voidaan tutkia, muodostaako vektorijoukko kannan jonkin toisen vektorijoukon virittämälle avaruudelle?”
- Vastaukset kootaan ja muutetaan numeeriseksi matriisiksi, jonka alkiot kertovat, kuinka monta kertaa jokin sana esiintyy yhdessä vastauksessa

Mitä tarkoittaa, että vektorit ovat lineaarisesti riippuvia?

- Luokittelutarkkuuden $ka = 53,7\%$
- Paras tarkkuus = $63,0\%$
- ”Jos jokin joukon vektori voidaan ilmoittaa joukon toisten vektorien lineaarikombinaationa, ne ovat lineaarisesti riippuvia.”
- Monta eri tapaa perustella oikea vastaus esim. rref-muoto

k-lähin naapuri, tehtävä 1

0	3	2		
1	2	5		1
2	1		5	
3	2	1	2	3
	0	1	2	3

True Class

Predicted Class

Miten voidaan tutkia, onko matriisi kääntyvä?

- Luokittelutarkkuuden $ka = 69,2\%$
- Paras tarkkuus = $70,4\%$
- ”Matriisi on kääntyvä, jos sen determinantti on erisuuri kuin nolla.”
- Pisteet painottuneet yläpäähän, josta luokittelun binäärisyys voi johtua

Satunnaismetsä, tehtävä 2

0	2			1
1				1
2	1			4
3	1			17
	0	1	2	3

True Class

Predicted Class

Todista luonnollisella kielellä, että nolla-avaruus on aliavaruus.

- Luokittelutarkkuuden $ka = 67,4\%$
- Paras tarkkuus $=77,8\%$
- ”Nollavektori toteuttaa nolla-avaruuden ehdon, joten nolla-avaruus on epätyhjä. Lisäksi nolla-avaruus on suljettu yhteenlaskun ja skalaarilla kertomisen suhteen.”

Deep Learn, tehtävä 3

0	10			1
1	4	3		
2		1		2
3		1		5
	0	1	2	3

True Class

Predicted Class

Miten voidaan tutkia, muodostaako vektorijoukko kannan jonkin toisen vektorijoukon virittämälle avaruudelle?

- Luokittelutarkkuuden $ka = 55,6\%$
- Paras tarkkuus = $70,4\%$
- Selkeästi vaikein tehtävä
- ”Jos vektorijoukon vektorien määrä on sama kuin vektoriavaruuden dimensio, ja vektorit ovat lineaarisesti riippumattomia, joukko muodostaa kannan.”

Satunnaismetsä, tehtävä 4

0	14	2		
1	3	1	1	
2	2	3		
3			1	
	0	1	2	3

True Class

Predicted Class

Mitkä sanat olivat yleisimpiä vastauksissa?

- Sanapilvet kertovat jokaisen tehtävän vastauksien yleisimmät sanat
- Suuret sanat ovat yleisempiä kuin pienet sanat
- Tehtävät ovat numerojärjestyksessä vasemmalta oikealle ja ylhäältä alas



Miten voidaan saada parempia tuloksia?

- Tulokset vaihtelevat pienen otoksen takia
- Täytesanojen poistaminen ja sanojen lemmatisoiminen (muuttaminen perusmuotoon) tehostaisi oppimista ja tekisi vastauksista samankaltaisempia
- Oliko pisteytysasteikko liian laaja/kapea?

Lähteet

[1] STACK

Dokumentaatio, <https://stack2.maths.ed.ac.uk/demo2018/question/type/stack/doc/doc.php/CAS/>

[2] Machine Learning Classifiers, Towards Data Science, Sidath

A., <https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623>

[3] Regression in Machine Learning, Data Driven Investor, Apoorva

D., <https://medium.com/datadriveninvestor/regression-in-machine-learning-296caae933ec>

[4] Introduction to k-Nearest Neighbours, Analytics Vidhya, Tavish

S., <https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/>

[5] What is Deep Learning?, Machine Learning Mastery, Brownlee

J., <https://machinelearningmastery.com/what-is-deep-learning/>

[6] Understanding Random Forest, Towards Data Science, Yiu

T., <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>