5.7.2021

Koulutin mallit potilailla, joiden tarkkailuaika (”Passed time”) oli enintään 5 vuotta ja pyrin ennustamaan muille potilaille käyttäen näin koulutettuja malleja 🡪 Kehnoja tuloksia.

Toin mukaan uusia malleja ”Ridge” ja ”SVR”. Ridgen koulutin alpha-arvolla 0.5 ja SVR kolmella eri muuttuja C arvolla: 0.3, 0.5, 1.0. Kaikki SVR muuttujat käyttivät kerneliä ”linear”, joskin ”rbf” on kokeiltu ja myös ”polynomial” tyylinen kernel tuli kokeiltua 3. asteella. Linear näyttää tuottavan parhaimpia tuloksia, mutta polynomial oli myös hyvin lähellä. Mahdollisesti eri asteinen polynomial voisi myös toimia. Ridge tuotti samantasoisia tuloksia ”LinRegin” kanssa.

Lasson ohella voisi kokeilla myös ”ElastiNet” mallia.

Pääkomponenttianalyysia harkitsin, mutten usko sen tuloksia pystyvän suoraan soveltamaan koulutusaineistoon, vaan sen tuloksia on tulkittava ja valittava vain joitain yksittäisiä muuttujia malliin.

Voisiko clustering algoritmit avittaa jollain tapaa? Suuri ongelma puuttuvien PET-muuttujien kanssa 🡪 mahdollista vain potilaille, joille tehty PET-tutkimus?

6.7.2021

Pääkomponenttianalyysi paljasti, että potilaat voidaan jakaa kolmella tavalla: iän, painoindeksin ja CTA-muuttujien mukaan eli yleisen suonistun ahtauman mukaan. PET-muuttujat olivat viidennes komponentti erittäin pienellä selittävällä osuudella (0.4% selitys varianssista).

Toin mukaan muuttujat, jotka mittaavat suurimman CTA-tyypin ja suurimman poikkeaman tietyissä segmenttikokonaisuuksissa (LAD, LCX ja RCA). Nämä muuttujat eivät tuottaneet suurta muutosta. Lisäksi poistin käytöstä vastaavat minimiarvot.

Pääkomponenttianalyysin tulokset pysyivät samana, vaikka keskityttiin potilaisiin, joille oli tehty PET-tutkimus. Olisi mielenkiintoista tietää, pysyykö PCA:n tulokset samana, kun muunnellaan datajoukkoa jotenkin muuten, esimerkiksi tutkimalla vain potilaita, jotka ovat olleet tarkkailun alla enintään 5 vuotta. Voi silti olla mielenkiintoista tutkia malleja, jotka ovat koulutettu näillä komponenttimuuttujilla.

Sivussa koulutin myös ElasticNet mallin täysin hyödyttömillä tuloksilla.

8.7.2021

Ensimmäisten PCA-muuttujien implementointi ja testaus.

13.7.2021

PCA-muuttujille unohtui tehdä standardointi, ja analyysi piti siis aloittaa alusta.

Standardoinnin jälkeen saadaan erilaisia PCA-muuttujia riippuen siitä, kohdistetaanko PCA-analyysi vain potilaille, joille on tehty PET-tutkimus, vai kaikille potilaille. Vaikka se kohdistettaisiin vain niille, joille on tehty PET-tutkimus, merkittävämmäksi muuttujaksi (selitysaste 0.31) jää muuttuja, joka reagoi vahvasti PET-muuttujiin, erityisesti suurimpiin poikkeamiin pet-arvoissa. Myös poikkeavien pet-segmenttien lukumäärä on merkittävä tekijä tässä muuttujassa. Erityisesti tämä muuttuja kohdistaa huomion potilaihin, joilla on poikkeavia pet-arvoja, mutta pieni lukumäärä tyyppien 2 ja 3 stenooseja ja pieni sis-arvo. Myös CTA-segmentit LCX-A, RCA-A, RCA-B, RCA-C korostuvat tässä muuttujassa hieman.

Toinen PCA-muuttuja (selitysaste 0.11) taas erottaa potilaat, joilla on normaalit PET-arvot, mutta korkeat CTA-arvot, ja jälleen pieni lukumäärä tyyppien 2 ja 3 stenooseja ja pieni sis-arvo. Muuttuja lisäksi reagoi lievästi ikään ja diabetekseen.

Kolmas PCA-muuttuja (selitysaste 0.044) vahvasti reagoi potilaan CTA-tulokseen LPD (0.74), LOM2 (0.48), ja LPL (0.49) segmenteissä. Erityisesti muuttuja pyrkii erottelemaan potilaat, joilla on korkeat arvot edellämainitissa segmenteissä ja pienet arvot segmenteissä RCA-C, RPD, RPL ja IM.

Neljäs PCA-muuttuja (selitysaste 0.040) erottelee potilaat, joilla on korkeat RPD, RPL (ja hieman LPL ja LPD) arvot, mutta pienet arvot segmenteissä LM, LAD-A, LAD-B, LAD-C, D1, LCX-A, LCX-B, LCX-C, ja IM.

Viides PCA-muuttuja(selitysaste 0.032) erottelee potilaat, joilla on korkea CTA-arvo segmenteissä LOM2 (0.8334), D2 (0.31), mutta pienet arvot segmenteissä LPD, LOM1, IM ja LPL. SIS ja SSS muuttujat myös hieman korreloivat tämän kanssa.

20.7.2021

PCA-muuttujat, kun PCA-analyysi kohdistetaan kaikille potilaille, eikä vain niille, joille on tehty PET-kuvaus:

PCA1 (selitysaste 0.28):

Stenosis type 2 count, max cta type, max high risk cta type, kaikki max str seg muuttujat ja poikkeavien pet-segmenttien lukumäärä saivat kaikki yli 0.2 arvoja. Lisäksi LAD-A, LCX-A ja RCA-A saivat lähes 0.2. Stenosis type 1 ja SIS saivat alle -0.2 arvoja. Muuttuja erottelee siis potilaat, joilla on vakavia ahtaumia niistä potilaista, joilla on kokonaisvaltaisia, mutta lievempiä ahtaumia.

PCA2 (selitysaste 0.12):

Max str seg -muuttujat, poikkeavien pet segmenttien lukumäärä ja tyyppien 2 ja 3 stenoosien lukumäärä saavat arvoja 0.2 lähellä. Lisäksi kaikki pet segmentit saavat arvoja yli 0.15. Sen sijaan tyyppien 4, 5 ja 6 stenoosien lukumäärä, max cta type ja max high risk cta type saavat kaikki arvoja alle -0.15. Muuttuja erottelee siis potilaat, joilla on pet-tuloksissa havaittavissa merkittäviä poikkeamia useissa segmenteissä, mutta cta-tulokset saavat pieniä arvoja niistä potilaista, joilla on merkittäviä ahtaumia useissa segmenteissä, mutta tämä ei näy pet-tuloksista.

PCA3 (selitysaste 0.037)

Muuttuja erottelee potilaat, joilla on korkeita cta-arvoja segmenteissä LCX-C, LPD, LOM1, LOM2, LPL, RPD, RPL, LCX-B, ja joilta löytyy hypertension, dyslipidemia, ja dyspnea. Lisäksi näillä potilailla on pienet arvot muuttujissa ikä, LAD-A, max cta type ja max high risk cta type. Eli muuttuja erottelee potilaat, jotka ovat vanhoja ja joilla on kohonneita cta-arvoja proksimaalisimmissa suonissa, mutta jolla ei ole hypertension, dyslipidemiaa tai dyspneaa.

PCA4 (selitysaste 0.033)

Muuttuja kohdistaa huomion potilaisiin jotka ovat: nuoria tupakoivia miehiä, joilla on kohonneita arvoja segmenteissä LPD, LOM2, LPL, ja normaaleja arvoja segmenteissä RCA-C, RPD, RPL. Nämä erotetaan siis potilaista, jotka ovat vanhoja tupakoimattomia naisia, joilla kohonneita arvoja segmenteissa RCA-C, RPD, ja RPL ja normaaleja LPD, LOM2, LPL.

PCA5 (selitysaste 0.030)

Sama kuin edellä: Vanhat tupakoimattomat naiset, joilla ei ole hypertension, dyslipidemia, dyspnea, kohonneita CTA-arvoja segmenteissä LPD, LOM2, LPL ja normaaleja segmenteissä IM.

23.7.2021

PCA1-komponentti tekee selvän jakauksen potilaille. Vain 0.5%-5% potilaista, joilla on pienet PCA1-arvot saavat tapahtuman tarkasteluvuosista riippuen, kun taas 9%-25% potilaista, joilla on korkeat PCA1-arvot saavat tapahtuman. PCA1-komponentti oleellisesti jakaa potilaat niihin, joilla on kokonaisvaltaista, mutta lievää tukkeutuneisuutta suonistossa, ja niihin, joilla on useita vakavia ahtaumia, joten tämä ei ole yllätys. Jako on kuitenkin niin radikaali, että tätä kannattanee ottaa huomioon malleja luodessa tulevaisuudessa.

PCA2-komponentti taas jakaa potilaat niihin, joilla on kohonneet PET-tulokset, mutta normaalit CTA-tulokset. 2 ja 3 vuoden sisällä tarkasteltaessa, tapahtumat ovat tiheimmillään PCA2-komponentin ääriarvoissa, kun taas jos tarkasteluvuotta ei rajoiteta, ainoastaan pienillä PCA2-komponentin arvoilla näkyy vastaavaa ilmiötä.

Edellä mainitut tulokset pysyvät samoina, kun rajoitutaan vain potilaisiin, joille on tehty PET-tutkimus.

**Ensi kerralla:**

* **PCA-tarkastelua**

**Tulevaisuudessa:**

* **Pyri ennustamaan potilaan elinaikaa (eli regressio kohdemuuttujalla ’passed time’)**
* **Kehitä Deep Learning modelia.**