



Universiteit
Leiden
The Netherlands

Scalability and uncertainty of Gaussian processes

Hadji, M.A.

Citation

Hadji, M. A. (2023, January 25). *Scalability and uncertainty of Gaussian processes*. Retrieved from <https://hdl.handle.net/1887/3513272>

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/1887/3513272>

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

Samenvatting

Dit proefschrift onderzoekt de frequentistische eigenschappen van Bayesiaanse procedures, met name het gebruik van Gauss-proces als *a-priori*-verdeling in Bayesiaanse niet-parametrische statistiek. In Bayesiaanse statistiek is er typisch geen vaste unieke parameter van het model, maar eerder een realisatie van een stochastische variabele die als parameter zal fungeren. Daartoe wordt een *a-priori*-verdeling op de parameters aangenomen. Na waarneming van de data leidt dit tot een *a-posteriori*-verdeling op de parameters, welke gebruikt wordt om tot een schatting van de parameters te komen. Hoewel dit paradigma aanzienlijk verschilt van de frequentistische statistiek, is het toch interessant om te zien hoe de *a-posteriori*-verdeling zich gedraagt als een willekeurige maat die afhangt van de "ware" parameter.

Een zeer aantrekkelijk voordeel van het Bayesiaanse framework is dat het gemakkelijk ingebouwde kwantificering van onzekerheid biedt. Aangezien het mogelijk is een steekproef te nemen uit de *a-posteriori*-verdeling, is de constructie van geloofwaardigheidsgebieden die een fractie van de *a-posteriori*-verdelingsmassa bevatten relatief eenvoudig. In hoofdstuk 2 bestuderen we de dekking van die geloofwaardigheidsgebied die het resultaat zijn van Gauss-proces als *a-priori*-verdelingen met gekwadraterde exponentiële covariance kernel. Omdat de steekproefpaden van het proces oneindig glad zijn, bestaat de gangbare praktijk erin de schaal te wijzigen om de onderliggende functionele parameter terug te vinden. De optimale schaling hangt af van de gladheid van deze parameter, die meestal onbekend is. De hyperparameter voor de schaling wordt dus meestal uit de gegevens geleerd met behulp van hierarchical Bayes of Empirical Bayes technieken. Helaas leiden beide methoden aanvankelijk tot overmoedige, onbetrouwbare onzekerheidsuitspraken voor een grote klasse parameters in de context van het Gaussian white noise model. Door echter de geloofwaardigheidsgebieden op te blazen met een logaritmische factor of de geschatte hyperparameter aan te passen met een logaritmische term kan een goede frequentistische dekking worden verkregen met behoud van een redelijke adaptieve grootte.

De volgende hoofdstukken richten zich op de schaalbaarheid van Bayesiaanse methoden in de context van niet-parametrische regressie op basis van Gauss-procesen. Vanuit het Bayesiaanse paradigma maakt Gauss-procesregressie het mogelijk probabilistische uitspraken te doen over de regressiefunctie op basis van de gegevens. Wanneer de ruis in het model Gaussisch is, is het bovendien mogelijk gebruik te maken van conjugatie om een relatief eenvoudig formule voor de *a-posteriori*-verdeling van de regressiefunctie. Dit model is echter zeer inhalig, want de rekencomplexiteit ervan schaalte kubiek met het aantal waarnemingen. Gedistribueerde methoden maken het mogelijk de gegevens te verdelen over verschillende machines die allemaal een lokale Bayesiaanse niet-parametrische regressie uitvoeren. De lokale oplossingen worden dan door een globale machine verzameld en samengevoegd tot een globale verdeling van de regressiefunctie.

De naïeve aanpak bestaat erin gewoon een niet-parametrische Gaussische regressie

uit te voeren met een willekeurige subset van de gegevens in elke machine en het gemiddelde van alle lokale verdelingen te nemen als globale verdeling. Deze aanpak toont al snel zijn frequentistische grenzen, doordat de convergentiesnelheid van de *a-posteriori*-verdeling afhangt van het aantal machines. Andere methoden zijn mogelijk: de *a-priori*-verdeling lokaal omlaag schalen en dan de resultaten middelen, of de aannemelijkheid omhoog schalen en bijvoorbeeld het Wasserstein-barycentrum van de resulterende lokale verdelingen vinden. Indien de gladheidsparameter van de prior van het Gaussische proces overeenkomt met de werkelijke regelmaat van de regressiefunctie, dan leiden beide methoden tot een bijna optimaal terugvinding en een goede kwantificering van de onzekerheid, mits het aantal machines niet te snel toeneemt ten opzichte van het aantal waarnemingen.

Een andere aanpak zou zijn om de ontwerpruimte van de regressie zodanig te partitioneren dat elke machine een regressie uitvoert op een van niet-overlappende resulterende subgebieden. Hoewel de uiteindelijke *a-posteriori*-verdeling discontinuïteiten bevat aan de grenzen van elke partitie, zal die optimaal samentrekken tot de "ware" regressiefunctie wanneer de herschalingshyperparameter van de prior goed gekozen is. Bovendien kan deze aanpak ook leiden tot adaptieve optimale contractiesnelheden. Zelfs wanneer de "ware" regelmaat onbekend is, is het mogelijk de hyperparameter te leren met behulp van een hiërarchisch framework. Ook dit leidt tot een optimaal herstel van de regressiefunctie.

Deze aanpak kan worden gezien als een aggregatie van de proefsteken van de *a-posteriori*-verdeling verkregen door de verschillende machines in combinatie gewichtsfuncties. De gewichtsfuncties zouden immers indicatorfuncties zijn van een subregio van de ontwerpruimte. De discontinuïteiten van deze laatste zijn dan het gevolg van de discontinuïteiten van de gewichten. Uit een grondige simulatiestudie blijkt dat het mogelijk is om door het kiezen van passende, gegevensgestuurde gewichten een adaptief bijna-optimaal herstel en dekking van de onderliggende regressiefunctie te bereiken.