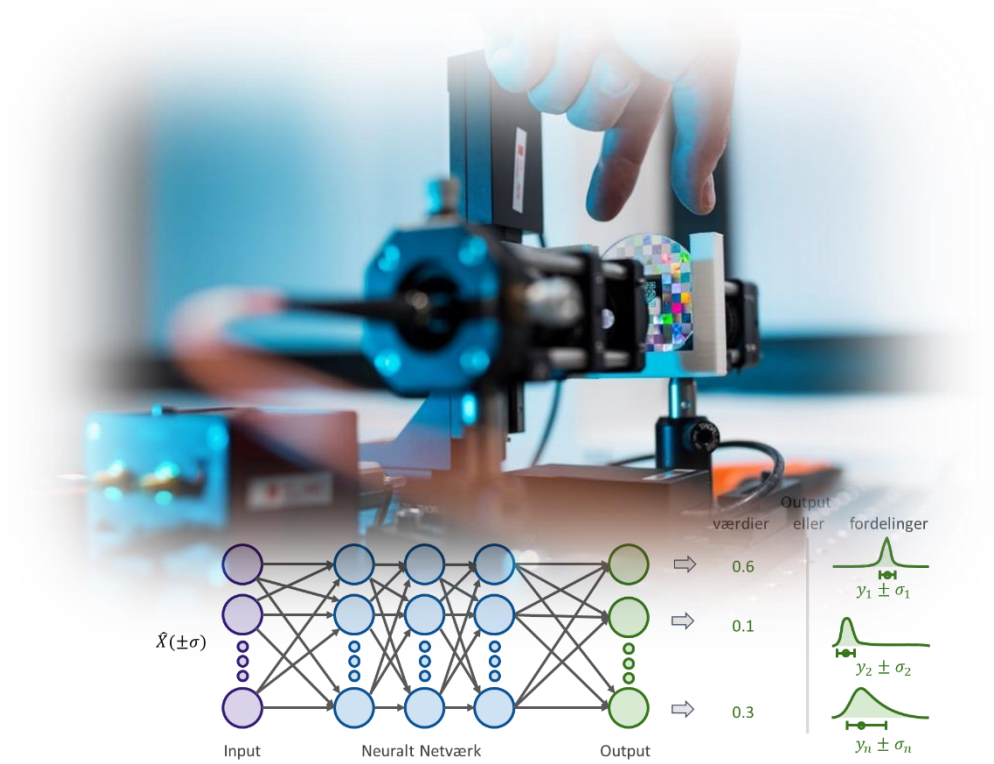
DFM
Danmarks Nationale Metrologiinstitut

Metrologi og Kunstig Intelligens

Modeller med konfidens

Af: Jesper Bjerge Christensen, Mathias Geisler, Anders Brusch og David Balslev-Harder



Dato: 2021-03-22
Ref.: DFM-PUB ID 175
Projekt: 4006-03
Metrologi i kunstig intelligens og digitale tvillinger
Dette arbejde er støttet af Uddannelses- og Forskningsstyrelsen.

DFM A/S - Danmarks Nationale Metrologiinstitut
Kogle Allé 5
2970 Hørsholm
Danmark
+45 7730 5800

Resumé

Denne rapport giver en oversigt over DFM's planlagte fokusområder ifm. GTS-aktiviteten "Metrologi i kunstig intelligens og digitale tvillinger", som er et statsstøttet projekt i samarbejde med FORCE Technology. Foruden en overordnet introduktion til arbejdsprocessen tilknyttet træningen af modeller inden for kunstig intelligens, lægges i denne rapport stor vægt på vigtigheden af, at de store modeller gøres i stand til at tilknytte en usikkerhed til deres forudsigelser. I denne forbindelse defineres begreberne aleatoriske og epistemiske usikkerheder, som beskriver hhv. usikkerheder fra data og model. Desuden gives et indblik i et udsnit af DFM's aktive forskningsaktiviteter, hvori digitale tvillinger benyttes til at lære et neuralt netværk at klassificere og kvantificere nanostrukturer baseret på optisk skatterometri.

Introduktion

Kunstig intelligens (KI) har i løbet af det seneste årti fundet anvendelse inden for et væld af forskellige sektorer, og det bruges i stigende grad som et værktøj til at optimere og automatisere industrielle processer og opgaver. Eksempler på anvendelser inkluderer kvalitetssikring mht. produktionsudsving, "smart factory" og forudsigende vedligeholdelse, "smart farming", modeller til prognoser for epidemier, assisteret patologi, samt billedanalyse og biometri. Der er efterhånden ingen tvivl om at vi kan se ind i en fremtid, hvor KI og maskinlæring-baserede algoritmer vil blive en mere og mere integreret del af menneskets samfund og hverdag.

Det er dog vigtigt at huske, at en KI stadig blot består af algoritmer udført af en computer – algoritmer der (med få undtagelser) hverken kan lave risikovurderinger eller forstå evt. konsekvenser af dårlige valg. I nogle tilfælde kan forkerte beslutninger være katastrofale, som fx i 2016 da en passager i en selvkørende bil omkom, efter bilens autonome system fejlklassificerede en hvid trailer som værende en del af himlen (NHTSA, 2017). Eller i 2020 da en uskyldig mand blev beskyldt for en forbrydelse grundet en fejl begået af et ansigtsgenkendelsessystem (Hill, 2020). På www.incidentdatabase.ai/ findes oversigter over (sandsynligvis kun en meget lille del af) de uheld, der er forekommet grundet beslutninger taget af KI-systemer.

En helt fundamental fejl i rigtig mange KI-modeller er deres manglende evne til at svare: "*Tak fordi du spørger, men jeg ved det ikke*". Som mennesker har vi en helt naturlig evne (kognitiv sans) til at vide, hvornår vi *er i tvivl om noget,* og hvornår vi *ved noget*. KI-modeller har, desværre, ikke helt den samme sans for, hvornår de er sikre, og hvornår de er usikre, og til det helt centrale spørgsmål om, hvorvidt "KI-modeller ved, hvad de ikke ved", er svaret oftest nej. Dette er især tilfældet, når en model pludselig skal tage stilling til et nyt datasæt, hvor der ikke har været lignende fortilfælde. I stedet for at svare noget tilfældigt, ville det være passende, hvis modellen gav et svar, der så var behæftet med en meget stor *usikkerhed*.

Metrologi er læren om målinger af fx længde, masse, optisk effekt og temperatur. De fleste har den opfattelse, at et instruments visning ved en måling er et eksakt resultat, men denne opfattelse overlever kun, fordi der eksisterer et, for de fleste usynligt, netværk af kalibreringer (sammenligninger), som sikrer, at udstyret er indstillet til at måle rigtigt. Nationale metrologiinstitutter (NMI'er) har som primær opgave at sikre adgang til pålidelige målinger, der er globalt accepterede. Dette realiseres gennem internationalt samarbejde mellem NMI'erne, hvor der fortløbende udføres direkte sammenligninger af målemetoder og

normaler for at sikre overensstemmelse. DFM udbreder denne globale sporbarhed til danske aktører via kalibreringer. Til hver kalibrering tildeles *usikkerheder* for måleresultaterne, og bruges dette udstyr igen til at kalibrere i andre sammenhænge, stiger usikkerheden. De metrologiske forskere ved NMI'er er eksperter i at kvantificere usikkerheder samt identificere alle bidragende usikkerhedsfaktorer. For at KI-systemer kan opnå samme tillidsniveau i forhold til måletekniske og produktionsmæssige sammenhænge er de metrologiske betragtninger højst aktuelle.

At KI-modeller oftest ikke både spytter et svar og en forbundet usikkerhed på svaret ud er generelt et stort problem, da det i princippet er umuligt at forholde sig til et tal, som ikke har en usikkerhed. Uden en usikkerhed kan man ikke lave en individuel vurdering af, om man skal stole på modellens output. I stedet bliver en KI-model næsten udelukkende bedømt på metrikker som "accuracy" og "precision" – metrikker som uden diskussion er brugbare til at vurdere modellens overordnede performance, men som intet siger om modellens konfidens i de enkelte tilfælde. Når en models overordnede performance bruges til at tage stilling til det enkelte tilfælde, svarer det i virkeligheden lidt til et spil hasard – og det er, desværre, sådan det er blevet kutyme at bruge KI.

Et af de helt store kritikpunkter, som KI igennem årene er blevet mødt af, er, at modellerne er for sort boks-agtige og ugennemskuelige. Men mon ikke en del af denne kritik ville forstumme, hvis modellernes forudsigelser ikke bare var en numerisk værdi, men i stedet et fuldbyrdet statistisk resultat med både en estimeret værdi og en tilhørende usikkerhed?

Med GTS-aktiviteten 'Metrologi i kunstig intelligens og digitale tvillinger' (se bedreInnovation.dk) arbejder DFM og FORCE Technology på at sammenkoble metrologiske metoder med den seneste udvikling inden for KI, digitale tvillinger og Internet of Things. I de følgende afsnit fremføres hvordan DFM, med metrologiske briller, ser muligheder for at udvide pålideligheden af KI og dermed øge anvendeligheden af KI i kritiske processer.

Innovation af målemetoder vha. kunstig intelligens

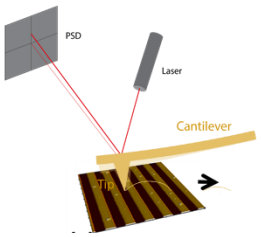
KI er mest kendt for sin anvendelse inden for billed- og stemmegenkendelse/generering, diagnostisk analyse, finansielle analyser og ikke mindst selvkørende biler. Fælles for de kendte metoder er adgang til meget store datasæt, eller muligheden for at systemet kan træne sig selv enten rent digitalt eller i den fysiske verden (fx robotsystemer).

Ved DFM har udgangspunktet været at udvikle nye målemetoder med innovativ KI. Et eksempel på dette er DFM's optiske spektrometersystem, som kan udføre in-line karakterisering af strukturer på nanoskala i almindelige produktionsomgivelser. Konventionelt anvendes fx Atomic Force Mikroskopi (AFM) til måling af nanostrukturer, men AFM er meget følsomt over for vibrationer, er tidskrævende at udføre og beror på stærkt specialiseret arbejdskraft.

Den optiske metode benyttet ved DFM baserer sig på at måle, hvordan nanostrukturerne påvirker spektret fra en hvid lyskilde, og den kan operere selvstændigt. For at opnå dette har DFM opbygget en stor database som kobler gitterstrukturers dimensioner til modsvarende optiske spektre. Som udgangspunkt anvendte DFM disse datasæt som opslagsbibliotek til at bestemme hvilket spektrum i databasen, der passede bedst til et målt spektrum, men denne metode var stærkt begrænset af opslagstiden grundet databasens størrelse.

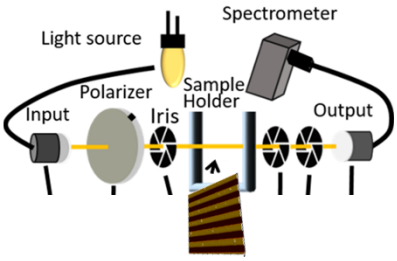
Ved i stedet at bruge KI-metoder har DFM trænet et neuralt netværk, som kan behandle målte optiske spektre fra nanostrukturer og ud fra spektret bestemme størrelsen på strukturernes dimensioner på nanoskala. Med denne metode kan analysen give forbedrede estimater især for størrelser som ikke ligger tæt på elementer i datasættet. En ekstra fordel ved denne metode er, at når KI'en først er trænet, så er dataprocesseringstiden stærkt reduceret, når systemet anvendes i nye målinger. Aktuelt arbejder DFM med metoder som tillader det neurale netværk at tildele usikkerheder på størrelserne af de beregnede nanostrukturer.

Måling af nanostrukturer



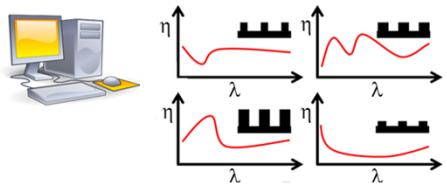
Klassisk bruges avanceret mikroskopi, som tager mange timer og kun måler et lille område.

Optisk skatterometri



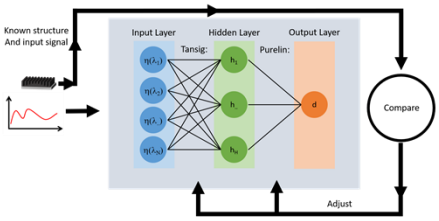
Metode med optisk spektrometer som måler prøven på <1 s og dækker et stort område.

Data: strukturer vs. spektre



Et målt spektrum sammenlignes med et bibliotek af relationer mellem gitterstrukturer og spektre. Det bedste match vælges som resultat.

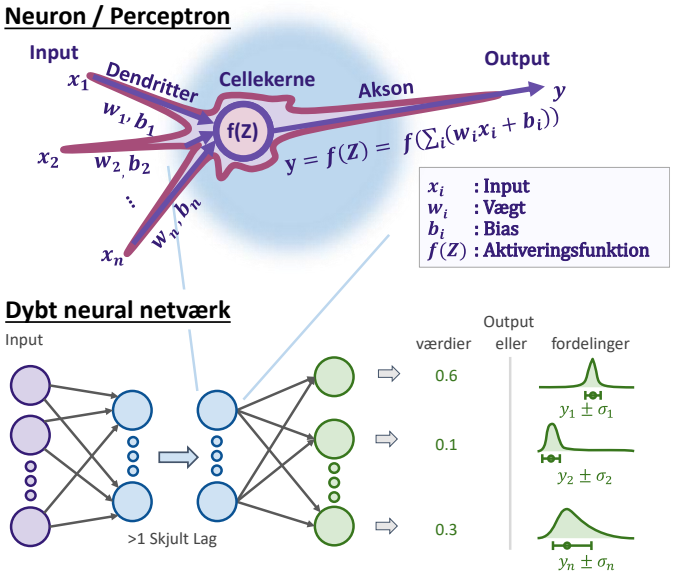
Forbedrede anlyser med KI



Ved at konstruere et neuralt netværk forbedres beregningstiden med mere nuancerede analyseresultater.

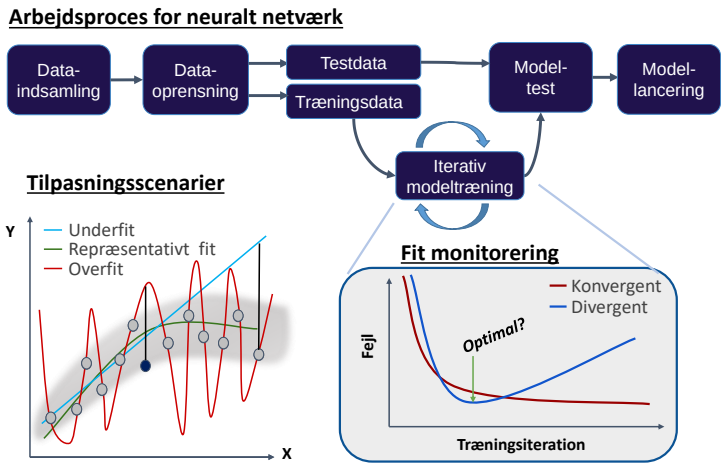
Metrologi i kunstig intelligens

Neurale netværk er opfundet med inspiration fra nervesystemer i biologiens verden. I biologiens verden består nerveceller (forsimplet set) af netop ét akson (transmitteren, Tx), et antal dendritter (modtagere, Rx) og en cellekerne som er med til at bestemme aktiveringen af aksonet (se Figur 1). Nervecellerne repræsenteres i KI ved en model (også kaldet Perceptron), som tager et antal numeriske input $x_1, x_2, \ldots, x_n$ og genererer én outputværdi y. Udgangsværdien bestemmes af aktiveringsfunktionen f, som godt kan være ulineær, og som input typisk har $\sum_i(w_i x_i + b_i)$, hvor parametrene w_i og b_i hhv. er vægte og bias, som justeres under træningen. Et kunstigt neuralt netværk opbygges ved at forbinde store antal af disse kunstige neuroner, typisk designet i lag, således at inputværdier indføres i det første lag og derfra videreføres til efterfølgende lag, hvor et sidste lag slutteligt definerer outputværdierne. Kunstige neurale netværk kan opbygges på et utal af måder og indgår ofte som proprietær del af virksomheders løsning på en opgave.



Figur 1. Kunstige neuroners funktionalitet og deres rolle i opbyggelsen af et neuralt netværk.

En overordnet arbejdsproces for implementering af neurale netværk går dog igen (se Figur 2). Fælles for alle maskinlæringsmodeller er behovet for store datasæt. Datasættet bearbejdes således, at hvert datapunkt har relevant inputdata og et tilsvarende output. Fordi modellen kun lærer fra det tilgængelige data, er korrektheden og kvaliteten af datasættet af stor betydning for netværkets resulterende funktionalitet. Datasættet bearbejdes og opdeles i et træningssæt, et valideringssæt og et testsæt. Træningssættet danner grundlag for at indstille parametrene (træningen) i det valgte neurale netværksdesign (antal lag, antal neuroner i lagene osv.). Valideringssættet anvendes efterfølgende til at kontrollere, hvordan den nuværende konfiguration netværk fungerer mht. datapunkter, som det ikke har set før med henblik på at optimere netværkskonfigurationen. Når en tilfredsstillende konfiguration for det neurale netværk er opnået, anvender man slutteligt testsættet for at evaluere netværkets præstationsevne.



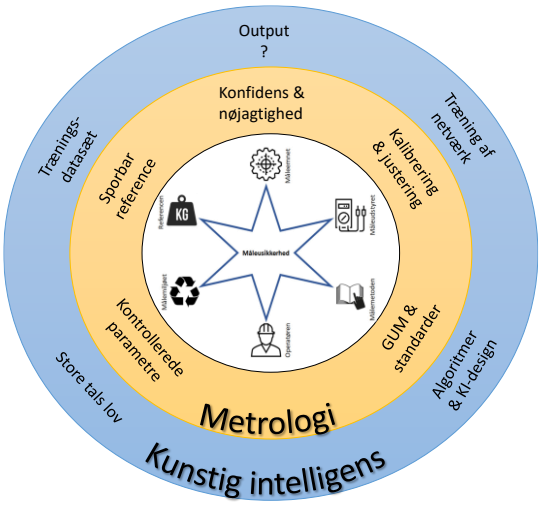
Figur 2. Arbejdsproces for træning af neuralt netværk.

Træningen foregår ved at udføre en regressionsanalyse, der søger at minimere fejlen mellem træningsdatasættets forventede outputdata og de forudsigelser, som det trænede netværk udregner på basis af træningsdatasættets

inputdata. Under træningen kan denne fejl overvåges løbende. Valideringsfejlen kan enten udvikle sig konvergent dvs. blive ved med at nærme sig en mindste værdi, eller den kan divergere og derved ikke finde en løsning. Nogle gange findes et minimum inden divergensen, og dette bliver da ofte betragtet som den bedste løsning, men der er ingen garanti for, om dette er et globalt eller lokalt minimum. Den efterfølgende sammenligning med validerings- og testsættet er essentiel i forhold til at lave denne vurdering, men også i forhold til at vurdere, om den optimerede indstilling af parametre overfitter eller underfitter i forhold til datasættet. Dog mangler denne validering ofte at kontrollere for performance som ligger langt uden for det område, som er defineret af træningssættet. De afgørende faktorer i forhold til en KI-kvalitet ligger således i designet af det neurale netværk, træningen af netværket, samt størrelsen og kvaliteten af det datasæt som anvendes til træningen. Således er adgang til store datasæt allerede blevet en del af børsfondes vurdering af virksomheder, der arbejder med KI. Der foreligger derfor en opgave i at kunne dokumentere kvaliteten af disse forhold på en sammenlignelig måde, og her byder metrologi både på tilsvarende erfaringer fra andre sammenhænge, samt en stor matematisk og metodisk værktøjskasse som er relevant ifm. evalueringen af de implementerede neurale netværk.

Det kunstige neurale netværk kan betragtes som en funktion $\mathcal{F}(\boldsymbol{d}, \boldsymbol{\theta})$, hvor $\boldsymbol{d}$ er et inputdatapunkt og $\boldsymbol{\theta}$ repræsenterer alle netværkets interne parametre. Med metrologiske briller svarer neurale netværk til de modelfunktioner, som anvendes i udregning af usikkerhedsbudgetter ifm. måletekniske kalibreringer. Der er således en række erfaringer fra metrologiske tilgange, som med fordel kan overføres til KI.

Idet neurale netværk kan implementeres på vilkårlige måder er standardisering en meget kompleks opgave, især fordi forskellige designs kan anvendes til samme type opgaver. Figur 3 viser i centrum de væsentligste faktorer, som bidrager til usikkerheden i en måleteknisk kalibrering nemlig målemetoden, måleudstyret, måleemnet, referenceemnet, målemiljøet og operatøren. Alle disse faktorer kan relateres over til KI. Hvor KI beror på store datasæt og store tals lov, anvender metrologi systematisk kontrol af alle relevante parametre og referenceobjekter, som er sporbare til andre laboratorier. Hvor KI er baseret på træning af netværk og sandsynligheder, supplerer metrologi med justering og kalibrering samt konfidensintervaller og nøjagtighed. Hvor KI bygger på design og algoritmer supplerer metrologi med standarder og metoder. Alle disse relationer kan drages i spil for at udvikle metoder til at dokumentere kvaliteten af KI, det være sig både for kvaliteten af de anvendte datasæt i træningen, men også hvordan Guide to Uncertainty of Measurements (GUM, 2008) kan anvendes til at udvide analyser til at inkludere usikkerhedsberegninger. Dette behandles i de følgende to afsnit om aleatoriske (bidrag fra data) og epistemiske (bidrag fra KI-modellen) usikkerheder.



Figur 3. Analogi mellem metrologiens etablerede metoder og metoder som stadig er i udvikling inden for kunstig intelligens.

Aleatoriske usikkerheder

Den aleatoriske usikkerhed er det, vi også kender som statistisk usikkerhed. En sensor, der passivt monitorerer CO_2-koncentrationen i luften, giver et output som vil fluktuere en smule over tid. Disse fluktuationer er ikke nødvendigvis et resultat af ændrede CO_2-koncentrationer, men skyldes typisk intern elektrisk støj i sensoren selv. Dette er et eksempel på aleatorisk usikkerhed – de naturlige variationer som forekommer i repetitioner af forsøg, der er foretaget under samme forhold. Et andet eksempel kan ses i Figur 4, der illustrerer de statistiske variationer af temperaturmålinger foretaget i et oliebad med en fastholdt temperatur.

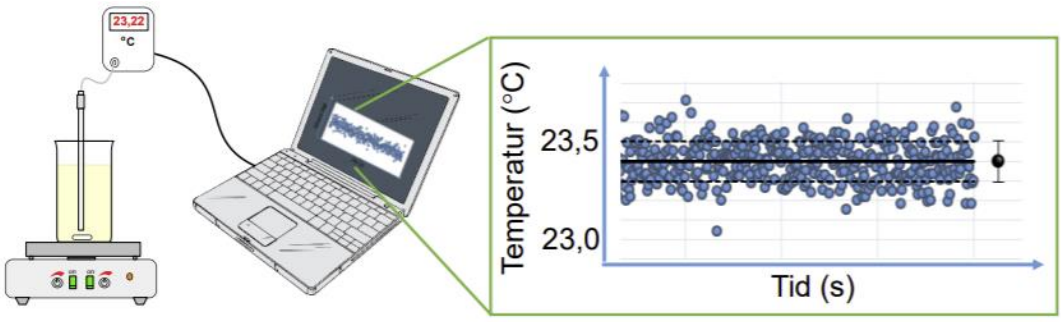
ALEATORISK USIKKERHED

Den aleatoriske usikkerhed er den usikkerhed, som er forbundet med den statistiske variation i data som et resultat af fx elektrisk eller termisk støj. I modeller indenfor kunstig intelligens vil vi typisk referere til de aleatoriske usikkerheder som inputusikkerheder.

Aleatoriske usikkerheder er vigtige at forstå i forbindelse med KI, da de indtræder som usikkerheder på den data, der bruges som input i modellen. Den statistiske støj reduceres ikke med opsamlingen af mere og mere data (i modsætning til den epistemiske usikkerhed, se nedenfor), og fastsætter derfor ofte en øvre grænse for, hvor godt en given KI-model kan klare sig i sidste ende.

Usikkerhed på inputparametrene er vigtige at kunne kvantificere i modellens træningsfase. I træningsfasen af en klassificeringsalgoritme er det kutyme at bruge dataaugmentering til at skabe (pseudo)nye træningsdatasæt. Usikkerhederne for modellens inputparameterrum bør danne rammen for, hvordan denne dataaugmentering foregår, men i praksis foregår det i dag på en meget løst defineret måde. Derudover er kontradiktatorisk (eng. adversarial) træning (Goodfellow, 2015) blevet et populært værktøj til at gøre sin model mindre følsom over for afvigende eller ekstreme punkter i inputdata, men det er endnu uklart, hvordan implementeringen af nye og mere raffinerede træningsformer bør udføres i praksis for at tage korrekt højde for inputusikkerheden.

Den aleatoriske usikkerhed er en egenskab af data, men den kan være afgørende at forstå for at kunne optimere sin model. Det er derfor interessant, at man, ved at lave en lille ændring i sin træningsprocedure, kan lære, hvordan usikkerhederne fra ens input propagerer igennem netværket og bliver til usikkerheder på ens forudsigelser (Kendall, 2017). Udvikling og standardisering af principperne omkring usikkerhedspropagering gennem neurale netværk spiller en helt central rolle i indeværende projekt.



Figur 4. Et elektronisk termometer benyttes til at aflæse temperaturen fra et oliebad, der har en konstant temperatur. Som et resultat af de statistiske måleusikkerheder fra instrumentet, fluktuerer aflæsningerne omkring en værdi på 23,4 °C med en standardafvigelse på 0,1 °C. Figuren er adapteret fra (Balslev-Harder, 2017).

Epistemiske usikkerheder

Den epistemiske usikkerhed er den usikkerhed, som vi direkte forbinder med vores model – et mere sigende navn er derfor modelusikkerhed. Inden for KI findes to primære bidrag til modelusikkerheden: (1) *Usikkerhed grundet modellens principielle begrænsninger* og (2) *usikkerhed i modellens parametre grundet en endelig datamængde.* Den første slags modelusikkerhed kan fx eksemplificeres ved, at man med en ekstremt præcis vægt forsøger at finde massen af et blylod uden at korrigere for luftens opdrift på loddet; resultatet vil aldrig kunne blive helt nøjagtigt, fordi den anvendte model er ufuldstændig. Inden for KI er denne slags usikkerhed relateret til modellens arkitektur og kan tænkes på som værende relateret til modellens absolutte potentiale. Den anden slags modelusikkerhed er statistisk – når man påbegynder træningen af et neuralt netværk er usikkerheden på hver enkelt neuronforbindelse stor, men i takt med, at modellen ser mere og mere data, bliver den mere og mere sikker på dens "valg" af modelparametre.

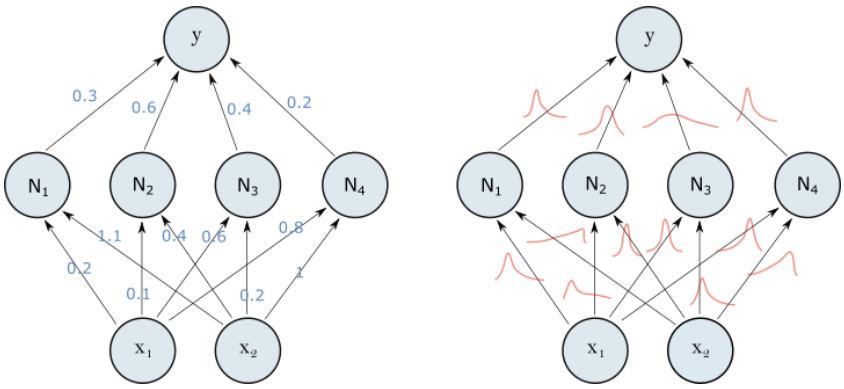
EPISTEMISK USIKKERHED

Den epistemiske usikkerhed (ikke at forvirre med den epidemiske usikkerhed vi lever i) er den usikkerhed, som er forbundet med vores model enten som resultat af, at modellen er begrænset eller grundet manglende observationer til at opbygge den.

Modelusikkerhed beskriver altså vores uvidenhed i forhold til at vælge den helt rigtige model. I takt med at modellen trænes, bliver modellen mindre og mindre usikker på de enkelte modelparametres værdier. Denne fortolkning synes at foreslå, at der er noget fundamentalt forkert med måden neurale netværk ofte trænes på i dag – for burde vi i virkeligheden ikke hellere træne *fordelinger* end enkelte numeriske vægte, som skitseret i Figur 5? Med sandsynlighedsfordelinger for modellens vægte kan modellens output y samtidig blive behæftet med en standardusikkerhed $u(y)$.

I de seneste 5-10 år har der været et stigende fokus på at implementere usikkerhedsvurderinger i KI. Der er især sket fremskridt inden for metoder baseret på *bayesiske neurale netværk*. Et af eksemplerne på vigtige bidrag findes i (Blundell, 2015), hvori der beskrives en algoritme (Bayes by Backprop) til at træne både gennemsnitsværdier og standardafvigelser på samtlige vægte i netværket.

En stor fordel ved at benytte bayesisk-baseret KI er måden, hvorpå disse kan håndtere data, som ikke er indeholdt i træningsparameterrummet. Her er de bayesiske neurale netværk nemlig gode til at behæfte en stor usikkerhed med det endelige svar. Mere om dette kan findes, ved forespørgsel, i (Christensen, 2021).

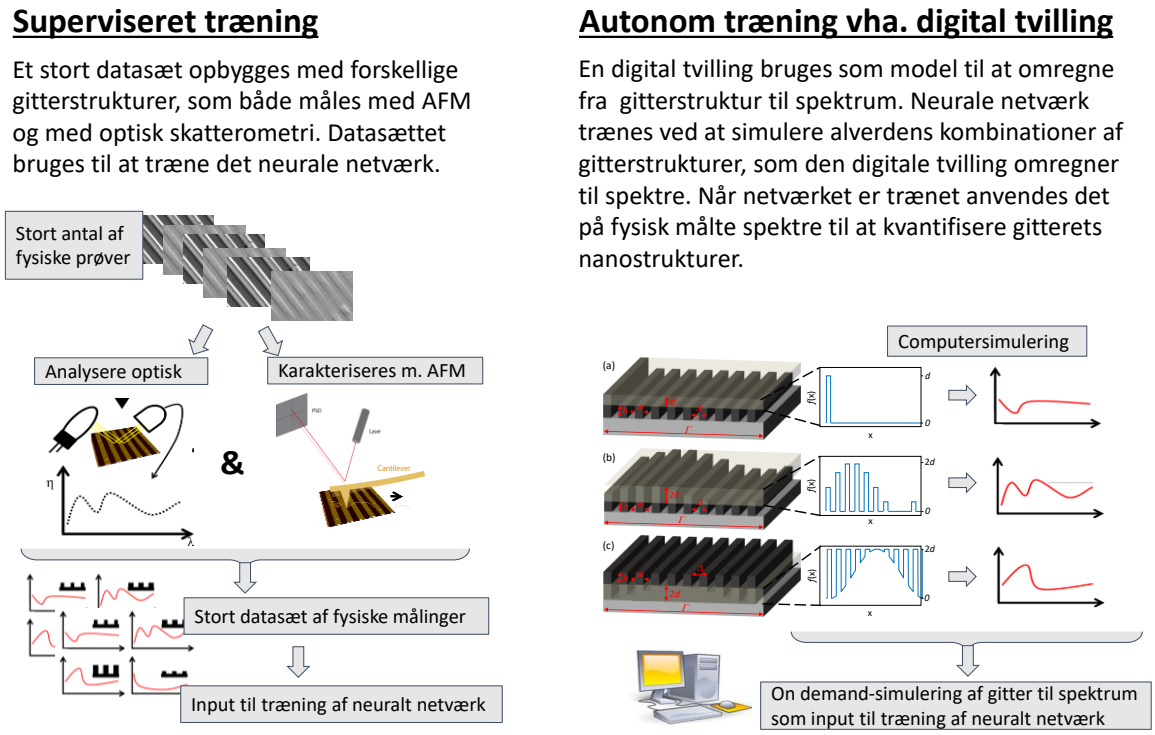


Figur 5. Forskellen mellem et klassisk neuralt netværk via baglænspropagering (t.v.) og et netværk trænet via "Bayes by Backprop" (t.h.). Inspireret af (Blundell, 2015).

Superviseret vs. autonom træning

Ikke alle ting kan beskrives med en analytisk matematisk funktion, og nogle gange kan et system være så komplekst, at den underliggende model er utiltalende at skulle stille op. Det er i disse situationer, KI har et kæmpe potentiale som både små og store virksomheder i stigende grad er begyndt at få øjnene op for. Overordnet inddeles KI-modeller i to kategorier: Klassificeringsmodeller (fx identifikation af objekter på et billede) og regressionsmodeller (som giver numeriske estimater, fx fremskrivning af tidsserier). Yderligere findes modelkategorier som fx foldningsnetværk (convolutional), graf-foldningsnetværk, og tilbageførte (recurrent) netværk. Træningen af netværket inddeles også i kategorier: superviseret træning og autonom træning. Superviseret træning beror på træning med datasæt, der matcher inputparametre med outputparametre som beskrevet tidligere. Ved autonom træning, bruger træningsalgoritmen sine egne erfaringer til at optimere netværket. Et berømt eksempel er AlphaGo, hvor KI'en blev sat til at spille brætspillet Go mod sig selv og på den måde trænede autonomt. Go er for komplekst til en udtømmende udregning af alle mulige træk, og AlphaGo lærte på egen hånd med et avanceret neuralt netværk en måde at finde de bedste træk på. Algoritmens selvudviklede taktikker var så effektive og samtidig ukendte for menneskelige spillere, at de nu i dag kigger AlphaGo over skulderen for at aflure dens strategi.

Inden for måleteknologi kan autonom træning indgå ved, at træningen af det neurale netværk udføres med en digital tvilling. Pointen ved denne fremgangsmåde er, at det trænede netværk kan være beregningsmæssigt hurtigere i anvendelsesøjeblikket, samt at modelleringskompleksiteten ofte er asymmetrisk, dvs. det er nemmere at simulere i en retning end i den anden. Som eksempel på dette kunne være DFM's skatterometri-case (se Figur 6), hvor det ud fra en simuleret nanostruktur er muligt at modellere, hvordan det optiske spektrum vil se ud, hvorimod den modsatte beregning (fra spektrum til



Figur 6. Fremgangsmetode ved hhv. superviseret træning med store datasæt og autonom træning vha. digitale tvillinger.

struktur) er et langt sværere problem. Udfordringen ved at anvende digitale tvillinger af fysiske systemer er, at det er nødvendigt tage højde for alskens tænkelige og utænkelige fysiske parametre, som kan indvirke på resultatet samt usikkerhedsbidrag fra disse. Her byder metrologien igen på relevante metoder til at kvalitetssikre digitale tvillinger og vurdere usikkerhedsbidrag i de autonomt trænede netværk.

Perspektivering

I løbet af det sidste halve årti har der været et stærkt øget (akademisk) fokus på inklusionen, og forståelsen, af usikkerheder i forudsigelser lavet af modeller bygget på KI. Denne interesse er opstået, fordi en række forskere opdagede, at deres modeller gav meget sikre, men forkerte forudsigelser på eksempler, hvor modellerne ikke burde kende svaret grundet manglende træning. Nogle forskere præsenterede endda eksempler, hvor deres modeller gav meget sikre forudsigelser for inputdata bestående af ren støj. Disse tendenser medfører helt naturligt, at man må stille spørgsmålstegn ved modellernes troværdighed – især eftersom modellerne typisk ikke umiddelbart giver den menneskelige bruger mulighed for at forstå grundlaget for dens beregninger pga. netværkets kompleksitet.

Med usikkerhedsvurderinger på modellernes output er det muligt at få en forståelse af, hvornår modellernes output er sikre, og hvornår det modsatte er tilfældet. I forbindelse med den øgede interesse for inkorporering af usikkerheder i KI er der efterhånden udviklet en række forskellige metoder, som varierer en hel del i både kompleksitet og håndgribelighed. Nogle af disse metoder giver endda information om, hvorvidt en given model er begrænset primært af aleatoriske eller epistemiske usikkerhedsbidrag – information som belejligt kan bruges som et centralt værktøj til at forbedre modellen.

Med de mange nye metoder til usikkerhedsvurderinger inden for KI er det nødvendigt, at der opbygges standarder, som kortlægger metodernes anvendelsesområde og begrænsninger. Standardiseringsarbejde indenfor KI har allerede været i gang i flere år (ISO/IEC JTC 1/SC 42) med fokus på bl.a. gennemsigtighed og troværdighed. Der ligger i dette standardiseringsarbejde en enorm udfordring, i og med at KI har et uoverskueligt bredt anvendelsespotentiale.

I fremtiden vil KI være en integreret del af vores hverdag med anvendelser inden for alt fra forslag til ugens indkøb over procesoptimering til tidlig diagnosticering af alvorlige sygdomme. Ikke desto mindre vigtigt er det derfor, at de udviklede algoritmer i stigende grad tilegnes en forståelse for hvad de ved – og især hvad de ikke ved. Her er usikkerheder på modellernes forudsigelser en nøglekomponent til at kvalificere modellernes forudsigelser og dermed højne anvendeligheden og udbyttet af det potentiale, der ligger i neurale netværk.

Bibliografi

Balslev-Harder, D. (2017). *A2 – Introduktion til usikkerhedsbudgetter.* metrologi.dk.

Blundell, C. (2015). Weight Uncertainty in Neural Networks. *32nd International Conference on Machine Learning*.

Christensen, J. (2021). *Metrologi og kunstig intelligens - Intern DFM rapport.*

Goodfellow, I. (2015). Explaining and Harnessing Adversarial Examples. *3rd International Conference on Learning Representations*.

GUM. (2008). *JCGM 100:2008 Evaluation of measurement data - Guide to the expression of uncertainty in measurement.* BIPM.

Hill, K. (2020). *Wrongfully Accused by an Algorithm.* New York Times.

Kendall, A. (2017). What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *31st Conference on Neural Information Processing Systems*.

NHTSA. (2017). *Tesla Crash Preliminary Evaluation Report, PE 16-007.* U. S. Department of Transportation.