



Universiteit
Leiden
The Netherlands

Better predictions when models are wrong or underspecified

Ommen, M. van

Citation

Ommen, M. van. (2015, June 10). *Better predictions when models are wrong or underspecified*. Retrieved from <https://hdl.handle.net/1887/33204>

Version: Corrected 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/1887/33204>

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

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/33204> holds various files of this Leiden University dissertation

Author: Ommen, Thijs van

Title: Better predictions when models are wrong or underspecified

Issue Date: 2015-06-10

Stellingen

1. Effectieve modelselectie met als doel het voorspellen van y -waarden gegeven nieuwe x -waarden in regressie- en classificatieproblemen vereist enige kennis van die nieuwe x -waarden. (*Hfst. 2*)
2. Bayesiaanse methoden voor modelselectie en voorspelling kunnen zeer onbetrouwbare resultaten opleveren als de gebruikte modellen verkeerd zijn en de voorspellingen worden geëvalueerd met een verliesfunctie anders dan logaritmisch verlies. (*Hfst. 3–5*)
3. Met het SafeBayesiaanse algoritme kan de hierboven beschreven situatie herkend en gecorrigeerd worden. (*Hfst. 3–5*)
4. Bij een voorspeltaak zoals in het Monty Hall-probleem (waar een kansverdeling op een uitkomst bijgewerkt moet worden in het licht van nieuw bewijsmateriaal, maar waar het precieze verband tussen uitkomst en bewijsmateriaal onbekend is), dient bij een worst-case benadering rekening gehouden te worden met de ruimte van mogelijk te observeren waarden en met de keuze van verliesfunctie. (*Hfst. 6*)
5. Als in de hierboven beschreven voorspeltaak de ruimte van mogelijke observaties aan bepaalde structureisen voldoet, is de worst-case optimale voorspelling niet afhankelijk van de gebruikte verliesfunctie, en kan deze kansverdeling efficiënt berekend worden. (*Hfst. 7–8*)
6. Voor het verkrijgen van inzicht in het gedrag van een nieuwe statistische methode zijn theoretische analyse en experimenten op gesimuleerde data van grotere waarde dan experimenten op data uit de praktijk.
7. Ook statistische methoden die al lange tijd in de praktijk worden toegepast, kunnen nog steeds ruimte bieden voor verbetering.
8. Alle verliesfuncties zijn verkeerd, maar sommige zijn logaritmisch. (*Vrij naar G.E.P. Box*)
9. Het doen van goede voorspellingen en het vinden van de ware verdeling zijn twee verschillende, soms tegengestelde doeleinden.
10. ‘Er is geen God’ is een simplificerende aanname die het voorspellend vermogen van veel wetenschappelijke modellen ten goede komt. Dit biedt echter geen wetenschappelijke grond voor de conclusie dat deze aanname waar is.
11. Een zin zegt meer dan duizend woorden.
12. Alles heeft met elkaar te maken; de vraag is *hoe*.