

Noget om Machine Learning måske?

P5 Semestersprojekt - Efteråret 2016

Gruppe 5405

Titel:

Noget om Machine Learning
måske?

Tema:

Dunno.. kig i studieordningen?

Projektperiode:

P5, Efteråret 2016
??/??/???? - ??/??/????

Projektgruppe:

5405

Synopsis:

Medvirkende:

NAVN
NAVN
NAVN
NAVN
NAVN

Vejleder:

NANO

Sider: 73

Appendikser: 1

Afsluttet: 27/05/2016

Offentliggørelse af rapportens indhold, med kildeangivelse, må kun ske efter aftale med forfatterne.

Forord

Indholdsfortegnelse

1	Indledning	1
1.1	Initierende problemstilling	1
2	Problemanalyse	2
2.1	Overbelægning	2
2.1.1	Konsekvenser af overbelægning	3
2.2	Machine learning	4

Indledning 1

Der er i dag overbelægning på flere afdelinger på de danske hospitaler, hvoraf nogle afdelingerne berøres i hele og flere måneder ad gangen. [2015] Overbelægning resulterer i, at sundhedspersonalet får mindre tid pr. indlagt patient. Ifølge en undersøgelse fra Dansk Sygeplejeråd, mener hver anden regionalt ansat sygeplejerske på tværs af regionerne, at den travle arbejdsdag går ud over patienternes sikkerhed [Kjeldsen2015]. Et studie påviser, at ved blot én ekstra indlagt patient pr. sygeplejerske øges mortalitetsraten med 7 % indenfor en indlæggelse på 30 dage ¹ [Aiken2014].

Foruden sundhedspersonalets øgede risiko for at begå fejl ift. behandlingen af patienter, er der ligeledes en sikkerhedsrisiko forbundet ved overbelægning på hospitalerne. De ekstra patienter, der ligger på stuerne, gangene og vaskerummene, pga. overbelægning, er en større udfordring ved evakuering under brand. Pladsmangel, som medfører, at patienterne opholder sig i vaskerummene og på gangene, bevirker desuden til, at patienterne oplever et skærpet privatliv. [Madsen2014]

Aalborg universitetshospital har i et tidligere projekt indsamlet data fra 1.000 hospitalsindlæggelser. Disse data inkluderer blodprøveanalyser og knoglescanninger (DXA), hvilket formodes at kunne anvendes til udvikling af en prædiktiv model, der kan estimere indlæggelsesvarigheden blandt patienter. Denne rapport vil på baggrund af dette undersøge, hvorvidt det er muligt at forudsige indlæggelsesvarigheden ved brug af machine learning. ²

1.1 Initierende problemstilling

³ Hvordan påvirkes personalet og patienterne af overbelægning på hospitaler, og hvilke konsekvenser har overbelægning i forhold til sikkerheden på afdelingerne? Hvilke typer af machine learning findes der, samt hvor benyttes det i dag?

¹FiXme Note: Har stadig lidt svært ved om man kan forstå sætningen på den rigtige måde.

²FiXme Note: mangler stadig noget om begrundelsen for valg af machine learning

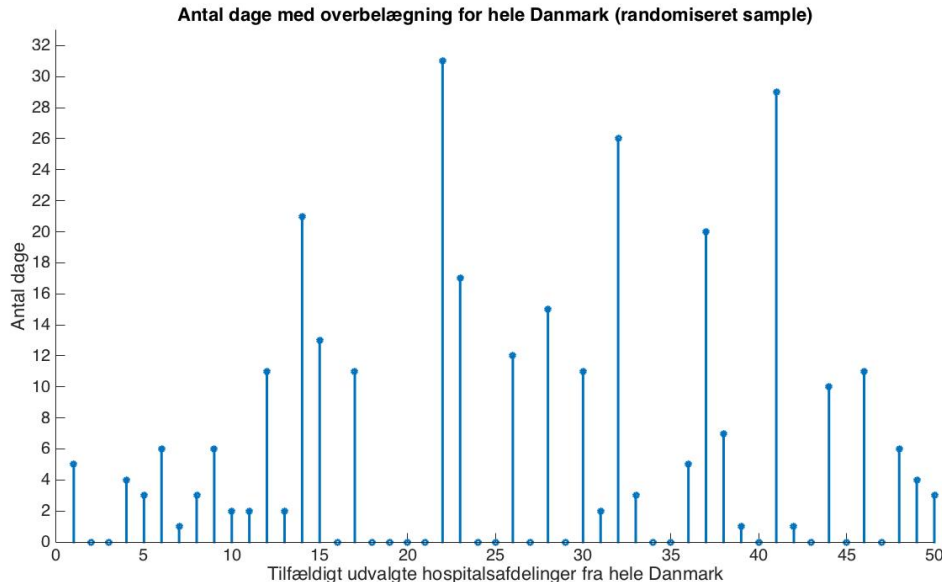
³FiXme Note: Dette er ikke den endelige problemstilling

Problemanalyse 2

2.1 Overbelægning

Overbelægning opstår, når en afdeling har en belægningsprocent på over 85. I perioden fra år 1996 til 2011 er sengepladserne på de danske hospitalsafdelinger reduceret med 30 %, for således at tilnærme sig en fuld udnyttelse af ressourcerne. Herunder udnyttelse af de tilgængelige sengepladser på hospitalsafdelingerne samt optimering af sundhedspersonalets arbejdstid [Madsen2014].

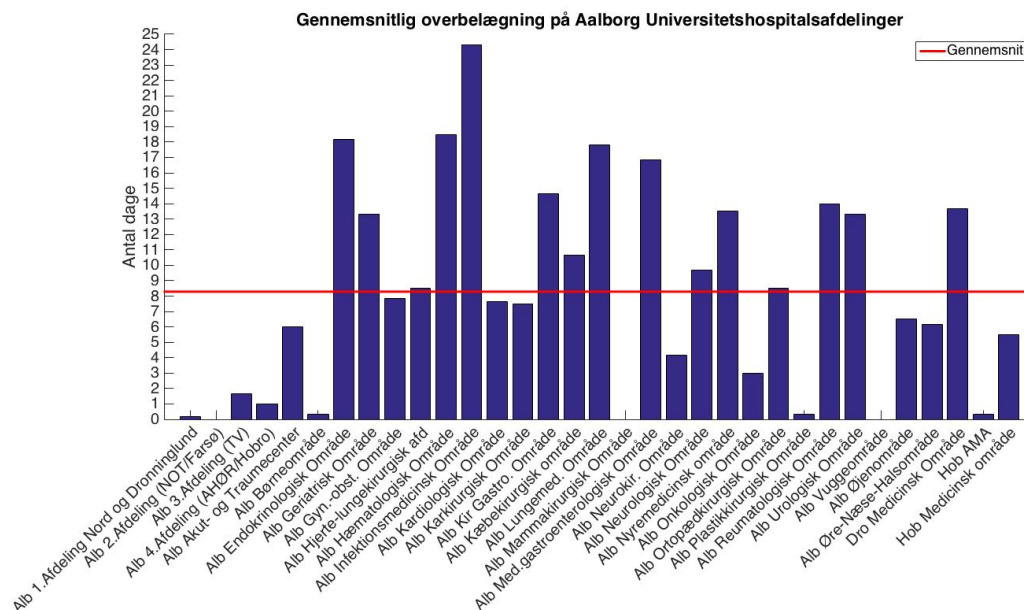
Antallet af indlagte patienter varierer hver måned, hvorfor tilgængeligheden af sengepladser ligeledes varierer. På figur 2.1 er det illustreret, hvordan de forskellige afdelinger på tværs af Danmark oplever overbelægning i januar måned år 2015.



Figur 2.1: Illustrerer antal dage med overbelægning på 50 tilfældige hospitalsafdelinger fra hele Danmark. De tilfældige data er taget fra januar måned år 2015. [SDS2015]

Det ses af figur 2.1, at der er en tendens til overbelægning på flere hospitalsafdelinger i Danmark. Ud af de 50 tilfældige hospitalsafdelinger har 35 af afdelingerne oplevet overbelægning i januar måned. Herudover ses det af figuren, at nogle afdelinger

har haft overbelægning op til 31 dage, hvilket svarer til, at afdelingen har haft overbelægning alle dage i januar måned.



Figur 2.2: Søjlerne illustrerer gennemsnittet af antal dage med overbelægning på Aalborg universitetshospitalsafdelinger fra januar til juni år 2015. [SDS2015] Den røde linje viser gennemsnittet af søjlerne. Denne er beregnet til en gennemsnitlig overbelægning på 8,29 dage.

Det fremgår af figur 2.2, at der ligeledes er en tendens til overbelægning på Aalborg universitetshospital. Der ses variation i antal dage med overbelægning på afdelingerne, dog ses overbelægning på 30 ud af 33 afdelinger i perioden fra januar til juni år 2015. På ambulatorisk infektionsmedicinsk område opleves en gennemsnitlig overbelægning på 24 dage, hvorimod eksempelvis ambulatorisk 2. afdeling ikke berøres af overbelægning. Gennemsnitlig er der en overbelægning på 8 dage for afdelingerne i perioden januar til juni år 2015.

2.1.1 Konsekvenser af overbelægning

Overbelægning på hospitaler medfører forlængede arbejdsdage, hvilket viser sig at have en negativ indvirkning på sundhedspersonalet. [Kjeldsen2015] [Dinges2004] Overbelægning resulterer i, at sundhedspersonalet har kortere tid til den enkelte patient, hvorfor risikoen for fejl øges. Ifølge et amerikansk studie fra år 2002 opstår fejlene hovedsageligt, når personalet har arbejdsdage på mere end 12 timer. [Dinges2004] Et andet amerikansk studie fra år 2014 indikerer en forøgelse af indlæggelsestiden ved forøget arbejdsbyrde for sundhedspersonalet [Elliott2014]. Overbelægning medvirker derfor til flere sengedage for patienterne. I takt med redueringen af sengepladser på 30 % som beskrevet i section 2.1, er sygeplejerskernes arbejdsbyrde øget med 40 % fra år 2001 til 2015 ifølge en dansk undersøgelse. Derudover viser under-

søgelsen, at dette resulterer i en stresset arbejdsdag og dermed en forringet kvalitet af behandlingen.[Kjeldsen2015]

Det erfares, at sundhedspersonalet bliver stressede og ikke er i stand til at give patienterne den optimale behandling ved forøget patientbyrde. [Aiken2002] Ved overbelægning er der flere patienter på stuerne, hvorfor det kan være nødvendigt at flytte nogle af patienterne ud på gangene og vaskerummene. Overbelægning kan derfor have en negativ effekt på både sundhedspersonalets og patienternes behandlingsforløb. Patientbyrden kan ligeledes være en hindring i tilfælde af brand under evakuering.

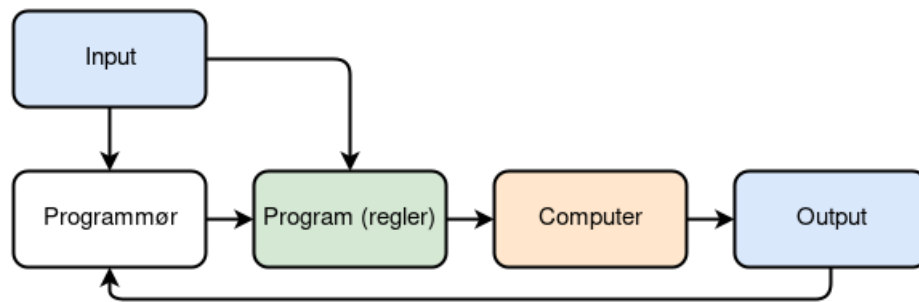
Ifølge et dansk studie fra år 2014, øges risikoen for mortalitet med 1,2 % ved en øget belægning på 10 %. [Madsen2014] Et canadisk studie understøtter, at en forøgelse med 10 % i belægning på akutafdelingen vil medføre en øget mortalitet og flere genindlæggelser.[McCusker2014] Disse tal sammenlignes ikke direkte med danske tal, da der er tale om forskellige sundhedsvæsener, men tendensen kan dog ses som en indikation for, at overbelægning på sygehuse har en negativ effekt på mortalitetsraten.

2.2 Machine learning

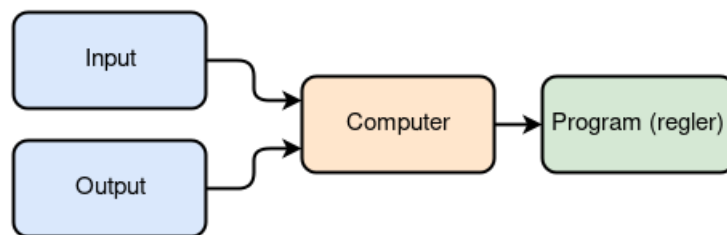
Machine learning er en underkategori af kunstig intelligens og er en metode, der kan anvendes til at gøre en computer i stand til at genkende mønstre i data. Derved er det muligt at bearbejde datasæt af en størrelse, der er umulige for mennesker at behandle. Det bruges uset mange steder og har mange formål, som f.eks. søgemaskiner, forudsigelse af aktiekurser og ansigtsgenkendelse.[DIKU2010]

En konventionel computeralgoritme arbejder efter et input og et regelsæt der afgør hvordan dette input skal behandles. Dette resulterer i et output, begrænset af programmørens evne til at opstille regler. Machine learning arbejder derimod typisk efter et input, hvilket kaldes data, og et output, hvilket kaldes label. Det betyder at machine learning selv skaber reglerne for hvordan man kommer fra data til label. Når systemet har et solidt regelsæt kan det selv lave labels til ny lignende data. Forskellen er illustreret på figur 2.3. Dette gøres i en logisk proces kaldet induktiv inferens, hvilket er en metode, hvor computeren tilnærmer sandsynligheden for et givent output, ud fra en sammenligning med lignende eksempler. Der findes forskellige måder at gøre dette på, og, men fælles for alle undersøges det, hvorvidt den nye data ligner den eksisterende data.[DIKU2010]

Traditionel programmering



Machine learning



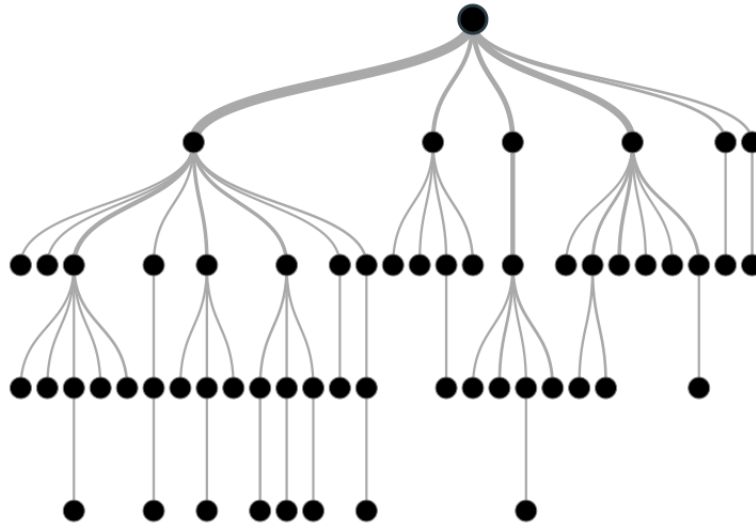
Figur 2.3: Det traditionelle programmeringsparadigme i forhold til machine learning

Figuren viser forskellen mellem den traditionelle tilgang til algoritmedesign i forhold til machine learning.

Når en algoritme designes til machine learning gøres dette på baggrund af et træningssæt. [DIKU2010] Træningssættet tager udgangspunkt i en datamængde der enten har kendte eller ukendte labels. Hvis indholdet af de forskellige samples i datasættet har en kendt label er der tale om supervised learning, altså opsætning af en genkendelse hvor systemet guides i den ønskede retning. [Brownlee2013] Der findes også unsupervised learning hvor indholdet af data samples ikke har nogen label. Ved unsupervised learning undersøges matematiske sammenhænge og tendenser. [Brownlee2013] I denne rapport fokuseres der fremadrettet kun på supervised learning.

Supervised learning forsøger først og fremmest at finde en model, hvor data afbilledes til labels og det bliver klassificeret. Det er vigtigt, at modellen afbilledes så korrekt som den kan, for både træningssættet og de nye data. En af grundene til at denne proces er vigtigt, er at den så kan generalisere, men samtidig kan blive stærkere og bedre rustet når der kommer ny data ind. Machine learning er i princippet en algoritme, som arbejder ud fra hypoteser og generalisering af data. Desto flere data den har at arbejde med igennem tiden desto bedre bliver træningssættet også, hvilket resulterer i lav risiko for at få forkerte labels ud. På denne måde bliver træningssættet bedre, og algoritmen bliver bedre til at rubicere nye data. [DIKU2010] Hvordan systemet træffer en beslutning kan variere mellem algoritmer, men overordnet kan det visualiseres som et beslutningstræ; et forsimplet, ensrettet flowchart som vist på figur 2.5, hvor hvert punkt er en indsnævring mod den label der passer

bedst på det pågældende data.¹



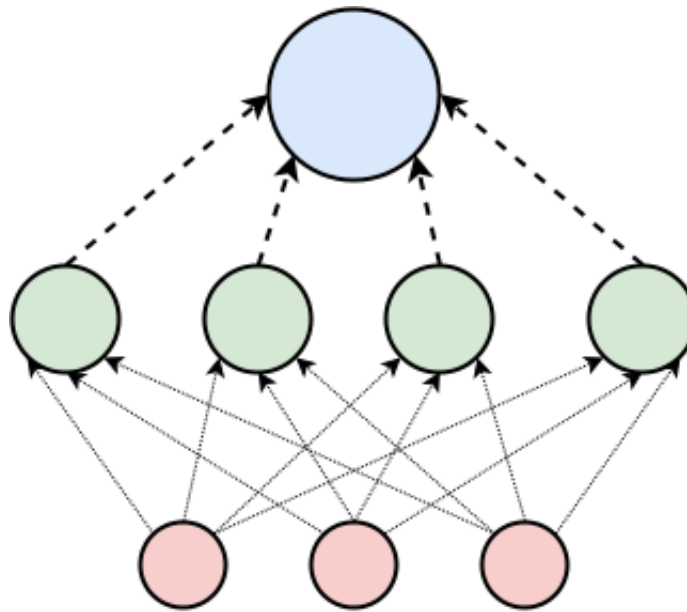
Figur 2.4: Visualisering af machine learning beslutningstræ

Figuren viser en forsimplet visualisering af et vilkårligt beslutningstræ fra en machine learning algoritme.²

Der findes flere metoder at udføre supervised learning på. En af de første, og stadig populære, er neurale netværk. Her tages der udgangspunkt i et netværk af celler, hvor disse celler får input hvor de i sidste ende summerer deres givne input til en neuron, hvilket kan ses på nedenstående figur. (skal lige have figuren ind) Fra at cellen får et input til der fås et output fra neuroner, forgår der en proces mellem de to stadier. Processen her kaldes hidden units, der er egentlig det fundamentale og kan næsten sammenlignes med hjernen. Dataene bliver klassificeret undervejs, af de enkelte celler, hvilket betyder at netværket minimeres for fejl, samt at generaliseringsevnen vil blive stærkere. Det gælder om for netværket, at få opbygget et struktur, så den kan forudsige de aspekter den bliver spurgt ind til. [DIKU2010] En af måderne netværket vil blive stærkere igennem tiden er, at den anvender begrebet feedforward netværk, der betyder at når netværket bliver anvendt bliver den trænet, samt også efter den er blevet anvendt. Dette skyldes blandt andet en mekanisme kaldt backpropagation. Netværket går ind og sammenligner det output den får ud sammenlignet med det der blev ønsket, og herefter kan der ud fra sammenlignings grundlaget gives feedback til de hidden units og output unitsene.³ Et eksempel på et neuralt netværk kan ses på figur ??.

¹FiXme Note: Bayesian Reasoning and Machine Learning. Mendeley virker ikke.. sætter den ind i morgen..

³FiXme Note: Har lagt kilden op i mendeley og hedder david, shai shalev- kunne ikke lige finde det navn man skulle citerer til



Figur 2.5: Neuralt netværk med skjult lag

På billedet ses et neuralt netværk med skjult lag. Datainput sker i de røde neuroner, beregnes i det skjulte grønne lag, hvorefter det outputtes i den blå neuron.

Andre metoder kan også anvendes til at udføre supervised learning. Support Vector Machines (SVM) er en populær algoritme til særligt praktiske problemer. Boosting-algoritmer, som f.eks. AdaBoost er også populære og anvender et princip hvor mange simple klassifikationer samles i en flertalsafstemning, frem for ét beslutningstræ per datapoint.

