Machine Learning and Bayesian modeling of rare diseases

Gabriel I Penagos – Londoño
Department of Economics, Pontificia Universidad Javeriana, Bogotá, Colombia.

Introduction

A rare disease affects 1 person in 2000[1], that small number disincentivize pharmaceuticals to develop treatments, *orphan medicines*.

Patients experience a long delay until a diagnosis is reached, spending resources in a process which is referred to as the "**diagnostic odyssey**"

The small population size **impairs large trial designs**, and the test power of standard statistics methods might not be sufficient for detecting a treatment effect.

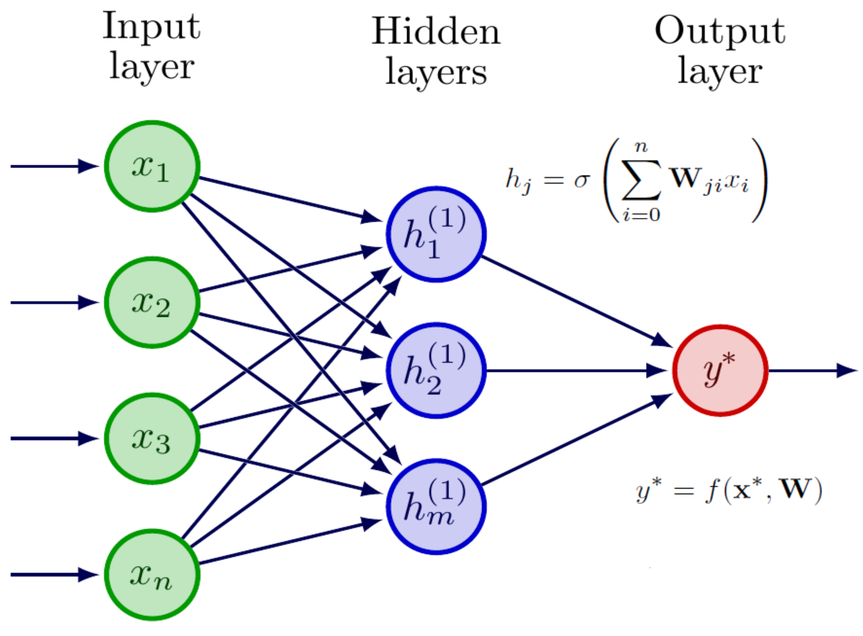
Even though the conditions are infrequent, they affect more than 36 million people in European countries imposing a **high disease burden**.

ANN

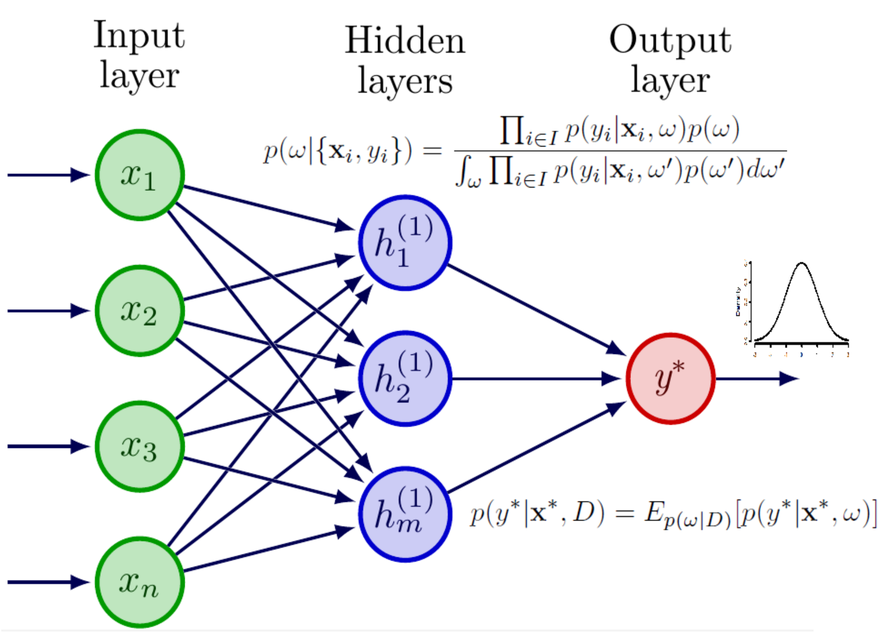
An **artificial neural network (ANN)** learns patterns and relationships from the data[2].

In theory, with hundreds of weights an ANN can approximate **any continuous function**, however this flexibility comes at a price, they would exhibit overfitting.

Overfitting results of a model fitting the training data, including random fluctuations, and therefore poorly performing on new data.



Methods – BNN



Bayesian neural networks (BNN), estimated with **Markov Chain Monte Carlo (MCMC)**, along with **evolutionary feature selection**.

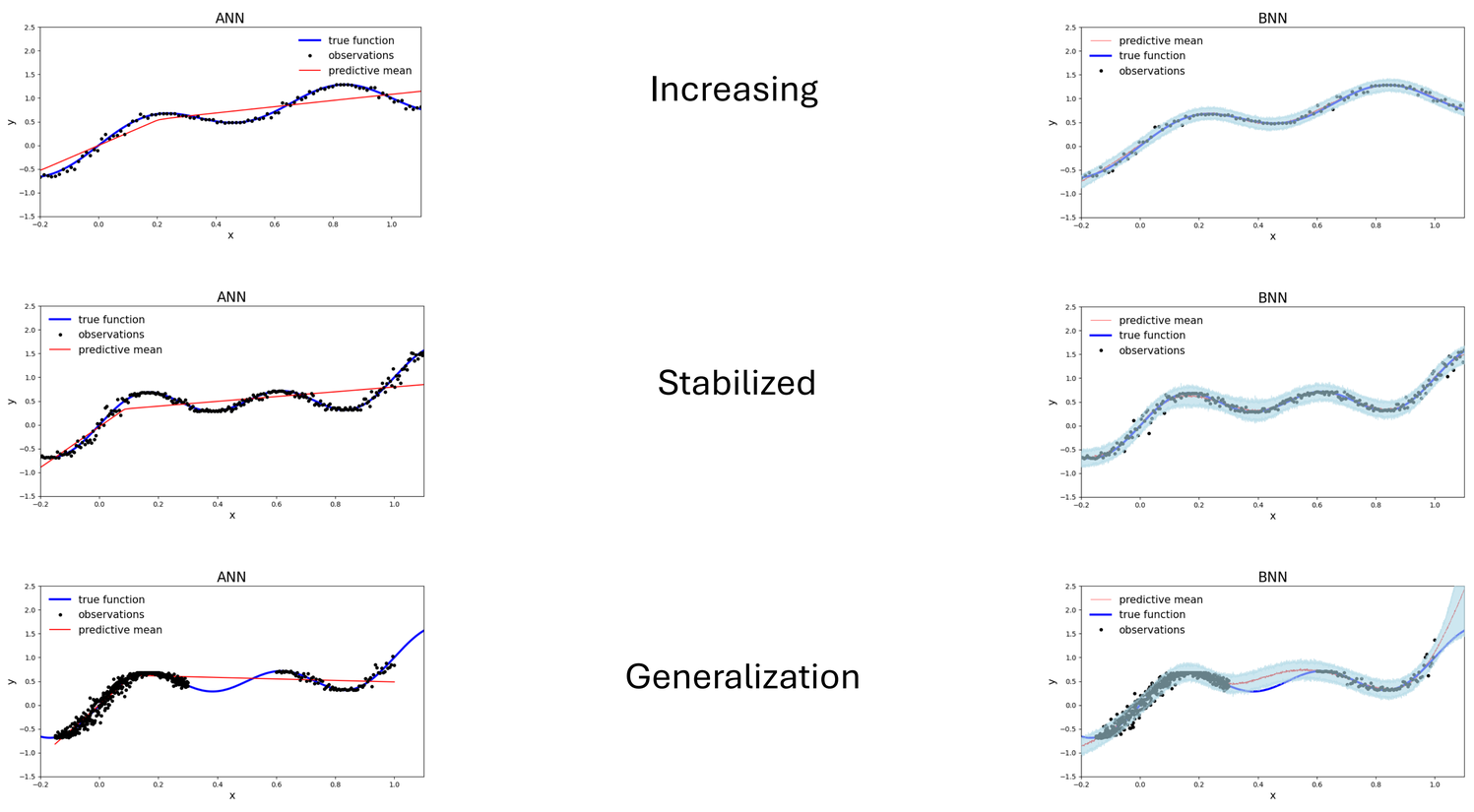
BNN addresses the issue of **explainability** integrating expert knowledge with the Bayesian model ability to detect subtle patterns.

Inputs include **structured data** such as genotypes/phenotypes, and **unstructured** data including written reports and photos.

BNN **handles small and noisy datasets.**

BNN provides a measure of model uncertainty to support **informed decisions taking**.

Performance with nonlinear functions[4]



Discussion

We propose a **hybrid Knowledge-based, Data-driven system**[3] that uses clinician's expertise on the disease to build the network and clinical data to estimate the parameters.

It addresses the issue of **transparency and explainability** and deals with the problem **of test power for small data samples**.

In a further study we will pursue to **predict the efficacy of therapeutic molecules**.

[1]European commission https://research-and-innovation.ec.europa.eu/research-area/health/rare-diseases_en, Accessed 27 November 2024.
[2]Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J Pharm Biomed Anal. 2000 Jun;22(5):717-27. doi: 10.1016/s0731-7085(99)00272-1. PMID: 10815714.
[3]Barcelona Declaration for the Proper Development and Usage of Artificial Intelligence in Europe. https://www.iiia.csic.es/barcelonadeclaration/. 2017 Mar;1-4.
[4]Inspired in UvA DL Notebooks, https://uvadlc-notebooks.readthedocs.io/en/latest/tutorial_notebooks/DL2/Bayesian_Neural_Networks/dl2_bnn_tut1_students_with_answers.html