

# Tekoäly ja koneoppiminen metsävaratiedon apuna

Tuomo Kauranne, Virpi Junntila, Katri Tegel ja Jarkko Suuronen

Arbonaut Oy ja LUT University

26. marraskuuta 2018

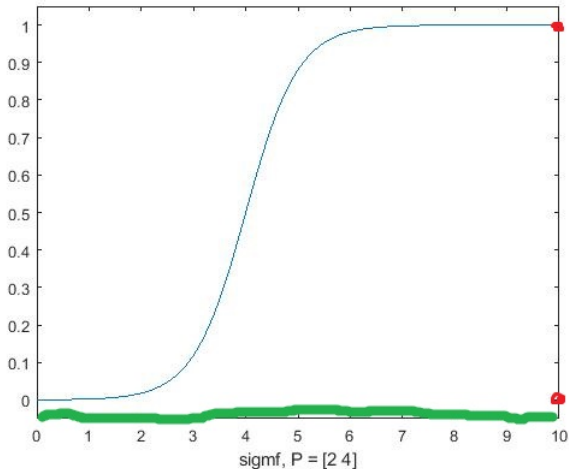
Metsätieteen päivä 2018

# Koneoppimisen kohteena ovat lukujen sijasta jakaumat

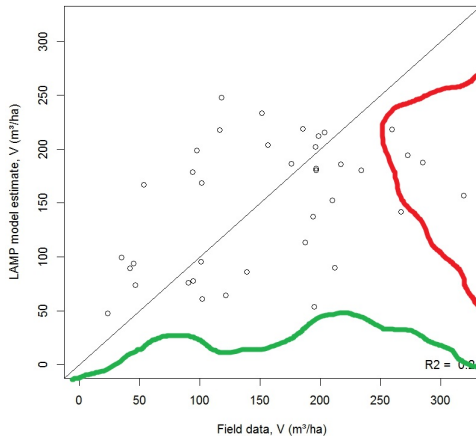
## Esimerkki 1

- *Koneoppimisessa ei yleensä käytetä populaation käsitettä eikä siksi otantateoriaa*
- *Jakaumia käsitellään aina jatkuvina ja äärettöminä*
- *Jakaumat ovat useimmiten moniulotteisia, jopa ääretönulotteisia - poikkeuksena luokitteluongelmat*
- *Lisäksi jakaumat on määritelty kaarevilla monistoilla, ei lineaarisella avaruudella*
- *Bayesilainen tilastotiede voidaan tulkita paikalliseksi lineaarisen avaruuden sovittamiseksi tähän monistoon*
- *Se on siis erityistapaus yleisestä koneoppimisesta - samoin SUR*

# Luokittelevan neuroverkon jakaumat



# Regressiomallin jakaumat



# Koneoppiminen on...

## Määritelmä 1

- *Kuvaus jakaumilta jakaumille, jossa*
- *Lähtöjoukko on jakaumat esimerkiksi kaukokartoitusdatan piirteiden avaruudella*
- *Kuvajoukko on jakaumat esimerkiksi mahdollisten spatiaalisten puustotunnusjoukkojen avaruudella*
- *Molemmat jakaumat voivat olla ääretönulotteisia*

# Koneoppiminen on...

## Määritelmä 2

- *Etsittyä kuvausta edustaa äärellisulotteinen **monisto (manifold)** ääretönulotteisessa tulo- ja lähtöjakaumien karteesisen tulon avaruudessa*
- *Tämä monisto määritellään kuvaajamoniston kiinnittävien **hyperparametrien** arvot kiinnittämällä*
- *Siten koneoppimisen kohteena on äärellisulotteinen avaruus jossa akseleina ovat eri hyperparametrit*
- *Ja opittu malli on yksi piste tai joskus pistejoukko tässä avaruudessa*
- *Silti tämä yksi pistekin vastaa kokonaista monistoa*

# Esimerkkejä koneoppimisen algoritmien hyperparametreista

## Esimerkki 2

- *Regressiomallin kertoimet*
- *Neuroverkon painot*
- *Radiaalisten Gaussin kantafunktioiden sijainnit, hajonnat ja painot*
- *Random Forest-algoritmin päätöspuiden joukko*
- *Kalman-suotimen kovarianssimatriisi*
- *GAN-verkon "Maansiirtofunktion" määrittävät todennäköisyysmassan siirrot*

# Informaatiogeometria

## Esimerkki 3

- *Informaatiogeometria esittää koneoppimisen differentiaaligeometrian kielellä*
- *Se mahdollistaa koneoppimisen algoritmien tulkinnan esimerkiksi optimointialgoritmien avulla*
- *Informaatiogeometrian perusteiden keskeinen kehitys tapahtui Japanissa 1980-luvulla, johtavana tutkijana Shun-ichi Amari*



# Informaatiogeometria

## Esimerkki 4

- *Informaatiogeometria rakentaa etäisyysmitan (metriikan) jakaumien välille*
- *Siinä jakaumien erilaisuuden neliölle annetaan määritelmä jakaumien välisen **divergenssin** kautta*
- *Tälle etäisyysmitalle saadaan euklidista vastaava käyräviivainen geometria jossa esimerkiksi Pythagoraan lause on voimassa*
- *Samoin suorakulmainen projektio*
- *Näiden avulla voidaan määritellä kriteeri optimaaliselle mallinsovitukselle*
- *Moniulotteisen jakauman keskinäiset korrelaatiot tulkitaan etsityn moniston kaarevuuksina*

# k-NN menetelmä koneoppimisen algoritmina

## Esimerkki 6

- *k-NN -algoritmissa tehdään interpolaatio piirteiden avaruudessa*
- *Hyperparametreina toimivat koalapankin koalojen puustotunnukset*
- *Etäisyysmittana toimii neliöityjen piirrekohtaisten poikkeamien summa*
- *Monisto on lokaalisti k-ulotteinen, mutta globaalisti n-ulotteinen missä n on koalojen määrä*

# k-NN menetelmä koneoppimisen algoritmina

## Esimerkki 8

- *k-MSN -algoritmissa tehdään myös interpolaatio piirteiden avaruudessa*
- *Hyperparametreina toimivat koeal pankin koealojen puustotunnukset*
- *Etäisyysmittana toimii neliöityjen piirrekohtaisten poikkeamien painotettu summa*
- *Painot määrätään koeal pankin piirteiden pääkomponenttianalyysillä*
- *Monisto on lokaalisti k-ulotteinen, mutta globaalisti n-ulotteinen missä n on koealojen määrä*

# SBSUR-menetelmä koneoppimisen algoritmina

## Esimerkki 10

- *SBSUR-algoritmissa lasketaan painotettu lineaarinen regressio piirteiden avaruudessa*
- *Hyperparametreina toimivat koelapankin koealojen puustotunnusten pääkomponentit*
- *Etäisyysmittana toimii neliöityjen piirrekohtaisten poikkeamien painotettu summa lisättynä sakkofunktiolla mallin dimensiolle*
- *Painot määrätään koelapankin piirteiden ja kaukokartoituspiirteiden parittaisella pääkomponenttianalyysillä yhteisjakaumassa*
- *Monisto on lokaalisti matalaulotteinen, mutta dimensio voi vaihdella kohdemetsikön ja koelapankin mukana*

# Random-Forest -luokittelu algoritmi

## Esimerkki 11

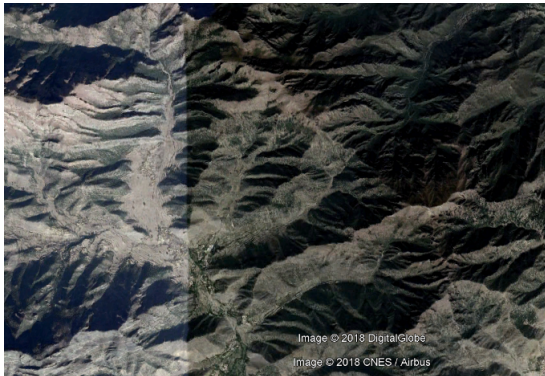
- *Random Forest -algoritmi kokeilee lukuisia erilaisia päätöspuita opetusaineistoon kunnes mielestään löytää kullekin luokalle parhaan päätöspuun.*
- *Sitä käytetään esimerkiksi maankäytön luokittelun algoritmina satelliittikuvilta.*
- *Ongelmana on usein satelliittikuvien optiset erot ja*
- *Maankäytön luokkien suuresti vaihtelevat pinta-alat, jolloin tarvitaan suuri opetusaineisto.*

# Random-Forest -luokittelu algoritmi

## Esimerkki 12

- *Baluchistan on pinta-alaltaan Suomen kokoinen kuiva ja vähämetsäinen Pakistanin maakunta.*
- *Sen metsien tunnistamiseen tarvittiin 1500 visuaalisesti tulkittua koealaa ja yli 20 validointikoealaa per maankäyttöluokka ja luokitteluvuosi, eli yhteensä noin 2000 koealaa.*

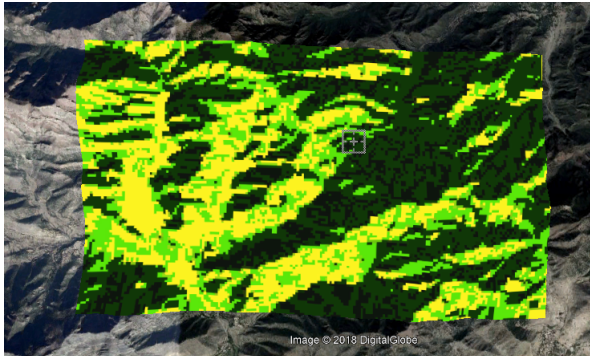
# Baluchistanin metsämaski



## Metsää Baluchistanissa

Kuva: Katri Tegel

# Baluchistanin metsämaski



## Metsää Baluchistanissa

Kuva: Katri Tegel



Numeroista jakaumiin

Koneoppimisen ydinpiirteet

Informaatiogeometria

Esim erkkejä koneoppimisesta metsänarvioinnissa

Maankäytön luokittelu Random Forestilla

"Peliverkot" (?) (GAN)

# Baluchistanin metsämaski



## Metsää Baluchistanissa

Kuva: Katri Tegel

Numeoista jakaumiin

Koneoppimisen ydinpiirteet

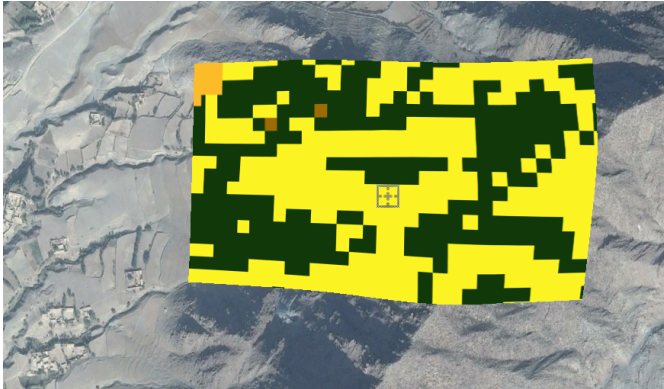
Informaatiogeometria

Esimerkkejä koneoppimisesta metsänarvioinnissa

Maankäytön luokittelu Random Forestilla

"Peliverkot" (?) (GAN)

# Baluchistanin metsämaski



## Metsää Baluchistanissa

Kuva: Katri Tegel



# Peliverkon periaate

## Esimerkki 13

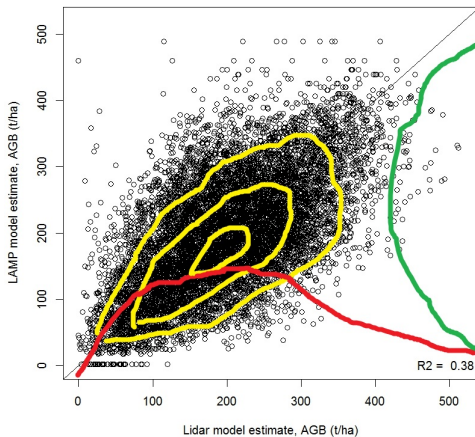
- *Asetetaan kaksi neuroverkkoa taistelemaan toisiaan vastaan*
- *Generaattorin tehtävänä on keksiä huijausaineistoa jota Tunnistaja ei osaa erottaa oikeasta opetusaineistosta*
- *Tunnistajan tehtävänä on luoda yhä hienostuneempia oppimissäntöjä joilla aidon erottaa väärästä*

# Peliverkon periaate

## Esimerkki 14

- *GAN-verkkoja on käytetty varsinkin kuvien tunnistuksessa*
- *Ne saavuttivat kuuluisuutta vuosi sitten kun Googlen GAN-verkko oppi vuorokaudessa tyhjästä luomaan Go-pelaajan joka voitti maailman parhaan ihmispelaajan*
- *Tällöin oppimisen kriteerinä on pelattujen pelien lopputulos - GAN-verkko harjoitteli pelaamalla miljardeja pelejä itseään vastaan*

# Peliverkon maansiirtometriikka



# GAN-verkot rakennusten tunnistamisessa satelliittukuvalta

## Esimerkki 15

- *Tehtävänä tunnistaa rakennukset "ei-rakennuksista" pikseliarvojen pohjalla*
- *Tulos ei ollut hyvä*
- *Generaattori oppi nopeasti liian hyvin tekemään "valerakennuksia" ja*
- *Tunnistaja jäi sen valheiden pauloihin...*

Numeoista jakaumiin

Koneoppimisen ydinpiirteet

Informaatiogeometria

Esimerkkejä koneoppimisesta metsänarvioinnissa

Maankäytön luokittelu Random Forestilla

"Peliverkot" (?) (GAN)

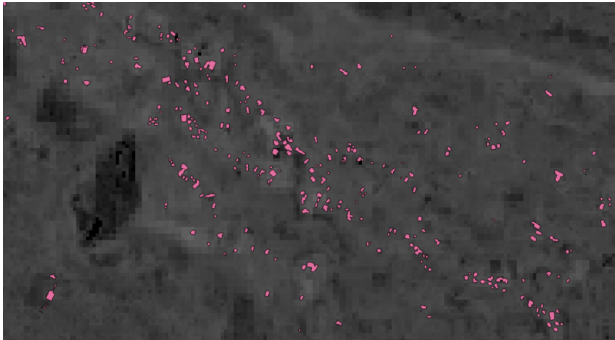
# GAN-verkot rakennusten tunnistamisessa



Rakennuksia vuorilla Gran Canarialla

Kuva: Jarkko Suuronen

# GAN-verkot rakennusten tunnistamisessa



## Rakennuksia vuorilla Gran Canarialla

Kuva: Jarkko Suuronen



Numeroista jakaumiin

Koneoppimisen ydinpiirteet

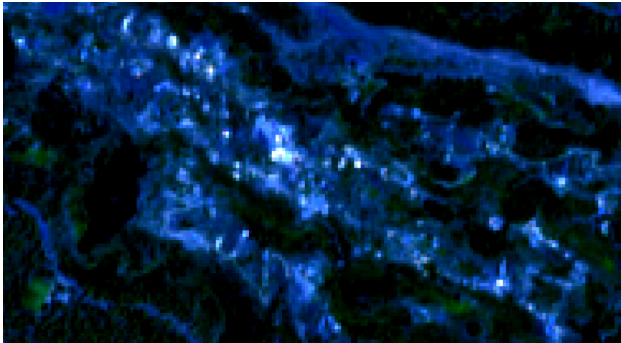
Informaatiogeometria

Esimerkkejä koneoppimisesta metsänarvioinnissa

Maankäytön luokittelu Random Forestilla

"Peliverkot" (?) (GAN)

## GAN-verkot rakennusten tunnistamisessa



Rakennuksia vuorilla Gran Canarialla

Kuva: Jarkko Suuronen