

Kommentar justeringer SOS2901

Emnesiden

Jeg har endret emnebeskrivelsen med det formålet å tydeliggjøre temaet for emnet, herunder hva fokuset er. Noen begreper er kanskje litt kryptiske for de som ikke kjenner til det fra før, men gir en bedre beskrivelse for f.eks. arbeidsgivere eller andre som lurer på hva de faktisk har lært.

Jeg har prøvd å fremstille emnet som et moderat teknisk nivå, men forutsette grunnleggende statistikk og databehandling tilsvarende det som dekkes på sosgeo1120. (Å ta kurset *før* sosgeo1120 vil være krevende, men samtidig vil nok kunne gå).

Undervisning

Det blir ingen tradisjonelle forelesninger unntatt en introduksjonsforelesning. Det vil derimot lages korte videoforelesninger til hver uke. Undervisningen kommer ellers til å være lagt opp med aktiv problemløsning med datasett og R.

Det vil ikke være noen krav for obligatorisk oppmøte eller aktivitet. Derimot vil undervisningen forutsette at studentene stiller forberedt. De skal levere inn noe i Canvas hver uke, og det er inngangsbilletten for å delta i undervisningen. De som ikke har levert i Canvas slipper altså ikke inn i rommet. Disse oppgavene vil ikke være arbeidskrevende, men jeg krever altså et minimum av forberedelse. Det kan være noen få quiz-oppgaver, et kort refleksjonsnotat, forberede noen spørsmål etc.

Pensum

Hovedboken er læreboken til Berk, men supplert med kortere kapitler for å dekke ROC curve, unsupervised learning (dvs. klusterteknikker og datareduksjon), og statistiske rettferdighetsmål. Studentene vil i undervisningen få noe mer presiseringer på hvilke delkapitler i Berk som vektlegges, da vi i praksis vil hoppe over enkelte underkapitler.

Boken til Berk og kapitlet om unsupervised learning baserer seg på R. Det er meget mulig jeg vil bruke andre funksjoner og praktiske løsninger enn det, bl.a. for å tilpasse til hvordan det er gjort på sosgeo1120. Oversikt over koder og script gis separat underveis i semesteret.

I tillegg til det rent metodiske er det en pensumbolk med mer kritiske perspektiver, og en pensumbolk med «casestudier». Casene er konkrete empiriske anvendelser som illustrerer bruken av maskinlæring. Jeg har forsøkt å spre eksemplene over ganske ulike typer temaer: kriminalitet, helse, HR, helse, kredittvurderinger og identifisering av Twitter-troll. Eksemplene bruker i hovedsak den typen algoritmer som kurset dekker: regresjon, klassifikasjonstrær, random forest og XGboost. Rettferdighet og tuning er i mindre grad dekket i casene, men det vil være en naturlig inngang i undervisningen.