Datalogiens Videnskabsteori 2024

Ugeseddel 6

Indhold

Fælles info uge 5

I denne uge er vi tilbage til det almindelige skema med øvelser mandag/fredag eller tirsdag og Flipped forelæsning mandag. Derudover har vi den anden klassiske forelæsning ved Mikkel fredag den 31. maj kl 10.15 – 12.00 i auditorium 4 på HCØ.

Alle materialer til uge 5

Forelæsningsvideoer

- Sørensen and Johansen (2021), kap. 6b: *At få computeren til at hjælpe os: Del II — datadrevne modeller*
- Video uge 6-1: Gæsteforelæsning v/ Irina Shklovski: Data construction and usefulness of models (optaget Zoom forelæsning)
- Video uge 6-2: Machine Learning epistemologi (optaget Zoom forelæsning)
- Video uge 6-3: Hvordan andre ser os - og hvad det betyder for os

Tilhørende kapitler fra grundbogen

- Sørensen and Johansen (2021), kap. 6b: *At få computeren til at hjælpe os: Del II — datadrevne modeller*

Grundbogskapitel til klassisk forelæsning

- Sørensen and Johansen (2021), kap. 10: *Etik, redelighed og privacy* (læs afsnit 10.6 – 10.10)

Tekster til øvelserne

- Justitsministeriet (2018) (ikke pensum)
- Miller et al. (2017)
-

Note til forelæsningvideoerne: Video 6-1 er en gæsteforelæsning fra 2021, som vi har fået lov til at genbruge. Video uge 6-2 henviser til en gammel video om Embodied AI, som vi har opdateret i år, så I ser altså ikke præcis den video, der henvises til. Derudover nævner video 6-2 en gammel eksamensform og en tekst om AI, der ikke længere er pensum.

Centrale begreber

- alt data er socialt konstrueret
- case: Google Flu trends
- epistemiske problemer med ML og big data:
  - de mange dimensioners problem
  - problemet med sammenfaldende fænomener
  - korrelation medfører ikke kausalitet
  - automation bias
- big data etik:
  - informeret samtykke
  - privacy
  - ejerskab
  - epistemologi
  - big-data skellet
- kontekstuel integritet
- algoritmisk diskrimination og fairness
- systemisk etik
- Rybergs undskyldninger
- videnskabelig uredelighed

Flipped Forelæsning 6 mandag den 27. maj

Til Flipped forelæsningen vil vi samle op på quiz-svarene fra uge 5 og lave aktiviteter om de centrale begreber fra uge 6.

Klassisk forelæsning fredag den 31. maj

Den klassiske forelæsning vil samle op på emnet fra sidste uge om etiske teorier og dække emnerne videnskabelig redelighed og professionelt ansvar.

Læringsmål

Efter denne uge skal du være i stand til at forklare de centrale epistemiske problemer ved Big Data-metoder og kunne bruge dem i en diskussion, herunder skal du kende casen om Google Flu trends. Du skal kunne forklare sammenhængen mellem epistemiske og etiske udfordringer ved modelbaserede eller automatiske beslutninger og du skal kunne reflektere over datalogens etiske og professionelle ansvar.

Øvelser uge 6

MANDAG 27. MAJ / FREDAG 31. MAJ ELLER TIRSDAG 28. MAJ

Emne 1: Explainable AI

Læringsmål

Efter dette emne skal du være i stand til at forklare hovedproblemerne i XAI-tilgange, som de præsenteres i Miller et al. (2017), samt diskutere dem i forhold til gældende lovgivning om AI-baserede beslutninger.

Litteratur (tilgængelig på Absalon)

- Justitsministeriet (2018) (ikke pensum)
- Miller et al. (2017)

Fremlæggelse gruppe 6: XAI

1. Forklar først forskningsspørgsmålet i den „metaanalyse“ (survey), som Miller et al. (2017) har udført.
2. Lav en figur, der viser 'den, der forklarer', 'den, der forklares til', 'det, der forklares' og 'det, hvorudfra der forklares' (samt evt. andre elementer).
3. Forklar derefter de forskellige „forklaringsbegreber“, som artiklen har identificeret. Kom med illustrationer af, hvor de forskellige forklaringstyper kan finde anvendelse.
4. Diskutér hvordan man kan forestille sig at tilpasse og udvikle machine learning og anvendelser deraf i retning af at kunne give bedre forklaringer.

Gruppearbejde emne 1

I Databeskyttelsesloven (opdateret i 2024) står der bl.a.

> Den registrerede har ret til ikke at være genstand for en afgørelse, der alene er baseret på automatisk behandling, herunder profilering, som har retsvirkning eller på tilsvarende vis betydeligt påvirker den pågældende. (Justitsministeriet, 2018, Artikel 22.1).

og

> Den registrerede har ret til at få den dataansvarliges bekræftelse på, om personoplysninger vedrørende den pågældende behandles, og i givet fald adgang til personoplysningerne og følgende information:
>
> h. forekomsten af automatiske afgørelser, herunder profilering, som omhandlet i artikel 22, stk. 1 og 4, og som minimum meningsfulde oplysninger om logikken heri samt betydningen og de forventede konsekvenser af en sådan behandling for den registrerede. (Justitsministeriet, 2018, Artikel 15.1).

Profilering er i lovteksten defineret som:

> (...) enhver form for automatisk behandling af personoplysninger, der består i at anvende personoplysninger til at evaluere bestemte personlige forhold vedrørende en fysisk person, navnlig for at analysere eller forudsige forhold vedrørende den fysiske persons arbejdsindsats, økonomiske situation, helbred, personlige præferencer, interesser, pålidelighed, adfærd, geografisk position eller bevægelser"(Justitsministeriet, 2018, Artikel 4.4).

Dvs. tilfælde, hvor indsamlede oplysninger anvendes til at lave profiler af en person til at forudse eksempelvis adfærd eller fremtidige behov.

Spørgsmål til diskussion

Diskuter i grupper følgende spørgsmål og hver klar til at præsentere jeres overvejelser i plenum:

1. Kan I forestille jer at lovkravet har en indflydelse på hensyn til modelpræcision? Er det et problem? Begrund jeres svar.
2. Er det overhovedet muligt at opfylde lovkravet for en virksomhed eller institution, givet "black-box"-opfattelsen af AI?
3. Diskuter hvordan man som modelkonstruktør skal forholde sig til de juridiske begrænsninger til brug i udviklingen af algoritmiske beslutninger baseret på machine learning. Er XAI den eneste vej frem? (se evt illustration nedenfor fra Richmond et al. (2024), s. 9)

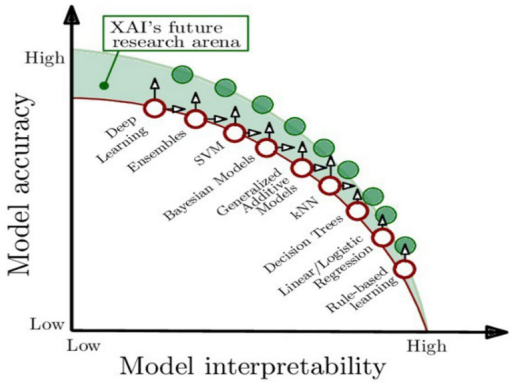

Fig. 2 The trade-off between explainability and performance, adapted from Barredo et al. (2020)

Emne 2: ML og bias: epistemologi

Læringsmål

Efter dette emne skal du være i stand til at forklare og udpege epistemiske problemer i ML-metoder, samt reflektere over de valg, der er en del af modelleringsprocessen.

Litteratur

- Kruse et al. (2017)
- Kruse (2018) (ikke pensum)
- Sørensen and Johansen (2021), kap. 6a og 6b
- Johansen and Sørensen (2018)

Fremlæggelse gruppe 7: ML og bias - forudsigelse af hoftefraktur

Gruppefremlæggelsen tager udgangspunkt i teksten Kruse et al. (2017), som beskriver konstruktionen af en model, der kan bruges til at forudsige hoftefrakturer. Teksten er en autentisk forskningsartikel, så du skal ikke regne med at du kan forstå alle de tekniske detaljer, og det forventer vi heller ikke, at du kan. Vi anbefaler, at du læser artiklen igennem for at få et overblik, og derefter går i dybden med de dele, der er relevant for fremlæggelsen. Du kan finde en populærvidenskabelig introduktion til forskningsprojektet i Kruse (2018).

1. Redegør med jeres egne ord for det overordnede mål og metoder i det forskningsprojekt, der beskrives i artiklen Kruse et al. (2017).
2. Beskriv med jeres egne ord nogle af de overvejelser, forfatterne gør sig om *over- og underfitting* (Kruse et al., 2017, 351). Hvordan kan det være, at man kan stå i en situation, hvor model A fungerer bedre på træningsdata end model B, mens model B fungerer bedre på valideringsdata end model A?

3. Beskriv nogle af de valg — fx af datakategorier mv. — som forskerne i projektet måtte gøre i forbindelse med konstruktionen af modellen. Bemærk, at det ikke er alle valg, der beskrives eller motiveres i artiklen. I må derfor gøre jer jeres egne overvejelser. I kan eventuelt inddrage modelfiguren fra Sørensen and Johansen (2021), kap. 6a, s 4

I afsnittet *Strengths and Limitations* står følgende:

> This study was designed and implemented to limit the human involvement and risk of information bias from the databases as much as possible. The aim was to separate the data from the interpretation. (Kruse et al., 2017, 357)

Forfatterne hævder her, at et bestemt skridt i modelleringsprocessen var objektivt i den forstand, at menneskelig fortolkning og indblanding var elimineret.

4. Diskutér, om modelleringsprocessen som helhed på samme måde var objektiv, dvs. renset for menneskelig fortolkning og vurdering. Inddrag fx delopgave 3 ovenfor i jeres diskussion.

Gruppearbejde emne 2

I skal I grupper overveje følgende spørgsmål, som bagefter diskuteres i plenum:

1. I teksten af Johansen and Sørensen (2018) og i forelæsningsvideoerne har vi hørt om tre overordnede problemer, der opstår når man modellerer med ML: Problemet med mange dimensioner, problemet med sammenfaldende fænomener og induktionsproblemet. Forklar kort hinanden hvad de epistemiske problemer går ud på.
2. Overvej om de epistemiske problemer ved modellering er til stede ved casen om hoftefrakturer. Hvordan optræder de? Er der andre problemer ved casen?
3. Diskuter hvorvidt det er muligt at lave en objektiv model. Beskriv herunder hvad I forstår ved objektivitet.
4. Vurdér overordnet, hvilke erkendelsesmæssige muligheder og begrænsninger, machine learning indebærer. I kan fx. vende tilbage til diskussionen af logisk positivisme. Her kan I for eksempel inddrage Chris Anderson og Google Flu trends casen.

Litteratur

Johansen, M. W. and Sørensen, H. K. (2018). Big Datas Titanic? *Aktuel Naturvidenskab*, 3.

Justitsministeriet (2018). Lov om supplerende bestemmelser til forordning om beskyttelse af fysiske personer i forbindelse med behandling af personoplysnin-

ger og om fri udveksling af sådanne oplysninger (databeskyttelsesloven). https://www.retsinformation.dk/eli/lta/2018/502.

Kruse, C. (2018). Big data er som en læge, der har set 3 millioner patienter. https://videnskab.dk/krop-sundhed/big-data-er-som-en-laege-der-har-set-3-millioner-patienter. videnskab.dk.

Kruse, C., Eiken, P., and Vestergaard, P. (2017). Machine learning principles can improve hip fracture prediction. *Calcified Tissue International*, 100:348–360.

Miller, T., Howe, P., and Sonenberg, L. (2017). Explainable AI: beware of inmates running the asylum or: How I learnt to stop worrying and love the social and behavioural sciences. *CoRR*, abs/1712.00547.

Richmond, K. M., Muddamsetty, S. M., Gammeltoft-Hansen, T., Olsen, H. P., and Moeslund, T. B. (2024). Explainable ai and law: An evidential survey. *Digital Society*, 3(1).

Sørensen, H. K. and Johansen, M. W. (2021). Invitation til de datalogiske fags videnskabsteori. Lærebog til brug for undervisning ved Institut for Naturfagenes Didaktik, Københavns Universitet. Under udarbejdelseyear.