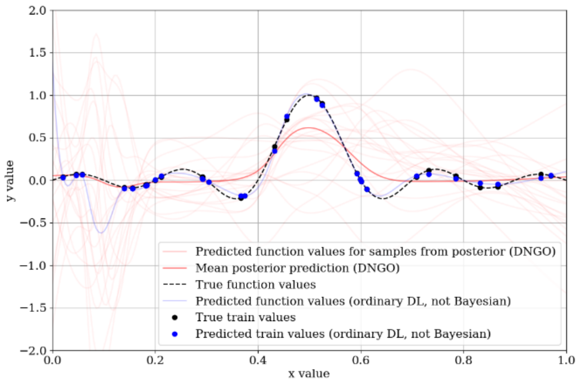
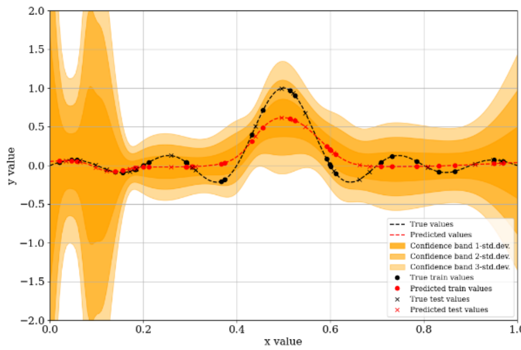
Sheet 11 – Team MTE

Marta Gulida - 5585808 - mg776
Erik Bode - 4505199 - kb301
Tillman Heisner - 4517815 - th273

Exercise 1.2) DNGO: Interpreting Uncertainty Predictions





Are the predicted uncertainties reasonable?

They are mostly reasonable: Uncertain, where few datapoints are, and more certain where more data is available.

Mention (at least) one positive and one negative aspect:

While the prediction does not match the true data, the confidence bands encapsulate the true data, with the worst being in the second band. This is not always the case, e.g. for x values around 0.1 the prediction matches the ground truth well.

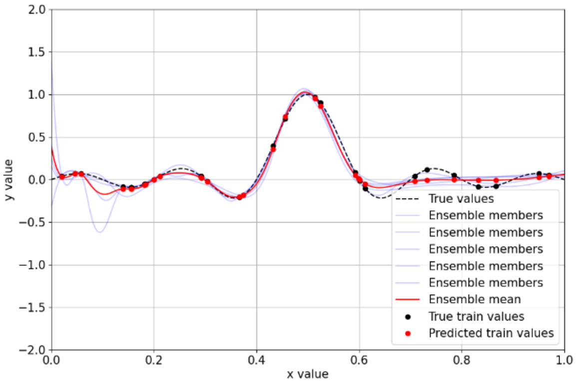
What happens to the predicted uncertainty as we make predictions for points outside of the training data range (i.e. extrapolation)?

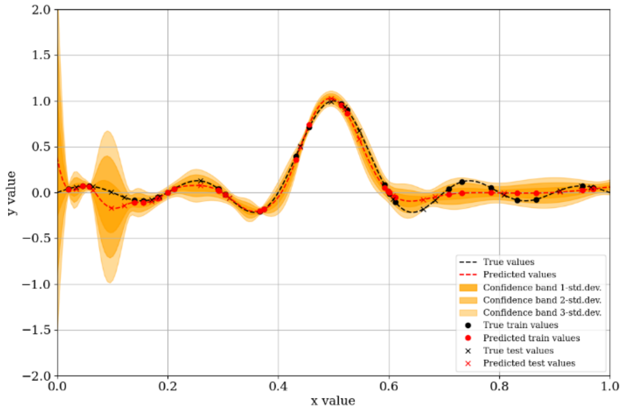
The uncertainty increases.

Assume our model would fit the training data perfectly. Would the predicted uncertainty for the training data be 0? Why (not)?

It would not, because we still have regions not covered by the training data, resulting in high uncertainty.

Exercise 2.2) Ensembles - Plot Multiple Predictions





How do these results compare to those of DNGO?

The results are way better than DNGO, the individual networks as well as the mean all match the ground with more precision.

We still have increased uncertainty in places with few training datapoints.

Is this comparison fair?

We are training more models in the ensemble, resulting in higher training time as well as higher inference time. For this reason, the comparison is not fair, even with better results.

Exercise 3) - MCMC sampling

How does the burn_in time affect the sampling process in general, and also specifically in the case of above 3 distributions?

When the burn_in time is decreased, the quality of the predicted distribution decreases. If the number of samples is also decreased sufficiently, the algorithm does not explore the entire underlying distribution, resulting in a very poor approximation.

In the first distribution, decreasing the burn_in time affects the approximation on the second peak. For extreme decreases, only the first peak is fitted.

In the second distribution, this is different: The first and third peaks are matched well, but the second peak isn't. For extreme decreases, only the first peak is fitted.

In the third distribution, the density of samples in the lower left part is explored less with decreasing burn_in time. For extreme decreases, only the upper right part is explored.

When increasing the burn_in time, a different effect occurs: For small increases, the first peaks are overfitted, but for larger increases this mostly returns to the initial distribution.

Notable are the effects on the second distribution: Increasing the burn_in time initially worsens the approximation on the second and third peak, but for a larger increase, the third peak is influenced more.