

Begrunnelse for forslag til programrådet for SOS2900

I forslaget til pensumrevisjon foreslo jeg om følgende:

Oppgaver og løsninger i seminarene er å betrakte som pensum.

En viktig bakgrunn her er at det finnes få lærebøker og annet materiell rettet mot studenter som ikke er fra langt mer tekniske fag enn sosiologi. Det er u hensiktsmessig å gå for langt den veien da det vil fortrenge annet mer substansielt stoff. Målsettingen er at studentene skal lære akkurat nok om viktige prinsipper slik at de skjønner hvordan algoritmene fungerer.

De oppgavene som er hands-on med data og pc har i all hovedsak det substansielle dekket i Berk (2016), herunder basis R-koder. Det er imidlertid en rekke små justeringer og måter å gjøre ting på som ikke står like tydelig. Dette handler mest om software, spesialfunksjoner osv.

Et eksempel er at Berk (2016) gir prinsippet for cross-validation og andre resampling metoder, men går ikke inn i k -fold-cross-validation med kode. Heller ikke hierarchical clustering og KNN-klassifisering er dekket av Berk (2016), men prinsippet er lett å forklare og krever bare et par linjer kode å gjøre.

Et annet eksempel er analyse av sosiale nettverksdata ved å predikere individers egenskaper som de ikke har oppgitt i egen profil. Dette er en enkel forlengelse av håndtering av nettverksdata (dekket i Luke 2015) og klassifiseringsteknikker (dekket i Berk 2016). Det som mangler er relativt enkle grep for å hente ut informasjon fra nettverksdata og putte inn i en algoritme.

Å legge opp svært tekniske pensumbidrag som gjerne dekker langt mer enn det som trengs på SOS2900 er primært unødvendig. Alle teknikker er for øvrig dekket i oppgavene på www.datacamp.com, og det kan jo eventuelt oppføres som en interaktiv lærebok der utvalgte kapitler er pensum.

Alternativ til mitt forslag om å la oppgaver definere deler pensum er å supplere med utdrag fra flere langt mer tekniske tekster som f.eks. følgende:

James, Witten, Hastie og Tibshirani (2014) *An introduction to statistical learning*, Springer

Aggarwal (2011) *Social Network Data Analytics*. Springer, Boston, MA

Begge disse er tilgjengelig online fra UiO og krever ikke kompendium. Men det vil kreve noe mer tid å plukke ut de relevante utdragene.