Can we alleviate the tensions using ANN?

Kostas Dialektopoulos

H.F.R.I. Hellenic Foundation for Research & Innovation

NAZARBAYEV UNIVERSITY

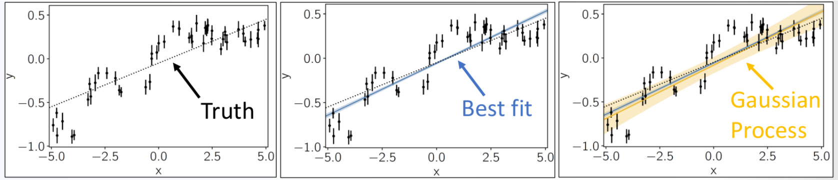
Tensions in Cosmology

Overview

1. Gaussian processes
   - What are they and how to use them?
   - H_0 estimation using GP
2. Artificial Neural Networks
   - Setup: how to construct an ANN
   - H_0 estimation using ANN

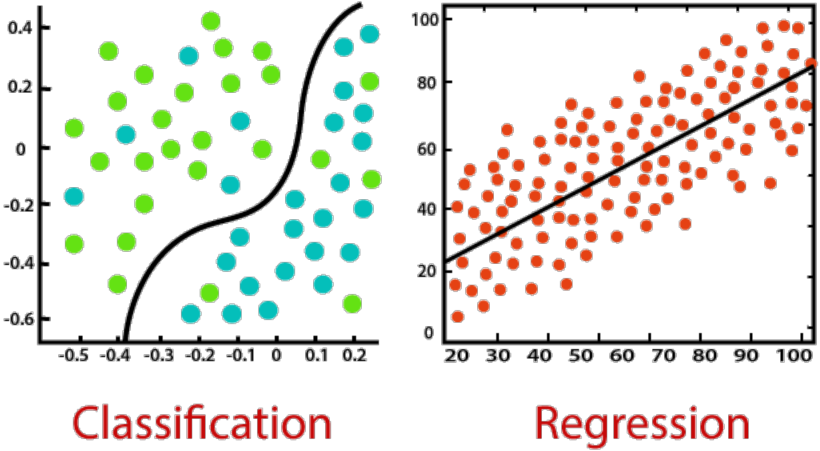
What are Gaussian processes?

A GP is a stochastic (random) process where any finite subset is a **multivariant Gaussian distribution** with mean $\mu(x)$ and covariance $k(x, x')$.



Setting each $\mu(x)$ to zero, the **covariance function** can be used to learn the behavior that produced the data points.

Regression and Classification



Gaussian Process Regression

The covariance function

consists of **hyperparameters**, which define the distribution $k(x, x')$.

The result is

a **model independent reconstruction** (in physics) of the behabior of some parameter.

Iterating over these values

using **Bayesian inference**, we optimize the hyperparameters.

Better than regular fitting

because it is **nonparametric** and it **assumes no physical model**.

Covariance functions

Squared Exponential

$$k(x,x') = \sigma^2 \exp\left[-\frac{1}{2}\left(\frac{x-x'}{l}\right)^2\right]$$

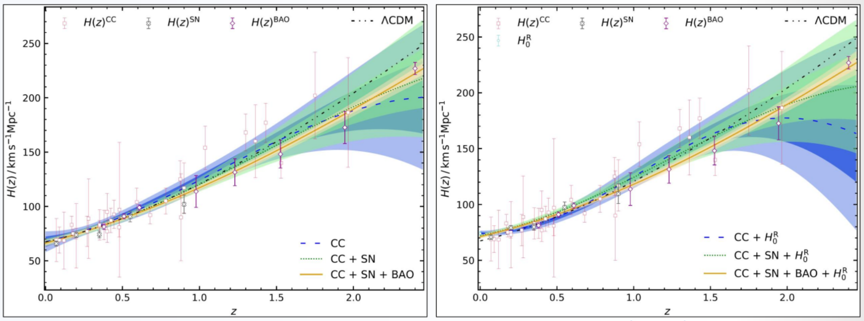
Rational Quadratic

$$k(x,x') = \sigma^2 \left(1 + \frac{(x-x')^2}{2\alpha l^2}\right)^{-\alpha}$$

Matérn

$$k(x,x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu(x-x')^2}}{l}\right)^{\nu} K_\nu \left(\frac{\sqrt{2\nu(x-x')^2}}{l}\right)$$

Squared Exponential H_0 GP (GaPP code: Seikel et al. (2012))



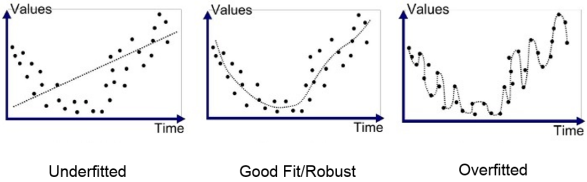
$H_0 = 67.539 \pm 4.772 \text{km/s/Mpc}$ $\quad$ $H_0 = 73.782 \pm 1.374 \text{km/s/Mpc}$

$H_0 = 67.001 \pm 1.653 \text{km/s/Mpc}$ $\quad$ $H_0 = 72.022 \pm 1.076 \text{km/s/Mpc}$

$H_0 = 66.197 \pm 1.464 \text{km/s/Mpc}$ $\quad$ $H_0 = 71.180 \pm 1.025 \text{km/s/Mpc}$

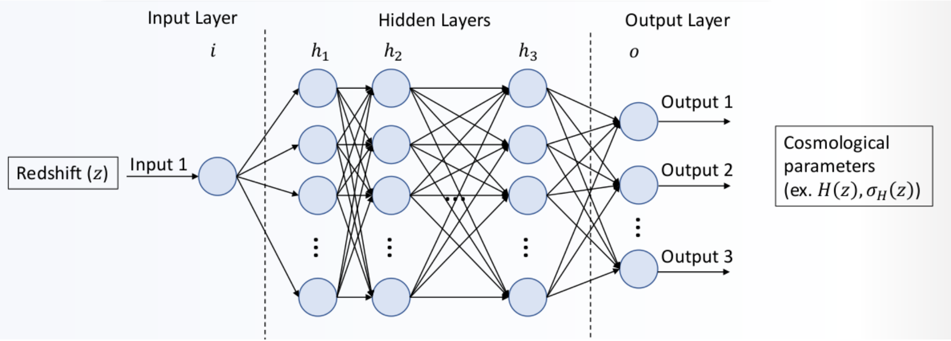
Open problems with GP reconstructions

- **Overfitting**: GP is very prone to overfitting for a limited set of data points.



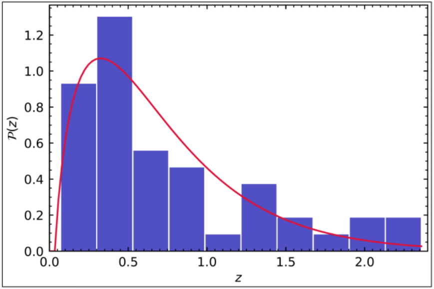
- **Kernel Selection Problem**: There is no physical reason to select the covariance function.

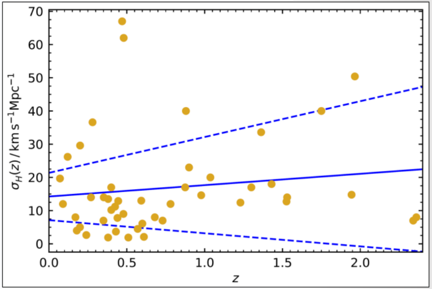
Artificial Neural Networks (ANN)



ReFANN code from Wang et al. (2020)

Training data for the ANN





$$P(z,\alpha,\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\lambda z}$$

Mean: $\sigma_H = 14.25 + 3.42z$
Upper error: $\sigma_H = 21.37 + 10.79z$
Lower error: $\sigma_H = 7.14 - 3.95z$

Designing the ANN

- **Risk**: Optimizes the **number of hidden layers and neurons** in an ANN

$$\text{risk} = \sum_{i=1}^{N} (\text{Bias}_i^2 + \text{Variance}_i) = \sum_{i=1}^{N} \left([H_{\rm obs}(z_i) - H_{\rm pred}(z_i)]^2 + \sigma_H^2(z_i) \right)$$

- **Loss**: Indicates the **number of iterations** a system needs to predict the observational data

1. Least absolute deviation (**L1**)

$$L1 = \sum_{i=1}^{N} |H_{\rm obs}(z_i) - H_{\rm pred}(z_i)|$$

2. Smoothed L1 (**SL1**)
3. Mean Square Error (**MSE**)

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (H_{\rm obs}(z_i) - H_{\rm pred}(z_i))^2$$

Building the ANN

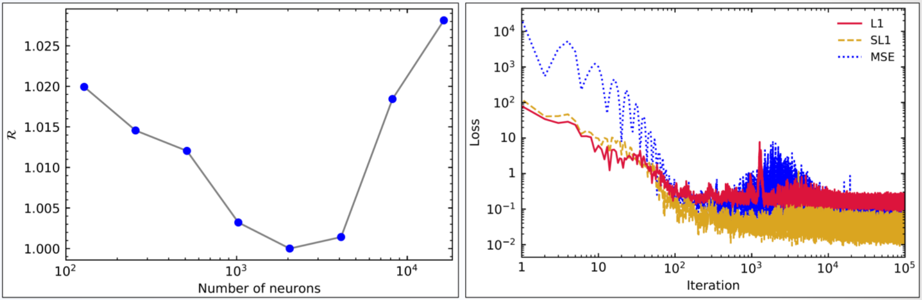


Figure: **Left:** Risk function for **one layer** (number of neurons 2^n, $n \in 7, ..., 14$), **Right:** Evolution of L1, SL1 and MSE loss functions

Using the ANN (Dialektopoulos, Levi Said et al. (2021))

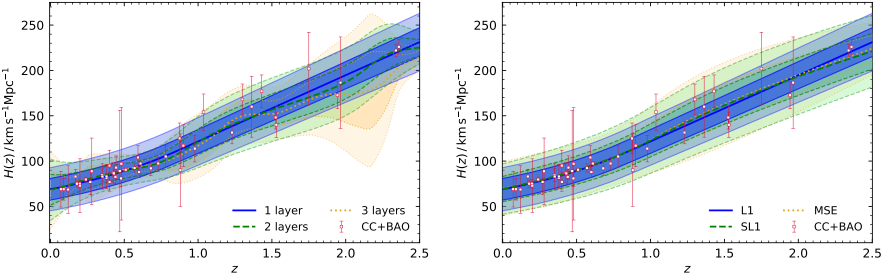


Figure: L1 $H(z)$ ANN reconstructions with different number of layers in the left panel. In the right panel we depict the $H(z)$ ANN reconstructions adapting the L1, SL1 and MSE loss functions.

Adding priors (Dialektopoulos, Levi Said et al. (2021))

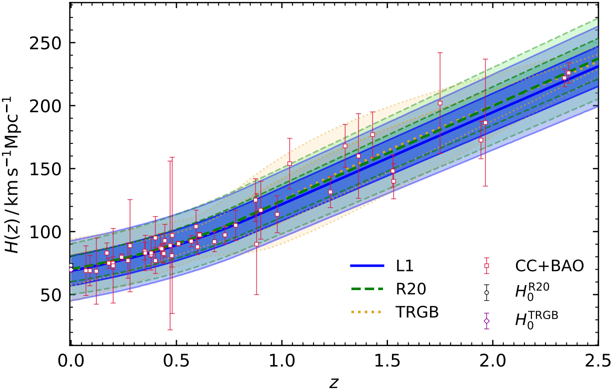


Figure: $H(z)$ ANN reconstructions when considering the L1 loss function without an H_0 prior (L1), with the R20 prior (R20) and the TRGB prior (TRGB).

Whisker plot of results

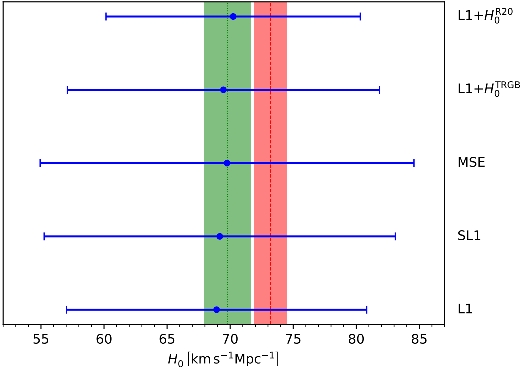


Figure: The inferred constraint on H_0 from the $H(z)$ ANN reconstructions as indicated on the vertical axis. The green and red bands illustrate the local measurements of $H_0^{\rm TRGB}$ and $H_0^{\rm R20}$, respectively.

Conclusion and Prospects

- GPs offer an interesting approach to tackle the **tension in** H_0.
- Using GP reconstruction of $H(z)$ and its derivatives we can constrain classes of modified theories of gravity.
- However, GPs suffer from **overfitting** and also introduce some bias with the **choice of the kernel**.
- With ANNs we can determine a completely **nonparametric reconstruction of the Hubble diagram.**


Thank you!
Christos