Submission for Deep Learning Exercise

Team: shallow_learning_group
Students: Batuhan Karaca

January 23, 2024

DNGO Model: Interpreting Uncertainty Predictions

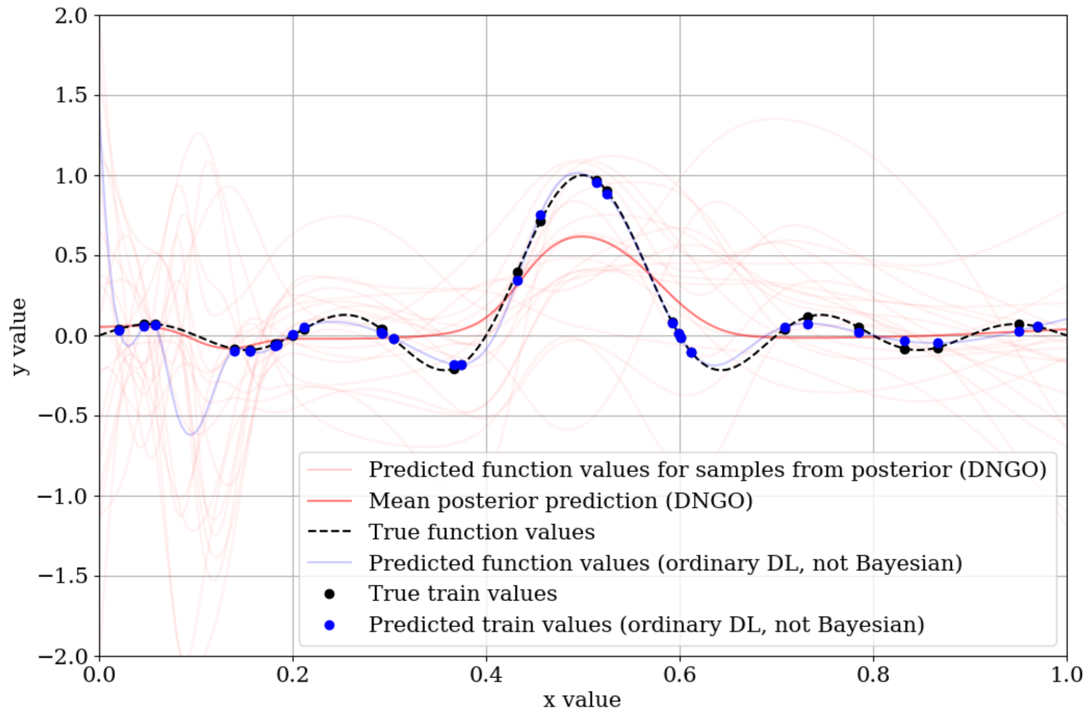


Figure 1: Graph of individual predictions

1

As seen in figure 2, as data points get more distant, the Bayesian model has more uncertainty between them as confidence bands grow, which is a downside. Furthermore in figure 1, the error rate is higher for BLR model as it assumes a prior with zero mean and noise with zero mean and constant variance (affecting likelihood). The distributions are also assumed to be Gaussian. The posterior suggested by BLR may not reflect the optimal solution as the noise may not abide by these conditions. Furthermore, sometimes the initial weights (of the network) may not be sampled from a Gaussian as people tend to use uniform weights.

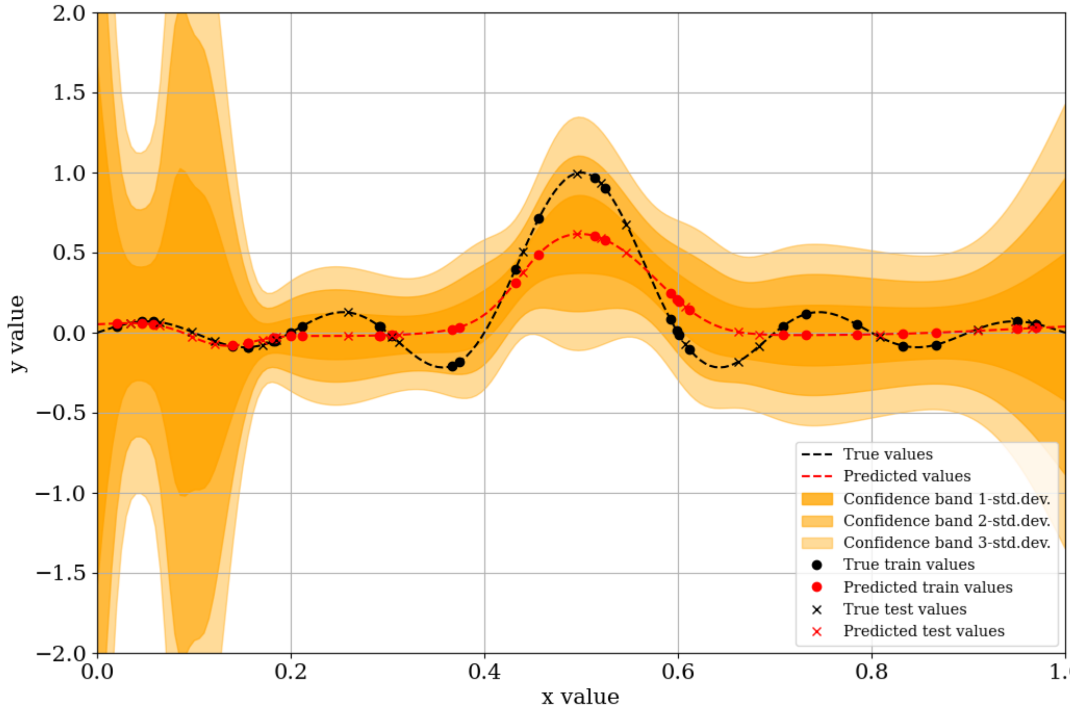


Figure 2: Graph of confidence bands of predictions. Note that for a data point, standard deviation σ of the distribution is taken. Confidence band k has perimeter of $k \cdot \sigma$

As seen in figure 1, Bayesian model finds the mean of the individual results, which is closer to the mid-point of the confidence-bands as in figure 2. This is a more moderate approximation compared to any instance, because an average distance of any point to the mean will be less than for an arbitrary distance. In this regard, Bayesian networks do some kind of regularization. If we had another data distribution, the Bayesian network would possibly perform better as it does not overfit to the data.

2

As told in the previous subsection, the uncertainty should grow between distant data points, as is the case in figure 2.

3

For the regions of uncertainty, the error rate should be higher, which is a problem for an overfitted model.

Ensembles: Plot Multiple Predictions

The ensemble takes the mean of a finite number of instances, whereas BLR model takes infinite number of instances into account. Therefore, variance for the BLR model could be higher as it includes more instances (assuming both the models were trained *well enough*).

are kind of Bayesian network as they approximate the posterior integral by summing over its instances.

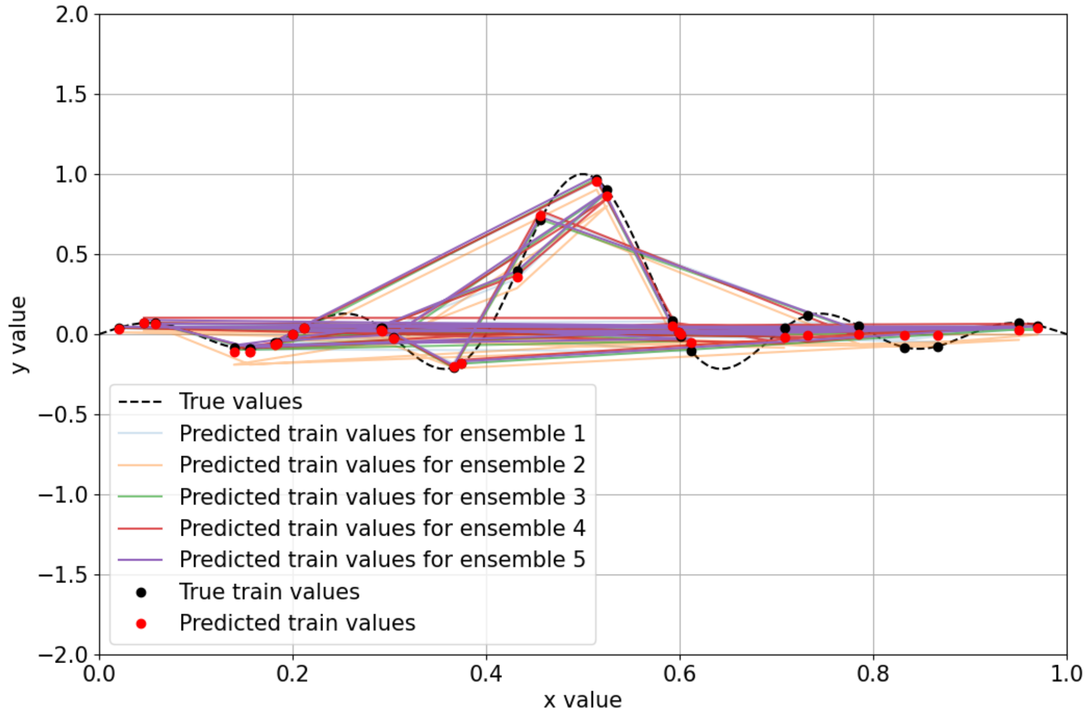


Figure 3: Graph of individual predictions

Therefore similarly, they suffer the same problem of *higher uncertainty between distant samples.*

MCMC Sampling: Plot Multiple Predictions

As shown in figure 5, the model performs the best when burn-in time is 3000. As number of samples grow, the upcoming samples depend on the previous samples in fixed time interval t, hence the initial sample. Since the initial sample may be biased, the model may perform worse with smaller burn-in time. However, it is hard to assess the difference between when $T = 3$ and $T = 3000$ for the 2D graphs.

The performance also deteriorated with larger burn-in time as well. It is evident in the first two graphs; however, (presumably as it is still hard as we may also need the third dimension) the values in the third graph wraps more nicely around the Gaussian curves

Uncertainties using PFNs

1

For the dataset, the PFN approximates GP with high precision as shown in figure 6.

2

The real dataset is only used during inference. For training, the model uses the data sampled from its prior.

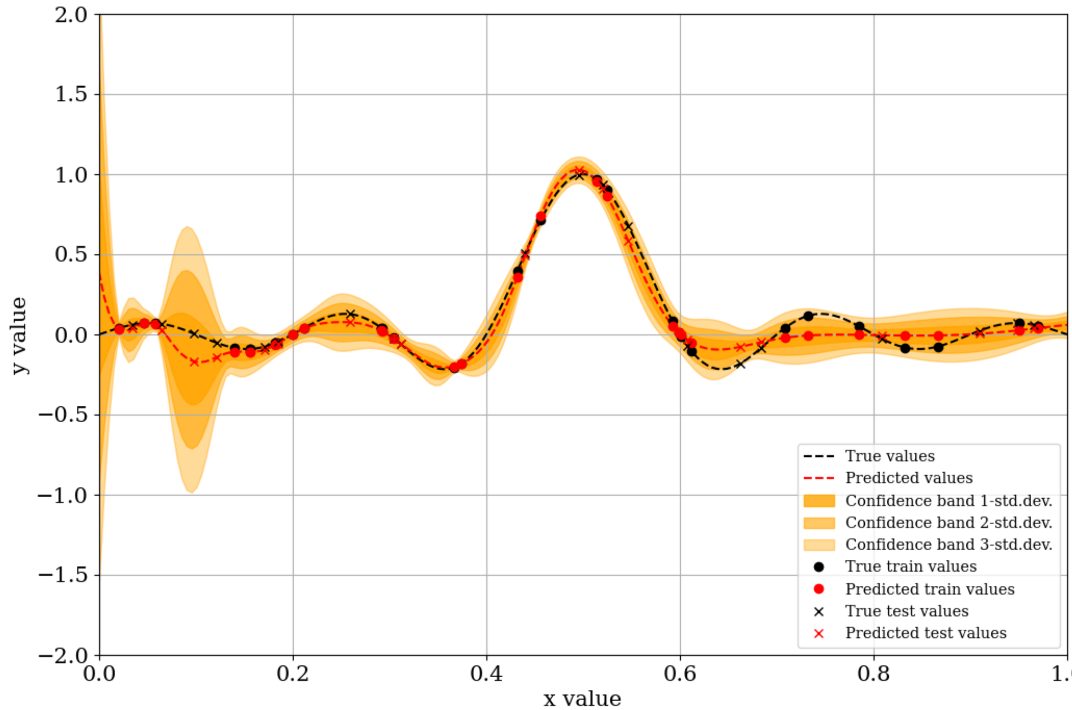


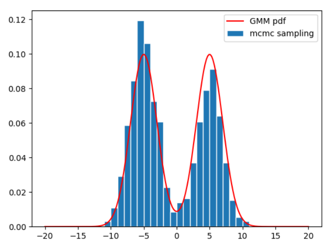
Figure 4: Graph of confidence bands of predictions.

3

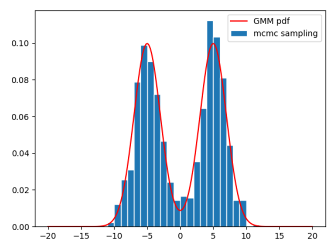
Gaussian processes assume the data has a normal distribution. PFNs does not make this generalization. The prior distribution can be chosen by the user.

4

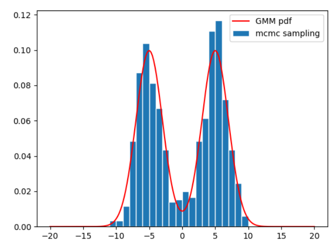
For the second dataset, the PFN is not a good approximation of GP as shown in figure 7.



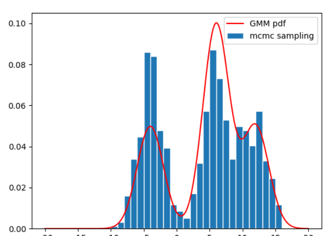
(a) T = 3



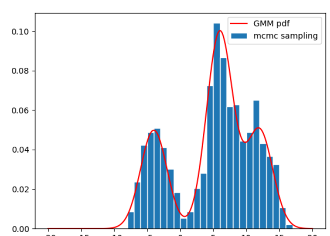
(b) T = 3000



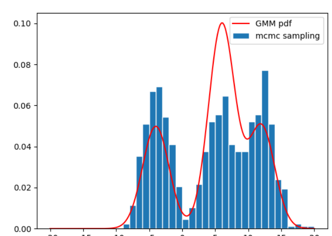
(c) T = 10000



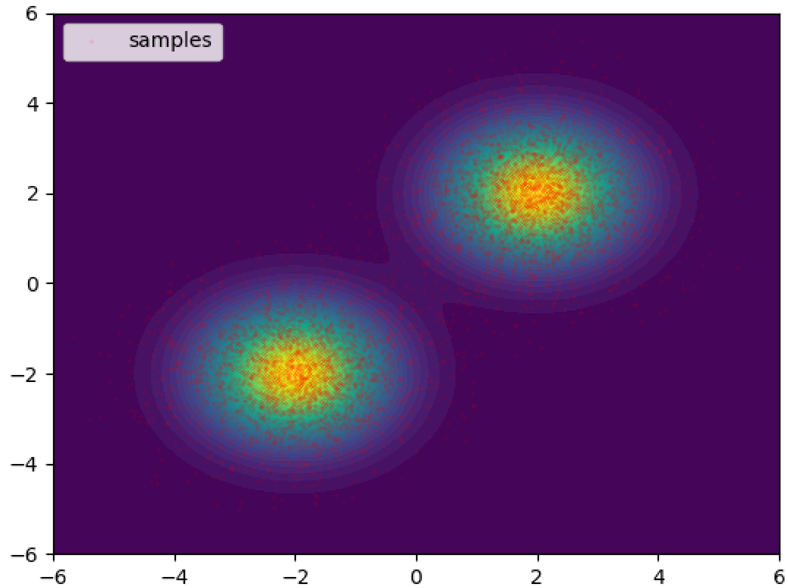
(d) T = 3



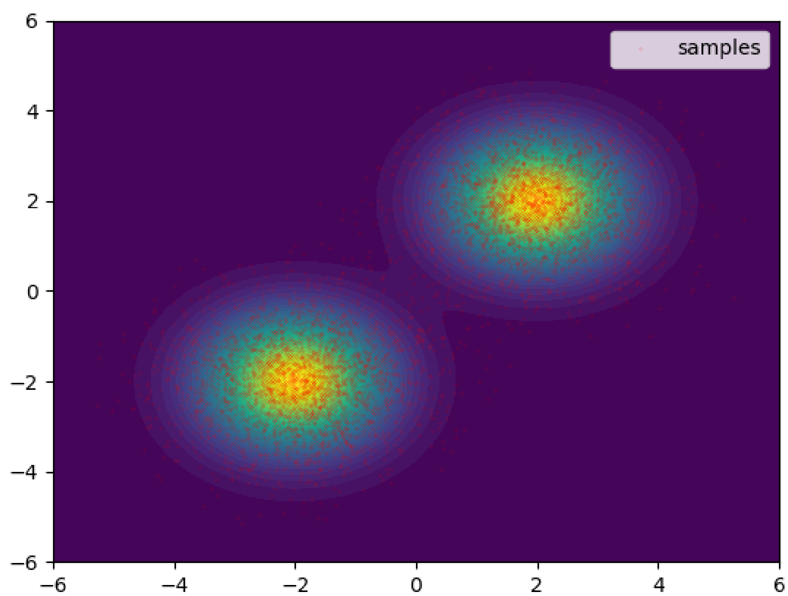
(e) T = 3000



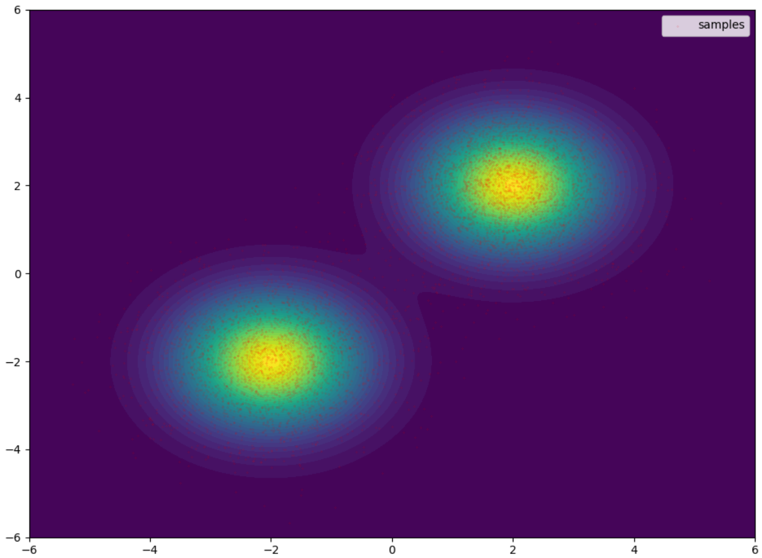
(f) T = 10000



(g) $T = 3$



(h) $T = 3000$



(i) T = 10000

Figure 5: Graph of true data and predictions. T denotes burn-in time. Triplets (a,b,c), (d,e,f), (g,h,i) show the results for the first, second and third distributions respectively

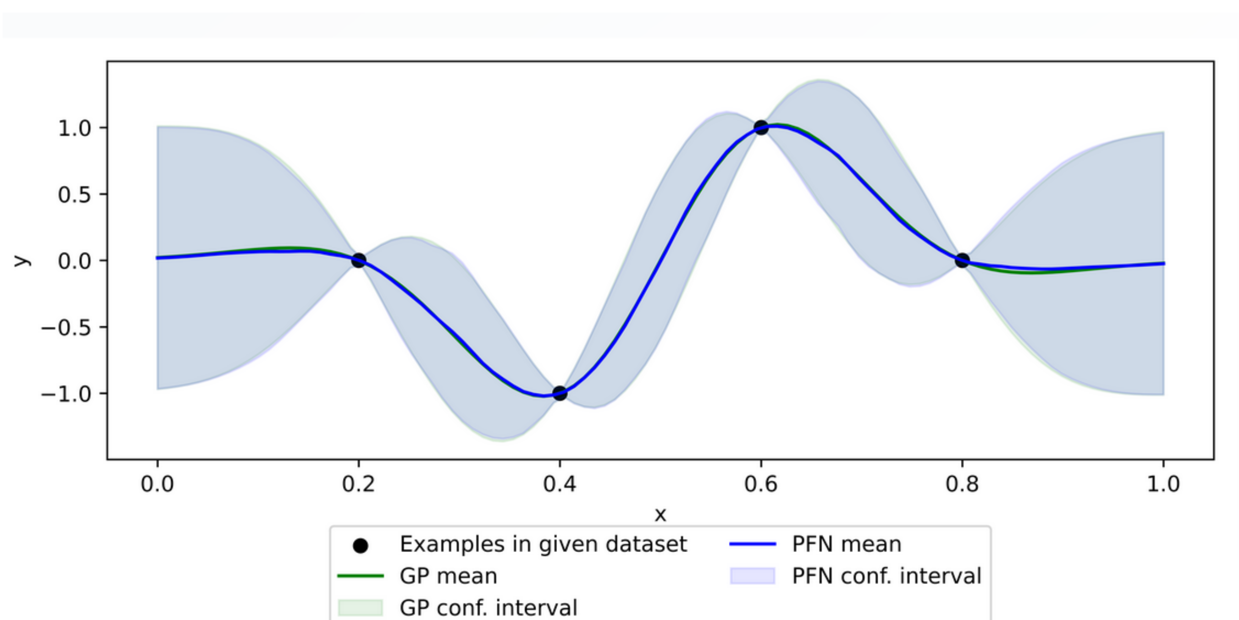


Figure 6: Plot of comparison between GP and PFN for the first dataset

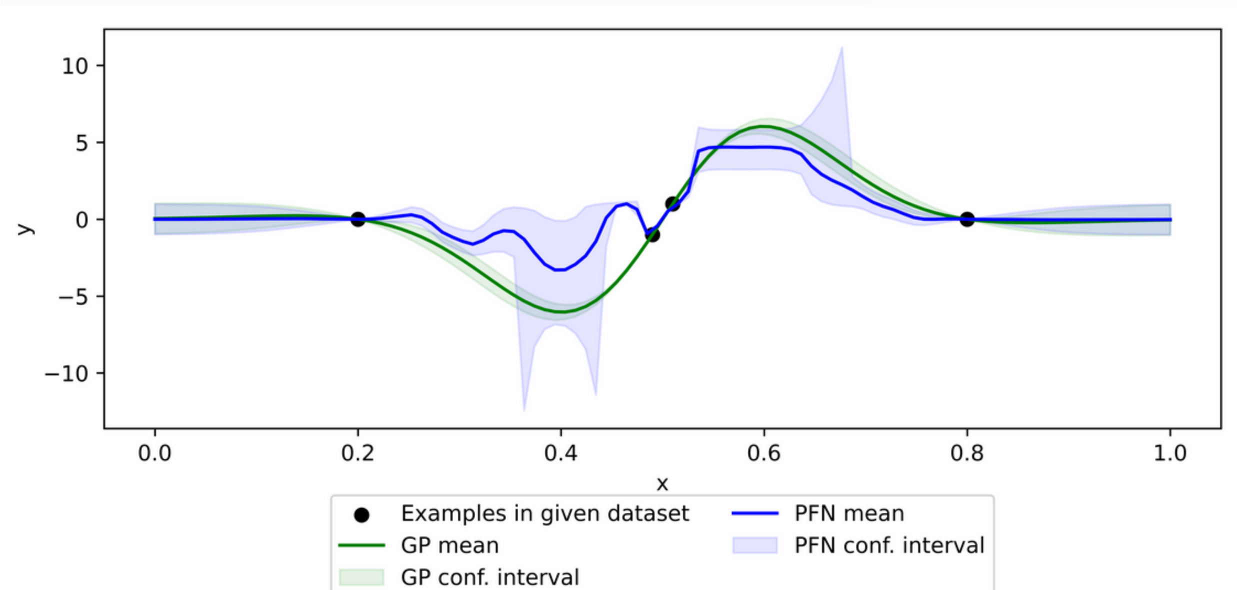


Figure 7: Plot of comparison between GP and PFN for the second dataset