

Hur maskininlärning möjliggör bildigenkänning

Ken Erikson

Kandidatavhandling i datateknik

Handledare: Hannu T. Toivonen

Åbo Akademi

Fakulteten för naturvetenskaper och teknik

Våren 2018

Referat

Maskininlärning är ett område som har utvecklats kraftigt och fått mycket uppmärksamhet under de senaste åren. Den fulla potentialen för vad maskininlärning kan uppnå är ändå ännu väldigt oklar. Nya framsteg görs hela tiden och endast framtiden kan utvisa var gränsen går för vad tekniken kan uppnå, eller om en sådan gräns ens existerar.

I den här kandidatavhandlingen tas först maskininlärningens historia kort upp. Maskininlärning har en lång historia och utvecklingen har skett i varierande takt främst beroende av hur hög prioritet marknads stora aktörer har satt på området. Vidare går diskussionen in på centrala begrepp inom området.

Fokus ligger sedan på djup maskininlärning och bildigenkänning. Dessa diskuteras både teoretiskt och i praktiken genom att enskilt presentera alla ingående steg, samt genom att ta upp ett praktiskt exempel på bildigenkänning.

Nyckelord

Maskininlärning, bildigenkänning, teckenigenkänning, ocr, mönsterigenkänning, djup maskininlärning

1. Inledning	4
2. Bakgrund	6
2.1. Maskininlärningens historia	6
2.2. Bildigenkännings historia	8
3. Maskininlärning med fokus på bildigenkänning	9
3.1. Mönsterigenkänning	10
3.2. Användningsområden för mönsterigenkänning	10
3.3. Förbehandling av data	11
3.4. Klassificering av data	12
3.5. Djup maskininlärning	13
3.6. Neurala nätverk	13
3.7. Typer av maskininlärningsalgoritmer	14
3.7.1. Övervakat lärande	15
3.7.2. Oövervakat lärande	15
3.7.3. Förstärkt lärande	15
4. Bildigenkänning med TensorFlow	17
4.1. Vad är TensorFlow?	17
4.2. Tillvägagångssätt	17
4.3. Softmaxregression	18
4.4. Exempel på indata	19
4.5. Kontroll av resultatet	20
5. Diskussion och sammanfattning	24
Källor	25

1. Inledning

Maskininlärning är ett område som utvecklats kraftigt under de senaste årtiondena och påverkar vardagen för så gott som varje människa, och eftersom maskininlärning ofta fungerar bakom ett användargränssnitt så är personen sällan medveten om hur stor inverkan maskininlärning har. Syftet med denna avhandling är att göra tydligt de fördelar som maskininlärning och bildigenkänning har jämfört med andra tillvägagångssätt. Detta görs genom att se på hur maskininlärning har utvecklats till vad det är i dag, samt vilka användningsområden som tekniken lämpar sig för och hur tekniken fungerar.

Maskininlärning diskuteras i kapitel 2 först ur ett historiskt perspektiv. Det som behandlas mest ingående är vad artificiell intelligens har betytt för maskininlärning och vilka framsteg som har lett till den teknik vi i dag har tillgänglig. Därefter diskuteras vanliga användningsområden för maskininlärning och nya användningar av djup maskininlärning.

Centrala begrepp inom maskininlärning tas i kapitlet 3 upp på ett teoretiskt plan och förklaras för att ge en behövlig inblick i alla de steg som ingår i bildigenkänning. Mönsterigenkänning och förbehandling av data är väsentliga inom alla typer av maskininlärning, och är därmed även viktiga vid bildigenkänning. Djup maskininlärning är även det vitalt för bildigenkänning och har varit en viktig del i det senaste uppsvinget av intresset för maskininlärning.

Ett praktiskt exempel på bildigenkänning behandlas i kapitel 4 där Googles egna maskininlärningsramverk, TensorFlow, används för att visa hur maskininlärning kan användas i praktiken. Fokus läggs på att poängtera vilka styrkor och svagheter maskininlärning har i sin nuvarande form.

I kapitel 5 diskuteras slutligen de punkter som tagits upp och maskininlärningens nuvarande läge, samt hur framtiden eventuellt kan utvecklas inom området.

2. Bakgrund

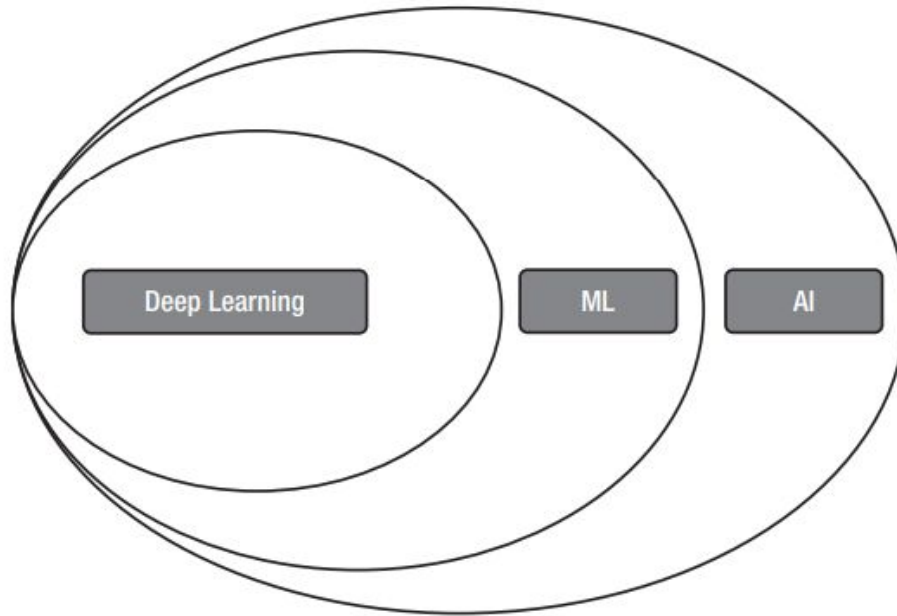
2.1. Maskininlärningens historia

Alan Turing ställde år 1950 frågan: "Kan maskiner tänka?", som blev den centrala frågan inom artificiell intelligens och sedan maskininlärning [1]. Artificiell intelligens, förkortat AI, började på riktigt undersökas efter slutet av andra världskriget och artificiell intelligens som begrepp myntades 1956 [2]. AI har sedan dess använts för att uppnå betydande resultat inom olika fält genom att t.ex. vinna över schackmästare, köra bilar på allmänna vägar och diagnostisera sjukdomar [2]. AI är ett fält som är relevant för vilken intelligent uppgift som helst [2].

För att utveckla AI som kan agera och resonera likt en människa är det viktigt att systemet kan representera och spara den kunskap som den har tillgängligt och, genom automatiserad resonering, använda sig av denna kunskap för att svara på frågor och dra slutsatser [2]. AI kan med hjälp av allt detta sköta många uppgifter, förutsatt att förhållandena kring uppgiften är statiska. För uppgifter där förhållanden ändras och systemet måste anpassa sig kan man använda sig av maskininlärning, förkortat ML. Maskininlärning är ett delområde inom artificiell intelligens, vilket kan ses i figur 1. Maskininlärning fungerar genom att anpassa sig till nya situationer och genom att detektera mönster ur sammanhang [2].

Maskininlärning fick sin riktiga början ur artificiella neurala nätverk, förkortat ANN [3]. I mitten på 1980-talet steg intresset för ANN explosionsartat eftersom framsteg inom parallelliserade processorer krävde ett sätt att distribuera alla beräkningar på, och ANN ansågs vara ett möjligt svar [3]. ANN är inspirerade av den mänskliga hjärnan, hur den fungerar då den lär sig och hur olika lager av neuroner samarbetar med varandra [3].

De framsteg som skedde under 1980-talet inom artificiella neurala nätverk och därmed också maskininlärning ledde till att maskiner redan under sent 80-tal och tidigt 90-tal kunde känna igen talade ord, köra bilar på allmänna vägar, spela spel som backgammon på elitnivå och mycket annat.



Figur 1. Forskningsområden inom artificiell intelligens [1].

Intresset för neurala nätverk stärktes igen under 2000-talets början och djup maskininlärning har sedan dess varit ett väldigt aktivt forskningsområde [4]. Djup maskininlärning (eng. Deep Learning) är ett delområde inom maskininlärning, vilket kan ses i figur 1. Djup maskininlärning är det område inom AI och maskininlärning som banar vägen för modern maskininlärning och nya framsteg sker hela tiden eftersom stora företag som Google, Facebook och Microsoft satsar mer och mer på sina interna djupinlärningsteam [4].

Matthew Kirk tar i sin bok *Thoughtful machine learning with Python* upp att den renässans som djup maskininlärning haft under de senaste 10 åren har lett till att maskininlärningssystem har börjat kunna använda sig av data som tidigare inte varit

kvantifierbara [5]. Maskininlärning som inte har sin grund i data klassificerat av människor har ökat möjliga användningsområden samt minskat människans del i maskininlärningsprocessen [5].

2.2. Bildigenkännings historia

Bild- och teckenigenkänning är ett område som i stor omfattning är beroende av maskininlärning för att förbättra noggrannheten och automationen vid igenkänning och klassificering av bilder och tecken.

George Nagy påstår i sin artikel *At the Frontiers of OCR* att mycket av framstegen inom teckenigenkänningsområdet från 1955 till 1992 berodde på bättre processor- och digitaliseringsteknik och inte på förbättrade klassificeringstekniker [6]. Nagy ansåg att större framsteg skulle komma från att se på den typografiska enhetligheten hos paragrafer och andra layoutkomponenter [6]. För att kunna använda sig av dessa data effektivt och automatiskt måste självlärande teknik såsom maskininlärning tillämpas.

Förutom teckenigenkänning har maskininlärning även redan under 90-talet använts inom bildigenkänning för att möjliggöra både automatisk ansiktsigenkänning [7] och igenkänning av möjliga cancerceller [8]. Maskininlärning har sedan även spelat en signifikant roll inom självkörande bilar genom att möjliggöra behagligare körning, snabbare styrning och effektivare bensinförbrukning [3]. Bildigenkänning har använts inom samma område för att bilarna skall kunna känna igen fotgängare, cyklister, trafikmärken, o.s.v. [3]. Alpaypin redogör i sin bok *Machine Learning: The New AI* för hur maskininlärning och bildigenkänning i framtiden även kommer att kunna användas inom utforskning av rymden [3].

3. Maskininlärning med fokus på bildigenkänning

Ett datorprogram sägs lära sig från erfarenhet E med avseende på en viss klass av uppgifter T och en prestandamätning P , om dess prestanda vid uppgifter i T , mätt med P , förbättras med erfarenhet E .

Denna definition på maskininlärning, tagen ur Tom M. Mitchells bok *Machine Learning* [7], utgör grunden för maskininlärning och alla delområden bygger vidare på denna definition. Maskininlärning fungerar genom att anpassa sig till nya situationer och genom att upptäcka och extrapolera mönster ur sammanhang [2]. Detta kapitel kommer vidare att ta upp vad som ligger i bakgrunden för hur maskininlärning kan hitta och använda sig av dessa mönster.

Att analysera bilder är att hitta mönster och dra slutsatser från givna data och hur dessa data samspelar med varandra. Theodoridis och Koutroumbas presenterar i sin bok *Pattern recognition* [9] att mönsterigenkänning är en av utgångspunkterna inom bildigenkänning, men även inom andra områden, som t.ex. röstigenkänning, datautvinning och informationssökning. Maskininlärning bygger vidare på idén om hur människor lagrar den bästa strategin för en viss situation i sina hjärnor och sedan tar fram denna information då samma situation uppenbarar sig igen [3]. Att få maskiner att lära sig av tidigare mönster för att kunna lösa problem, som tidigare var möjligt endast för människor att lösa, skulle kunna effektivisera många användningsområden.

3.1. Mönsterigenkänning

Mönsterigenkänning är klassificeringen av objekt in i olika klasser [1]. Mönsterigenkänning har, enligt Theodoridis och Koutroumbas, börjat användas i praktiken mer och mer ända sedan 1960-talet och är i dag i absolut framkant av tekniska tillämpningar och forskning [1]. Alpaydin säger i sin bok *Machine learning - the new AI* [3] att huvudidén bakom maskininlärning är hur klassificeringen kan härledas från tidigare observationer. Mönsterigenkänning är nyckeln till att förstå dessa tidigare observationer och fatta statistiskt tillförlitliga beslut, vilket gör att mönsterigenkänning är ytterst viktig i alla typer av maskininlärning.

3.2. Användningsområden för mönsterigenkänning

Mönsterigenkänning i sig är ingen ny uppfinning. Mönsterigenkänning har tidigare i historien använts för att kunna förutse jordbävningar [10], beskriva planeternas rörelse [11] och för att diagnostisera patienter [3]. Skillnaden mellan förr och nu är att förr togs informationen direkt från tidigare erfarenheter och nedskrivna berättelser, medan informationen nu kommer från den ofantliga och ständigt växande mängd maskinlagrade data som finns tillgänglig. Nu på senare år har dessa data använts för att hitta kreditbedrägerier [7], bokrekommendationer [7] och förutse ångest och depression hos åldrande patienter [6].

Bildbehandling är ett område där mönsterigenkänning är av yttersta vikt [1]. Ett bildbehandlingssystem är ett system som tar en bild och ger ut en beskrivning av bilden [1]. Ett typiskt användningsområde för ett sådant system är visuell granskning av produkter på ett löpande band [1]. Användning av sådana system har en stor positiv inverkan på den totala produktionen jämfört med om en människa skulle

granska varje enskild produkt. Om bara systemet har tillgång till tillräckligt med data och feedback så kan även kvaliteten säkras. I vissa fall kan ett sådant system ge till och med högre kvalitet än en mänsklig motsvarighet.

3.3. Förbehandling av data

De data man har tillgänglig utgör grunden för vad man kan åstadkomma med maskininlärning. En aspekt som är viktig för att effektivt kunna använda alla tillgängliga data är att dessa är i rätt fördefinierat format. För bildigenkänning är det till exempel viktigt att alla bilder är i samma storlek och bildförhållande. Enligt det som Ethem Alpaydin presenterar i sin bok *Machine Learning: The New AI* så förbereds alla data oftast redan i ett separat förbehandlingsskede. Till exempel vid optisk teckenigenkänning normaliseras givna tecken till en fixad storlek utan att information går förlorad [2]. Efter att alla data har konverterats till rätt format så kan de användas för att lära en maskin att klassificera nya data genom att känna igen mönster.

Beroende av de indata som undersöks och det sökta användningsområdet måste givna indata förbehandlas i olika stor utsträckning. Enligt det som Netzer et al. presenterar i *Reading Digits in Natural Images with Unsupervised Feature Learning* är det ett viktigt problem att inte endast kunna identifiera siffror ur färdigt skalade bilder, utan även att kunna hitta siffror i bilder där siffrorna både varierar i storlek och ändrar positionering på bilden [12]. Författarna förstärker detta genom att ta fram att om ett system hittar den korrekta placeringen på siffrorna med endast 90 % tillförlitlighet, så kommer det slutliga resultatet för systemet omöjligen att uppgå till mera än 90 %, fastän teckenigenkänningen och segmenteringen skulle fungera perfekt [12].

Segmenteringen är ett viktigt slutsteg efter att en siffra eller annat teckens position i en bild har hittats av systemet. Vid segmentering av teckensegment, som kan bestå av ett eller flera tecken, måste systemet hitta möjliga lodräta gränser mellan olika tecken [12]. De enskilda tecknen kan sedan inom sina givna gränser klassificeras separat [12].

3.4. Klassificering av data

Maskininlärning fungerar genom att klassificera data in i olika klasser. Klasserna som data kan klassificeras som är väldigt olika från fall till fall. Vid tillämpning av textigenkänning kan varje tecken klassificeras som en bokstav i alfabetet medan vid tillämpning av ansiktsigenkänning så kan bilder klassificeras enligt t.ex. kön, ålder eller namn.

Beroende av de sökta klasserna är det viktigt att identifiera vilka attribut hos alla data som är viktiga och vilka som inte är relevanta. I sin bok *Machine Learning: The New AI* redogör Ethem Alpaydin för hur de attribut som är relevanta varierar kraftigt med vilka klasser programmet försöker sortera in alla indata i. Han tar som exempel upp att vid textigenkänning så ignoreras t.ex. storleken, stilen och författarens egenheter om det sökta resultatet är vilket tecken det är fråga om, medan dessa samma attribut är bland de viktigaste vid signatur- och handstilsanalys. Alpaydin påpekar även att vid ansiktsigenkänning så kan attribut som känslor och miner vara irrelevant om de sökta klasserna är identiteten på personen, men högst relevanta om de sökta klasserna är vilken känslöstämning personen befinner sig i [2]. Detta leder till att ett system som tillämpar maskininlärning måste lära sig vilka attribut som är viktiga i systemets användningsområde.

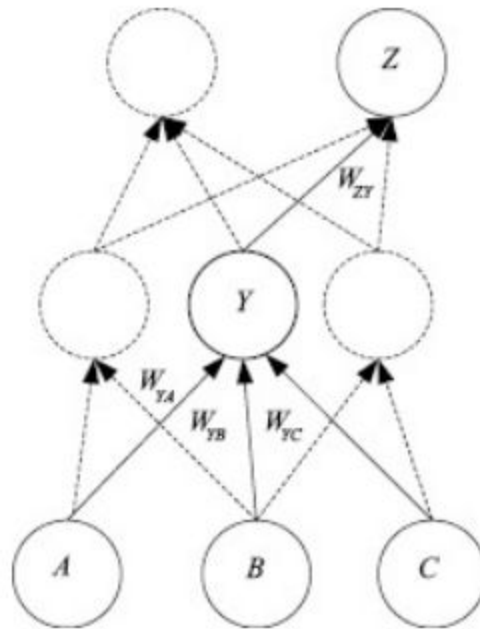
3.5. Djup maskininlärning

Djup maskininlärning är ett delområde inom maskininlärning [4]. Medan vanliga maskininlärningssystem endast lär sig genom att utgå från givna regler så lär sig djupa maskininlärningssystem utöver detta även genom att använda sig av givna mångdimensionella data och alla interpretationer av mönster som dessa data medför [4].

Djup maskininlärning är attraktivt i dagens läge mycket på grund av att information, utan mänsklig inblandning, kan fås ur de ofantliga datamängder som finns tillgängliga [3]. Vissa komplexa områden, som t.ex. datorsyn och naturlig språkbehandling, har tidigare uppvisat bristfälliga resultat vid försök att införa automation. Med hjälp av djup maskininlärning har automation åstadkommit inom dessa områden med en precision som kan konkurrera med, och i vissa fall till och med överträffa, den noggrannhet som uppnås av människor [4].

3.6. Neurala nätverk

De komplexa problem som människor har ställts inför har gjort att nya tillvägagångssätt undersökts och ny teknik och kunskap från andra forskningsområden har möjliggjort metoder som tidigare inte varit genomförbara. Senaste forskningen inom maskininlärning har fokuserat på att skapa modeller som påminner om hur våra hjärnor fungerar [4].



Figur 2. Exempel på ett neuralt nätverk uppbyggt av neuroner och synaptiska samband mellan dessa [3].

Grunden för hur vår hjärna fungerar är neuroner och de synaptiska samband som existerar mellan dessa [4]. En neuron är förenklat optimerad för att ta emot information från en neuron, behandla informationen på ett unikt sätt och sedan skicka iväg den behandlade informationen till en annan neuron [4]. Sättet som maskininlärning försöker efterlikna detta på kan ses i figur 2, där indata A, B och C ger ett visst värde Y i neuronerna i det mellersta lagret för att sedan ge ett värde Z som utdata. Vikterna som sätts på sambanden mellan noderna påverkar hur mycket sambandet beaktas i slutresultatet, t.ex. W_{ya} anger hur stor påverkan A har på värdet Y [4]. Dessa vikter justeras beroende av hur långt slutresultatet är från det sökta resultatet då modellen tränas för att sedan kunna ge trovärdiga resultat för okända data [4].

3.7. Typer av maskininlärningsalgoritmer

Det finns olika typer av maskininlärningsalgoritmer. I denna avhandling behandlas huvudtyperna som maskininlärningsalgoritmer delas in i, nämligen övervakat lärande, oövervakat lärande och förstärkande lärande.

3.7.1. Övervakat lärande

Vid övervakat lärande har de data som maskininlärningsalgoritmen har till förfogande förutom utgångsinformation även önskade mål eller resultat. Dessa data kallas märkta data och målet blir att hitta en generell regel som kartlägger indata till det önskade resultatet. Den nya regeln kan sedan användas för att märka nya omärkta data [13].

Märkt data är ofta tidskrävande att producera på grund av att människor ofta behöver delta eller åtminstone kontrollera automatisk klassificering. Om möjligt kan klassificeringen av data till viss del skötas av allmänheten. Övervakat lärande används flitigt inom t.ex. ansikts- och taligenkänning, filmrekommendationer och försäljningsprognoser [13].

3.7.2. Oövervakat lärande

Vid oövervakat lärande använder sig maskininlärningsalgoritmer av omärkta data, d.v.s. att ingen information existerar utöver utgångsinformationen och data har inte beskrivits på något sätt. Det är algoritmen som strukturerar givna indata och ser vilka gömda mönster som existerar för att sedan kunna märka ovanligheter. Oövervakat lärande kan användas för att lägga märke till bedrägeri och andra felaktigheter eller för att gruppera ihop kunder med liknande inköpsvanor [13].

3.7.3. Förstärkt lärande

Förstärkt lärande ger återkoppling till systemet så att det kan anpassa sig till dynamiska omständigheter och ge ut ett visst resultat. Systemet utvärderar sin egen prestation och eftersträvar att optimera den. Förstärkt lärande har använts flitigt bland annat inom självkörande bilar och för att möjliggöra schackmästaren AlphaGo [13].

4. Bildigenkänning med TensorFlow

Bildigenkänning och maskininlärning är i dagens läge inte endast för stora företag som har tillgång till den nyaste tekniken och mest kraftfulla hårdvaran. Även en enskild person kan använda sig av fritt tillgänglig mjukvara, som kan köras på ett brett allt från mobiltelefoner till storskaliga distribuerade system [8]. Ett praktiskt exempel på bildigenkänning är att ta ett färdigt klassificerat dataset och lokalt lära ett system att känna igen tecken baserat på detta. För att enkelt förevisa de enkla steg som ingår i ett sådant exempel används här Googles egna TensorFlow, som är ett fritt tillgängligt maskininlärningssystem med öppen källkod [14].

4.1. Vad är TensorFlow?

TensorFlow [15] är Googles fritt tillgängliga maskininlärningssystem som enligt Martín Abadi m.fl. fungerar både som ett gränssnitt för att uttrycka maskininlärningsalgoritmer och som en implementation för att exekvera sådana algoritmer [16]. TensorFlow är baserat på Googles DistBelief som är ett maskininlärningssystem som Google internt tidigare har använt sig av i stor utsträckning inom maskininlärningsforskning men även i ett brett utbud av erbjudna tjänster [16].

4.2. Tillvägagångssätt

För att förevisa bildigenkänning och teckenigenkänning i praktiken tas här upp ett praktiskt exempel. Som färdigt klassificerade data används MNIST (Modified National Institute of Standards and Technology) fritt tillgängliga databas med totalt

70 000 klassificerade handskrivna siffror i formatet 28x28 pixlar [17]. Li Deng framhåller i en artikel i *IEEE Signal Processing Magazine* att MNIST-databasen har blivit en standard för att snabbt och effektivt kunna jämföra olika maskininlärningsalgoritmer på en gemensam grund och med ett rättvist utgångsläge [17].

Som grund för exemplet används Googles och TensorFlows officiella handledningsmaterial *MNIST For ML Beginners* [18]. TensorFlow är gjort för att vara flexibelt och enkelt kunna uttrycka en mängd olika algoritmer [16]. Detta gör att ett relativt komplicerat problem så som teckenigenkänning kan lösas med godtycklig noggrannhet med endast ett tiotal rader kod, varav en stor del av raderna bara fungerar som initialisering av diverse inställningar och variabler [18].

TensorFlow-systemet tar in 100 slumpmässigt utvalda testdata ur MNIST-databasen och lär sig baserat på dessa hur handskrivna siffror borde se ut. Denna inlärningsprocess upprepas sedan 1 000 gånger för att få ett acceptabelt resultat. Inlärningsmetoden som TensorFlow-systemet använder sig av i detta exempel är softmaxregression.

4.3. Softmaxregression

Det är näst intill omöjligt att med fullständig tillförlitlighet bestämma vilken siffra en given bild representerar [4]. Buduma och Locascio framhäver därför i boken *Fundamentals of deep learning: Designing next-generation machine Intelligence algorithms* att det vid igenkänning av handskrivna siffror är en bättre lösning att

lägga fram resultatet enligt formeln nedan, där $\sum_{i=0}^9 p_i = 1$ [4]:

$$[p_0 \ p_1 \ p_2 \ p_3 \ \dots \ p_9].$$

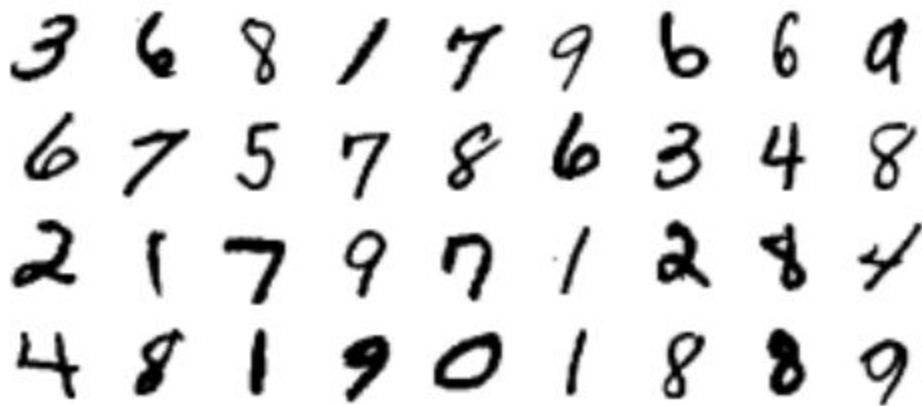
Denna vektor ovan är i praktiken det sista lagret, som i softmaxregression kallas softmaxlagret. Skillnaden mellan detta lager och alla övriga lager är att detta lager beror på övriga noder i lagret eftersom summan av alla noder bör bli 1 [4]. De slutgiltiga värdena i vektorn räknas ut enligt formeln nedan, där $z_i = \log\left(\frac{p_i}{1-p_i}\right)$ (p_i ur vektorn ovan) [4]:

$$y_i = \frac{e^{z_i}}{\sum_j e^{z_j}} .$$

Vektorns värden anger sedan vad systemet anser sannolikheten vara att den givna bilden representerar varje siffra mellan 0 och 9. Om värdet för en enda siffra ligger nära 1 ger systemet en stark gissning, och om värdet på gissningen är lika för flera siffror så är systemet osäkert på vilken siffra den givna bilden representerar [4].

4.4. Exempel på indata

För att få en inblick i hur MNIST-databasens bilder ser ut finns det i figur 3 ett utdrag ur databasen. Totalt testar systemet 70 000 färdigt klassificerade handskrivna siffror [17]. Dessa indata är normaliserade till 28x28 pixlars bilder i gråskala [19].



Figur 3. Exempel på siffror ur MNIST-databasen [19].

Systemet lär sig på basen av dessa indata och lär sig hur den känner igen siffror. För att inte påverka utseendet på de egna handskrivna siffrorna analyserade skribenten inte MNIST-databasens siffror på förhand.

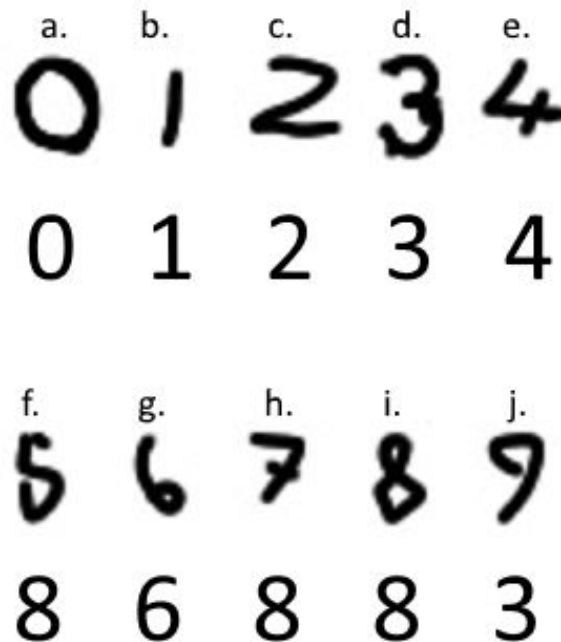
4.5. Kontroll av resultatet

TensorFlow-systemet lär sig på nytt varje körning och eftersom det finns en del slumpmässiga faktorer vid valet av inlärningsdata är resultatet olika vid varje nytt försök. Systemet testar sig självt efter varje inläring genom att köra ett test på en delmängd av MNIST-databasen som inte användes vid inläringen för att inte påverka resultatet. Systemet ger sedan ut en procent som anger hur stor del av dessa testdata som klassificerades korrekt med hjälp av det nyligen upplärda systemet. Systemet gav vid den körning som här behandlas att 92.09 % av MNIST-databasens testdata klassificerades korrekt.

För att testa om systemet även kunde klassificera egenproducerad indata så producerade skribenten två olika uppsättningar av 10 bilder för hand. Bilderna

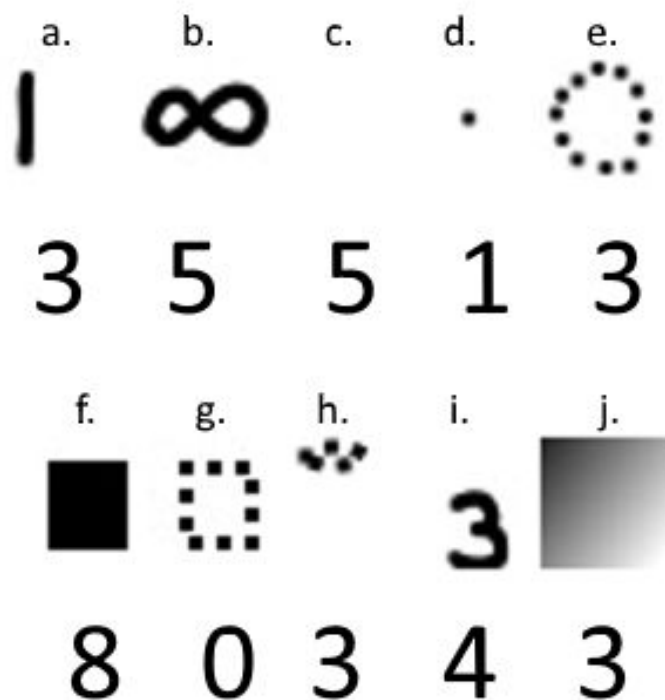
normaliserades till samma format som MNIST-databasens bilder. Den första uppsättningen är ett försök att producera normala siffror, medan den andra uppsättningen är för att se hur systemet klassificerar bilder som är roterade, skalade och även ett antal som helt saknar siffror.

Den första uppsättningen kan ses i figur 4, där systemets klassificering kan existera under varje handskriven siffra. 7 av 10 bilder är korrekt klassificerade, vilket inte riktigt är samma som de 92.09 % som systemet hade på MNIST-databasen under samma körning. Denna skillnad beror troligtvis främst på skillnader i stil och att en provstorlek på 10 inte kan ge tillförlitliga resultat. Stilskillnaden kommer fram tydligt om man jämför 5(f.) och 7(h.) ur figur 4 med samma siffror i figur 3.



Figur 4. Siffror producerade av skribenten och klassificerade av TensorFlow-exemplet.

Den andra uppsättningen kan ses i figur 5 och varje bild har en klassificering från systemet under sig. Första observationen som bör noteras är att TensorFlow-systemet klassificerar även bilder som klart inte innehåller en siffra. Detta påvisar att systemet inte vet vad en siffra är utan bara ger ut den siffra som bilden liknar mest, baserat på vad systemet har lärt sig på givna indata. Siffror som har en annan skalning eller rotation, som t.ex. i figur 5 a., b. och i., påvisar att förbehandling av bilder är viktigt för att systemet skall kunna klassificera bilder rätt. Om bilderna inte är i samma format som givna indata är det inte möjligt för systemet att trovärdigt klassificera bilderna.



Figur 5. Siffror och figurer producerade av skribenten och klassificerade av TensorFlow-exemplet.

Systemet är ett relativt enkelt system och förevisar en del av de möjligheter och begränsningar som maskininlärning för med sig. Ett system är endast så bra som de data det har studerat och hur indatan har anpassats. Om systemet inte har fått den vetskap det behöver för att veta att en tom bild, som i figur 5 c., inte ska klassificeras som en siffra, så kan systemet bara ge ut den siffra som verkar passa bäst baserat på den information det har tillgängligt.

5. Diskussion och sammanfattning

Maskininlärning är ett intressant forskningsområde just nu eftersom utveckling sker konstant och nya genombrott har skett så gott som varje år. Den fördel som maskininlärning har jämfört med andra tillämpningar är att det är så flexibelt och mångsidigt att det kan implementeras inom många olika områden. Eftersom maskininlärning ännu är ett relativt ungt område så existerar ännu många begränsningar, men dessa undersöks och minskar varje dag.

Bildigenkänning är ett viktigt område för framtiden eftersom mängden bilder hela tiden ökar och klassificering blir allt mera väsentligt. Inom allt från självkörande fordon och övervakning till ansikts- och teckenigenkänning är maskininlärning den enda lösning hittills som har visat lovande resultat även i praktiken.

Maskininlärning och bildigenkänning kommer i framtiden att bli allt viktigare i vardagen inom allt flera användningsområden. I och med att datamängden ökar, algoritmer optimeras och nya metoder tas fram kommer gränsen för vad maskininlärning är kapabel till att gång på gång överträffas. Var den slutgiltiga gränsen går återstår att se.

Källor

- [1] R. Kumar, *Machine Learning and Cognition in Enterprises: Business Intelligence Transformed*. Apress, 2017.
- [2] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach, Global Edition*. 2016.
- [3] E. Alpaydin, *Machine Learning: The New AI*. MIT Press, 2016.
- [4] N. Buduma and N. Locascio, *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*. “O’Reilly Media, Inc.,” 2017.
- [5] M. Kirk, *Thoughtful Machine Learning with Python: A Test-Driven Approach*. “O’Reilly Media, Inc.,” 2017.
- [6] G. Nagy, “At the frontiers of OCR,” *Proc. IEEE*, vol. 80, no. 7, pp. 1093–1100, 1992.
- [7] T. M. Mitchell, *Machine Learning*. 1997.
- [8] L. G. Koss, E. Lin, K. Schreiber, P. Elgert, and L. Mango, “Evaluation of the PAPNET™ Cytologic Screening System for Quality Control of Cervical Smears,” *Am. J. Clin. Pathol.*, vol. 101, no. 2, pp. 220–229, 1994.
- [9] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Academic Press, 2008.
- [10] “Luc Devroye, Laszlo Gyorfi and Gabor Lugosi, a probabilistic theory of pattern recognition (Springer, New York, 1996) 636 pages,” *Discrete Appl. Math.*, vol. 73, no. 2, pp. 192–194, 1997.
- [11] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer Verlag, 2006.
- [12] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, “Reading Digits in Natural Images with Unsupervised Feature Learning,” www-cs.stanford.edu. [Online]. Available: https://www-cs.stanford.edu/people/twangcat/papers/nips2011_housen umbers.pdf. [Accessed: 27-Mar-2018].
- [13] Y. (hayden) Liu, *Python Machine Learning By Example*. Packt Publishing Ltd, 2017.
- [14] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, Google Brain, “TensorFlow: A System for Large-Scale Machine Learning,” in *OSDI Vol. 16*, 2016, pp. 265–283.
- [15] *Programvara och dokumentation finns tillgänglig på tensorflow.org*. [Online]. Available: <https://www.tensorflow.org/>. [Accessed: 09-Mar-2018].
- [16] M. Abadi *et al.*, “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems,” 14-Mar-2016.

- [17] “The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web] - IEEE Journals & Magazine.” [Online]. Available: <http://ieeexplore.ieee.org/abstract/document/6296535/>. [Accessed: 13-Mar-2018].
- [18] “MNIST For ML Beginners | TensorFlow,” *TensorFlow*. [Online]. Available: https://www.tensorflow.org/versions/r1.1/get_started/mnist/beginners. [Accessed: 13-Mar-2018].
- [19] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.