

## Dirichlet Prosesser i pålitelighetsanalyser av maskete strømnettverk

Dirichlet prosesser faller inn under ikke-parametrisk Bayesiansk statistikk, og en populær modell er "The Infinite Gaussian Mixture Model". Denne modellen kan brukes både som grupperings- (clustering) og klassifiseringsmodell. For klassifisering blir denne modellen som en slags avansert utgave av "Naive/Idiot Bayes".

Innen pålitelighetsanalyser av maskete strømnettverk er det mest aktuelt å bruke denne modellen som er klassifiseringsalgoritme, hvor man klassifiserer driftstilstander. En driftstilstand er et stillbilde av nettverket, altså hvordan strømmen flyter i nettverket i et gitt øyeblikk. Flyten i nettverket finner man ved å løse et sett ikke-lineære ligninger, hvor inngangsdataene er produksjon og last på gitte punkter i nettverket. I pålitelighetsanalyser er man "kun" interessert i de driftstilstandene som fører til overlast på en eller flere linjer. Man er derfor interessert i å klassifisere driftstilstandene inn i to grupper, de som fører til overlast og de som er "ok", slik at man kan unngå å løse det ikke-lineære ligningssettet for "uinteressante" tilstander.

Oppgave/Prosjekt:

- Generell forståelse av ikke-parametrisk Bayesianske modeller, og spesielt "The Infinite Gaussian Mixture Model" (**Kommentar:** Jeg har laget en slags rapport på denne modellen, men denne er veldig "uferdig".)
- "The Infinite Gaussian Mixture Model" er implementert i C++, men koden kan forbedres og er ikke helt feilfri. Koden må derfor utvikles videre. Alternativt finnes det noen modeller implementert i MatLab/R som man kan se på. (**Kommentar:** Det krever en del å sette seg inn i disse modellene. Koden min i C++ kan forbedres mye, men jeg vil anbefale å programmere i C++. Håkon Tjelmeland har hjulpet meg med dette tidligere.)
- Kort innføring i pålitelighetsanalyser av maskete strømnett. (**Kommentar:** Dette er vel strengt tatt ikke nødvendig, men det er vel lurt å sette seg kjapt inn i dette så man vet litt om hva man holder på med.)
- Hvor effektiv er "The Infinite Gaussian Mixture Model" som klassifiseringsalgoritme i pålitelighetsanalyser? Her eksisterer det noen testdatasetter som kan brukes. I tillegg blir det viktig å tenke på hvilke variabler/inngangsdata man skal inkludere i modellen, slik at man unngår "The curse of dimensionality".
- Hvis disse metodene skal brukes i praksis kreves det at de er effektive: de må kunne klassifisere driftstilstander både effektivt og raskt. De fleste modellene som eksisterer i dag er ganske "treige", slik at en viktig del av oppgaven blir å finne en god/effektiv implementering. (**Kommentar:** Se på Bayesian hierarchical clustering?)