Risk bounds for deep learning

Bos, J.M.

Citation

Bos, J. M. (2024, June 19). *Risk bounds for deep learning*. Retrieved from https://hdl.handle.net/1887/3763887

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3763887

Note: To cite this publication please use the final published version (if applicable).

Samenvatting

In dit proefschrift wordt diep leren (in het Engels: deep learning) bestudeerd vanuit een statistisch perspectief. Bovengrenzen worden bewezen voor het risico in het slechtste geval van neurale netwerk schatters in de classificatie, kansdichtheidsschatting en lineaire regressie modellen. Neurale netwerken hebben in de praktijk veel belovende resultaten behaald voor hoog-dimensionale input problemen. Speciale aandacht wordt daarom gegeven aan de rol van de dimensie.

In hoofdstuk 1 worden non-parametrische statistiek en diep leren geïntroduceerd. Hoofdstuk 2 behandelt het probleem van het schatten van de voorwaardelijke categorielidmaatschapskansen in het classificatie model. Hiervoor wordt de kruisentropie verliesfunctie (in het Engels: Cross-Entropy loss) gebruikt. Deze verliesfunctie kan worden afgeleid van de aannemelijkheidsfunctie van de conditionele categorielidmaatschapskansen. Deze verliesfunctie kan oneindig groot worden dichtbij nul. Extra voorzichtigheid is daarom geboden voor conditionele categorielidmaatschapskansen dichtbij nul. We introduceren een afgeknotte versie van de verliesfunctie om de onbegrensdheid in de buurt van nul te verhelpen. Convergentiesnelheden voor het risico gebaseerd op deze afgeknotte verliesfunctie van een neuraal netwerk schatter worden bewezen. Deze snelheden hangen af van een nieuw criterium, de kleine waarde grens (in het Engels: small value bound). Dit criterium bepaalt hoe snel de conditionele categorielidmaatschapskansen naar nul mogen gaan. Het door ons gebruikte afgeknotte risico begrenst het risico gebaseerd op de Hellinger verliesfunctie van boven. Uit deze verwantschap volgt dat de bewezen convergentiesnelheden optimaal zijn als de conditionele categorielidmaatschapskansen wegbegrensd zijn van nul.

Hoofdstuk 3 gaat over kansdichtheidsschatting, een leertaak zonder voorbeeld/leraar (in het Engels: unsupervised learning). Een tweestapsmethode voor dit probleem wordt geïntroduceerd. Eerst wordt het probleem getransformeerd in een regressie probleem: een leertaak met voorbeeld/leraar (in het Engels: supervised learning). Een kernel kansdichtheidsschatter (in het Engels: kernel density estimator) met een te kleine brandbreedte wordt geconstrueerd op basis van de helft van de data. Deze

schatter wordt vervolgens gebruikt om uitkomstvariabelen te genereren voor de andere helft van de data. In de tweede stap van de methode wordt een diep neuraal netwerk getraind op de data die gegenereerd is in de eerste stap. Standaard theorie voor onafhankelijke identiek verdeelde (in het Engels: independent identically distributed, i.i.d.) variabelen is niet meteen toepasbaar. De uitkomstvariabelen zijn namelijk onderling afhankelijk van elkaar. Met behulp van een argumentatie gebaseerd op de Poisson-verdeling wordt een orakelongelijkheid bewezen. Convergentiesnelheden voor deze tweestapsmethode voor het kwadratische risico worden bewezen met behulp van bestaande benaderingsresultaten voor diepe neurale netwerken. Deze snelheden tonen aan dat als de kansdichtheid een samengestelde structuur heeft, dan is de tweestapsmethode in staat om hiervan gebruik te maken om een hogere convergentiesnelheid te bereiken. Een verkennende simulatiestudie bestudeerd de tweestapsmethode voor eindige data hoeveelheden. De simulatiestudie gebruikt hiervoor gestructureerde multivariate kansdichtheden uit de Bayesiaanse netwerk en copula modellen.

Voorwaartse gradiënt afdaling (in het Engels: Forward gradient descent) wordt bestudeerd in hoofdstuk 4. Dit is een biologisch gemotiveerd alternatief voor gradiënt afdaling. Deze methode bevat extra ruis ten opzichte van gradiënt afdaling. Het kan gezien worden als een tussenstap tussen gradiënt afdaling en afgeleide vrije nulde-order methoden (in het Engels: zero-order methods). Convergentiesnelheden voor het kwadratenrisico van deze methode worden bewezen in het lineaire regressie model met gerandomiseerd ontwerp. Deze snelheden zijn een dimensie afhankelijke factor $d\log(d)$ trager dan de convergentiesnelheden die bereikt worden door gradiënt afdaling. Echter, de bewezen snelheden zijn gelijk aan de convergentiesnelheden behaalt door nulde-order methoden.