Objective

The objective of this document is to demonstrate using a counter example that under the optimum variational posterior, the variance of the BNN's output (using the scaling 1/N) does not necessary converge to zero. By using the proposition 1, we can demonstrate that the variational posterior converges to the prior, and then that the BNN's output tends to the null distribution (null mean and null variance). However, this result is based on the un-tempered ELBO, when the ELBO is tempered by the right coefficient ($\eta_N = \tau p/N$), there is no raison for the variance of the BNN's output to converge to zero.

Data description

The problem studied is a toy regression $\{(x_i,\ y_i)\}$ where x_i is sampled uniformly from [-1, 1] and $y = x^3 +$, with $\epsilon \sim\ \mathcal{N}(0,\ 0.01)$. The number of data points considered in the training dataset is p = 200.

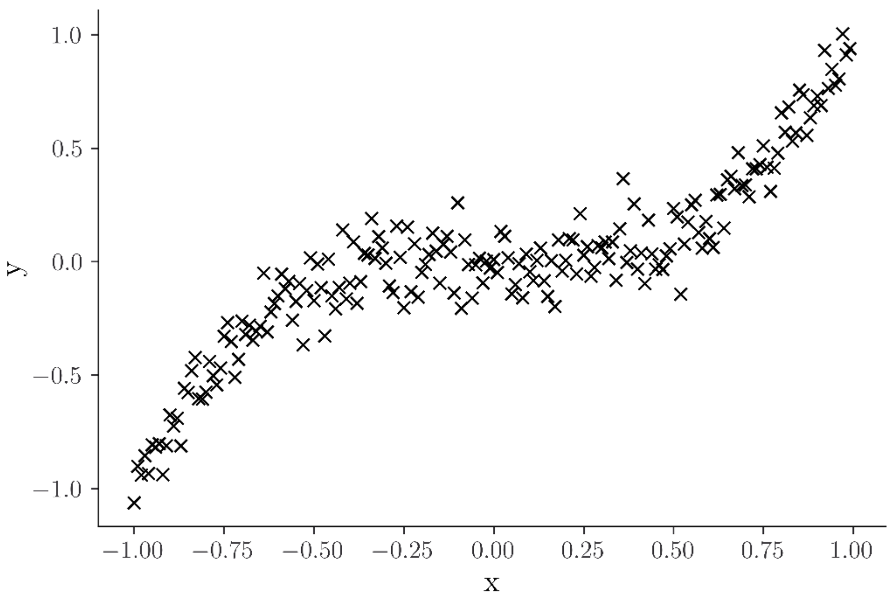


Figure 1 - Training dataset

Modelisation

The models considered are one hidden layer described by the following formula:

$$f_w(x) = \frac{1}{N}\sum_{i=1}^{N} \sigma(< b_i, x >)a_i,$$

We want to study the dependency of variance of the predictor ($\mathrm{Var}_{w\sim q_\theta}[f_w(x)]$) on the number of neurons N. Thus, we consider 9 possible values of N (10, 50, 100, 500, 1000, 5000, 10 000, 50 000 and 100 000), and for each model we will compute the empirical variance of the predictor. More precisely, to reduce the variability of our results , we will plot the expected empirical value after training:

$$E_{x\sim U[-1,1]}\left[Var_{w\sim q_\theta}[f_w(x)]\right],$$

The models are trained using the Bayes By Backprop algorithm: the ELBO is maximized using Adam optimizer (the expectations are approximated by Monte Carlo approximation with 5 samples). For each model, the following figure represents the output model after the training:

The figures representing the prediction of all models after the training can be found in the supplementary (Figure 2, Figure 3, Figure 4, Figure 5, Figure 6, Figure 7, Figure 8, Figure 9 and Figure 10). the training of all models seems to be well executed as they are able to correctly predict the true label.

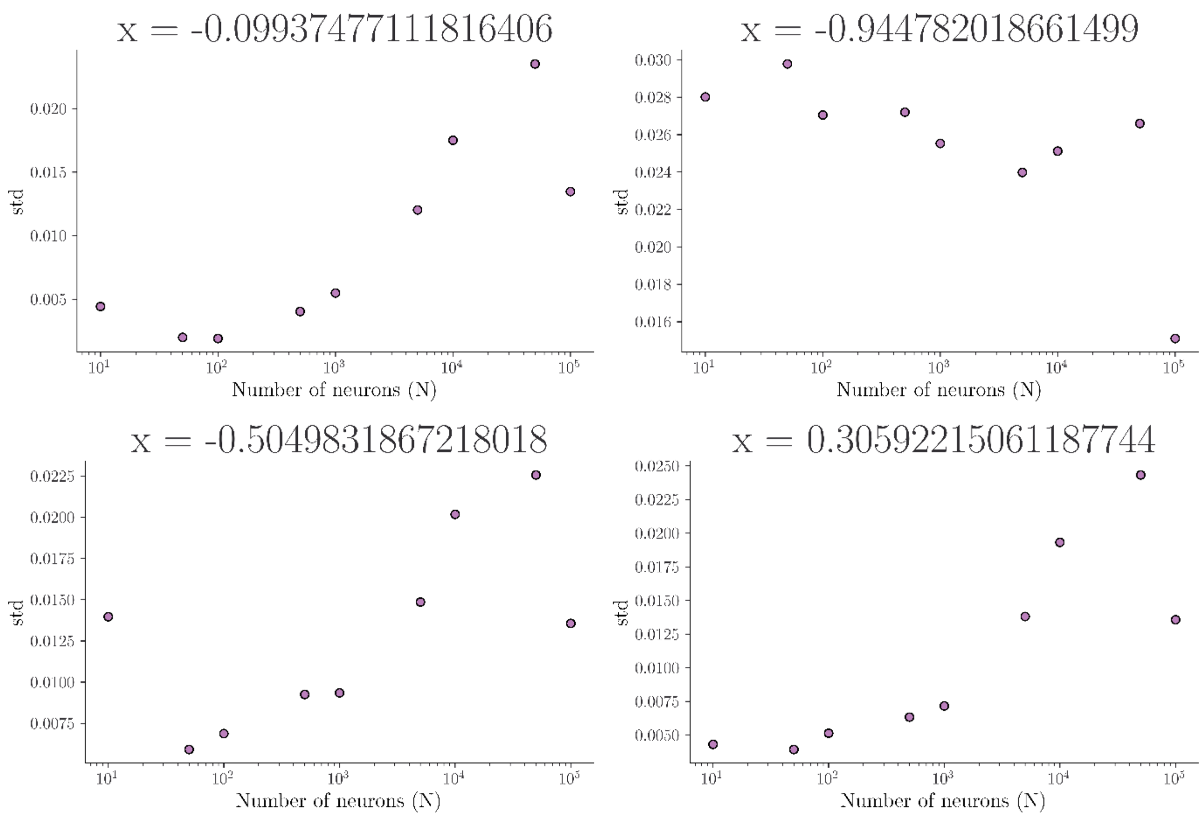
Results

Let's consider a testing point $x \sim U_{[-1,1]}$ the objective of this first experiment is to show that the variance of the predictor does not converges to zero. This variance is approximated by the empirical variance:

$$Var_{w\sim q_\theta}[f_w(x)] \approx \frac{1}{\mathrm{M - 1}}\sum_{j=1}^{M}(f_{w^j}(x) - E[f_w(x)])^2,$$

With M = 100 samples and $w^j \sim q_\theta$.

The following figures study the impact of N on the variance of the BNN's output for certain x sampled uniformly:

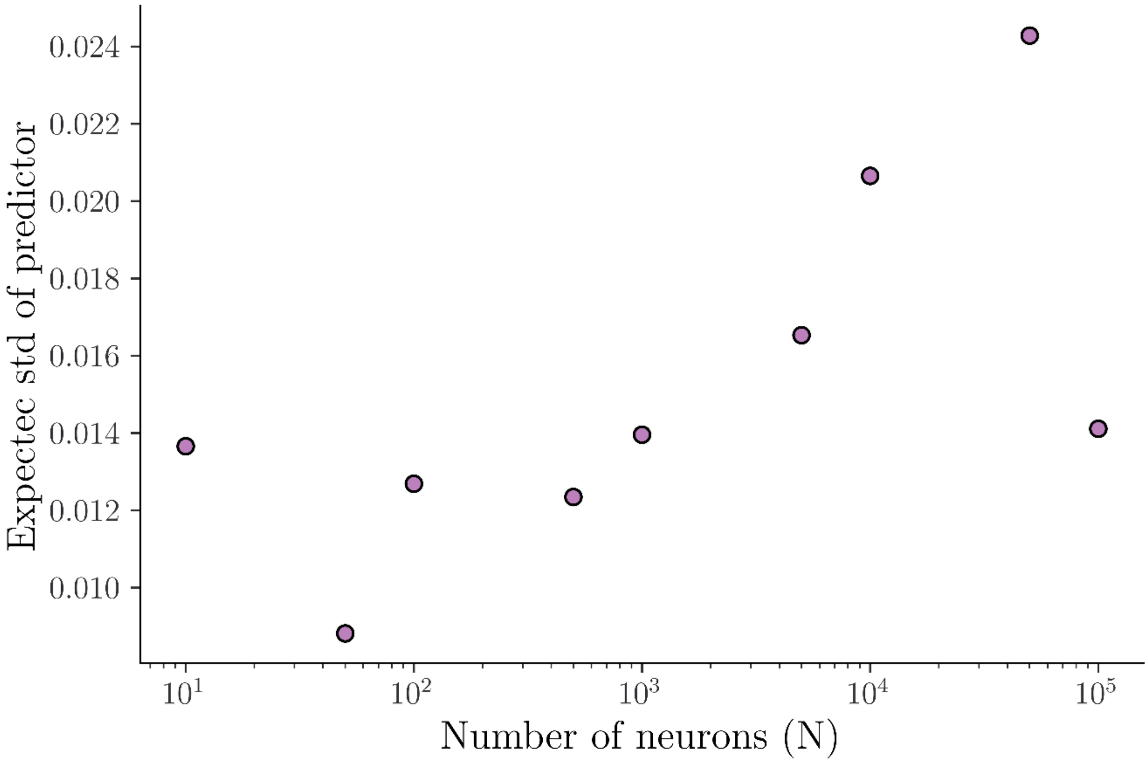


Comments:

For these inputs, we cannot see any pattern demonstrating that the variance of the BNN's output tends to zero as N increases. However, these results are dependent on the input x. To reduce the variability of our results, we will compute the expected variance of the BNN's output:

$$E_{x\sim U[-1,1]}\left[Var_{w\sim q_\theta}[f_w(x)]\right],$$

Where both the expectation and the variance are approximated by their empirical values using 100 samples. The following figure represents this expected variance for the several N studied:



Comments:

Exactly as expected, there is no pattern demonstrating that this expected variance tends to zero as N increases. The standard deviation obtained by our method is low (≈ 0.015) compared to the true standard deviation of the data (≈ 0.1), but it is well-known in the literature that BNN trained by variational inference under estimate the variance. The ratio p/N considered is $2.10^3 >> 1$, the overparametrized regime is clearly obtained.

Hyper-parameters

Learning rate	0.1
Mean initialization value	0.
Std initialization value	1.31
Loss	Square Loss
Dimension input	1
Dimension output	1
Std prior	1.
Mean prior	0
Number of epochs	300
Monte carlo samples	5
Number of data points (p)	200
Tau	10^{-5}
Number of sample for empirical mean	100
Number of samples for empirical std	100

Supplementary

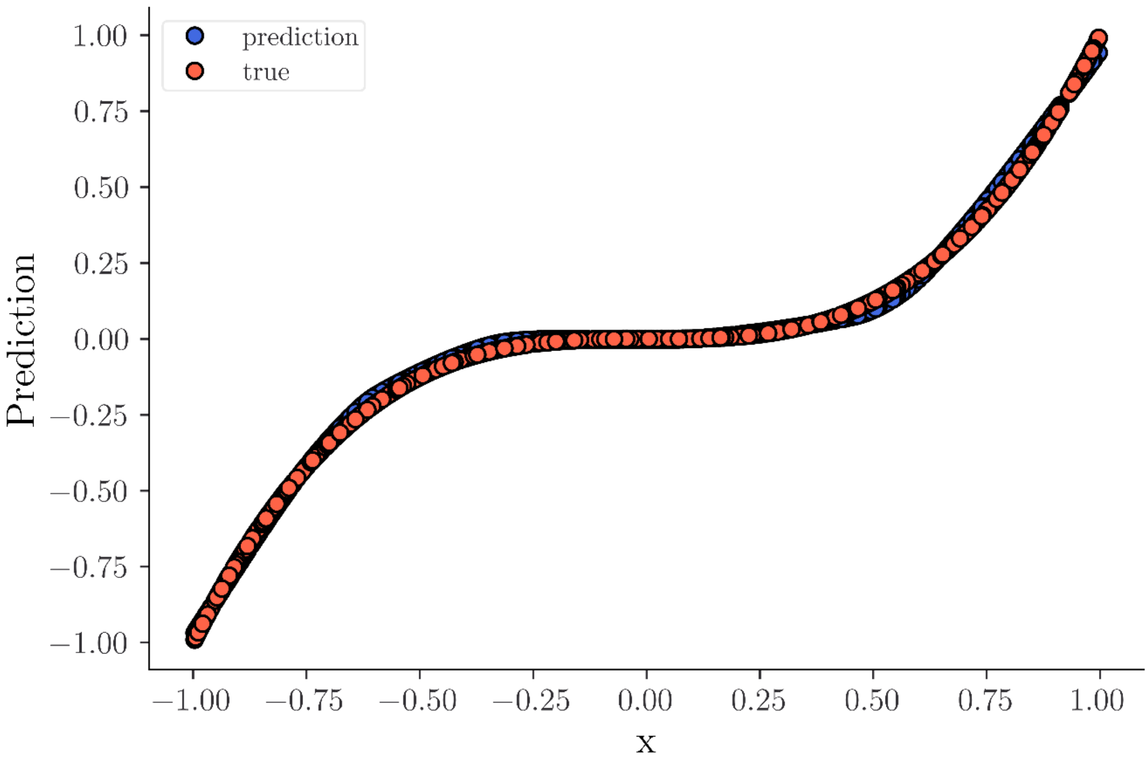


Figure 2 - N = 100 000

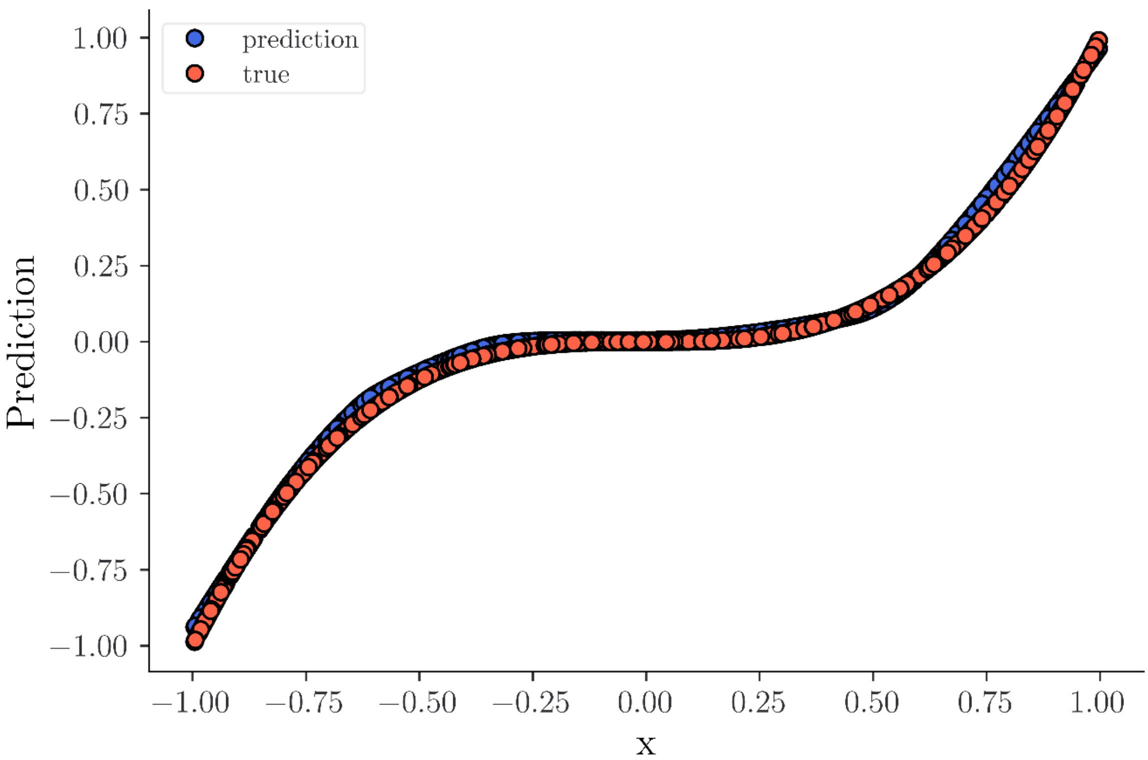


Figure 3 - N = 50 000

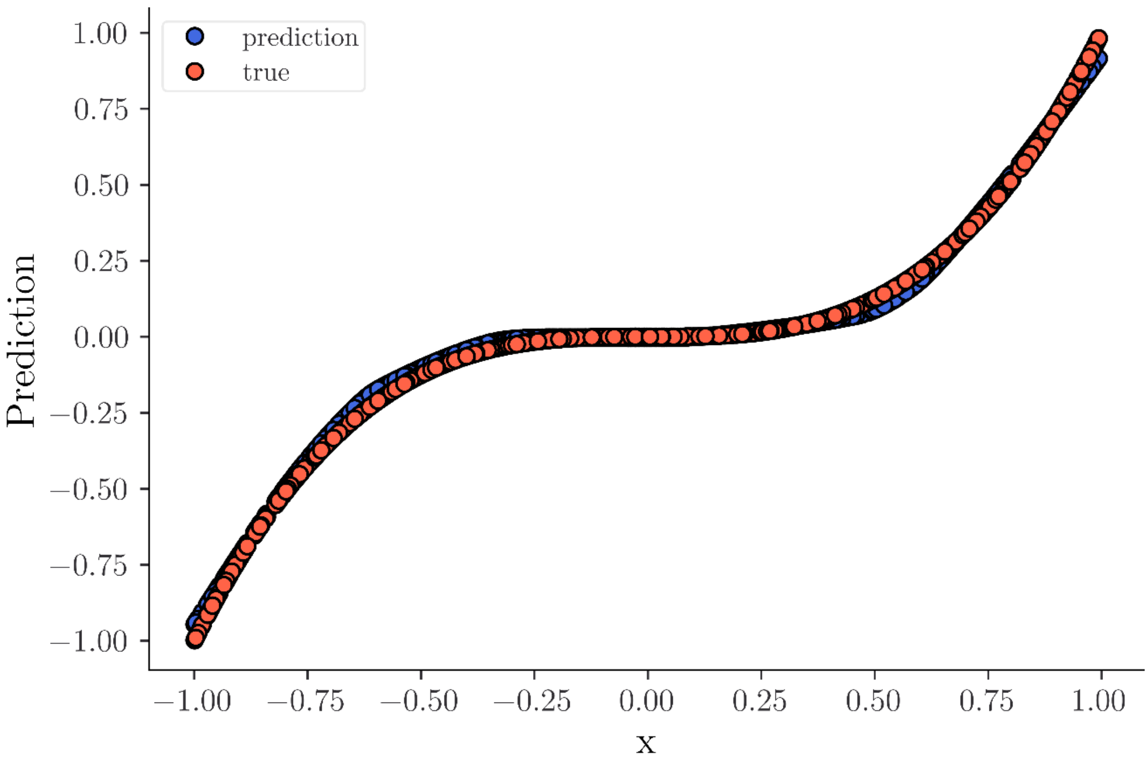


Figure 4 - N = 10 000

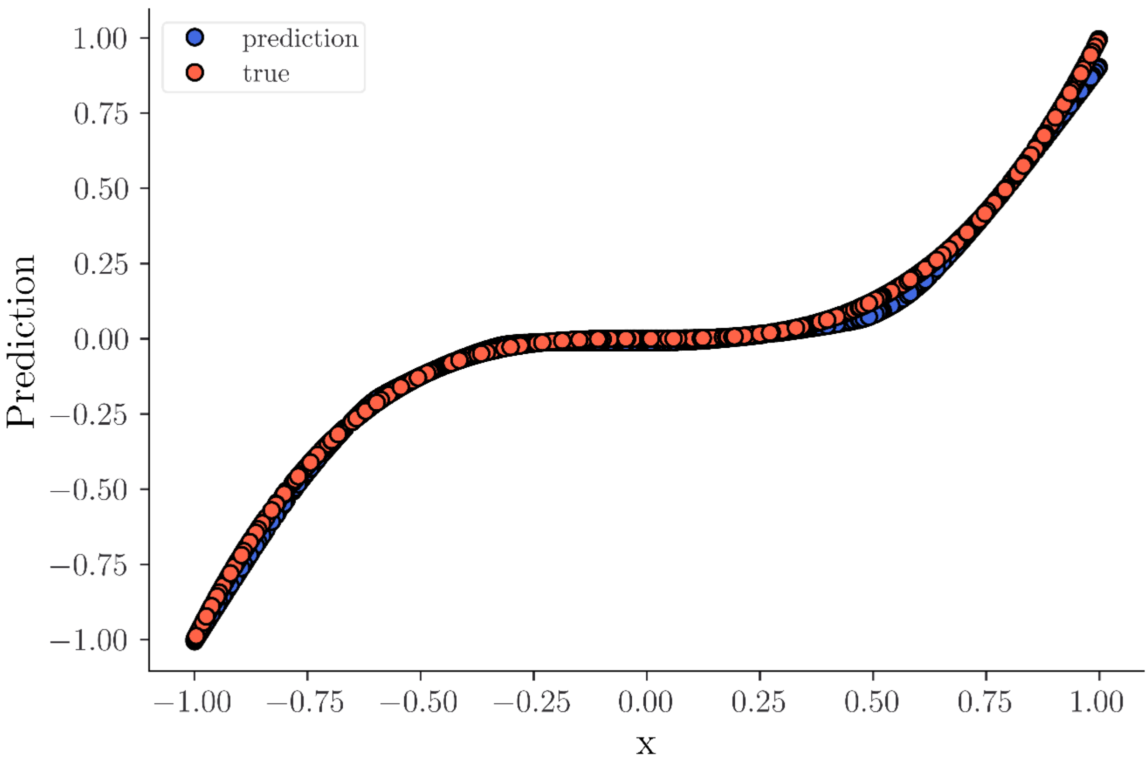


Figure 5 - N = 5000

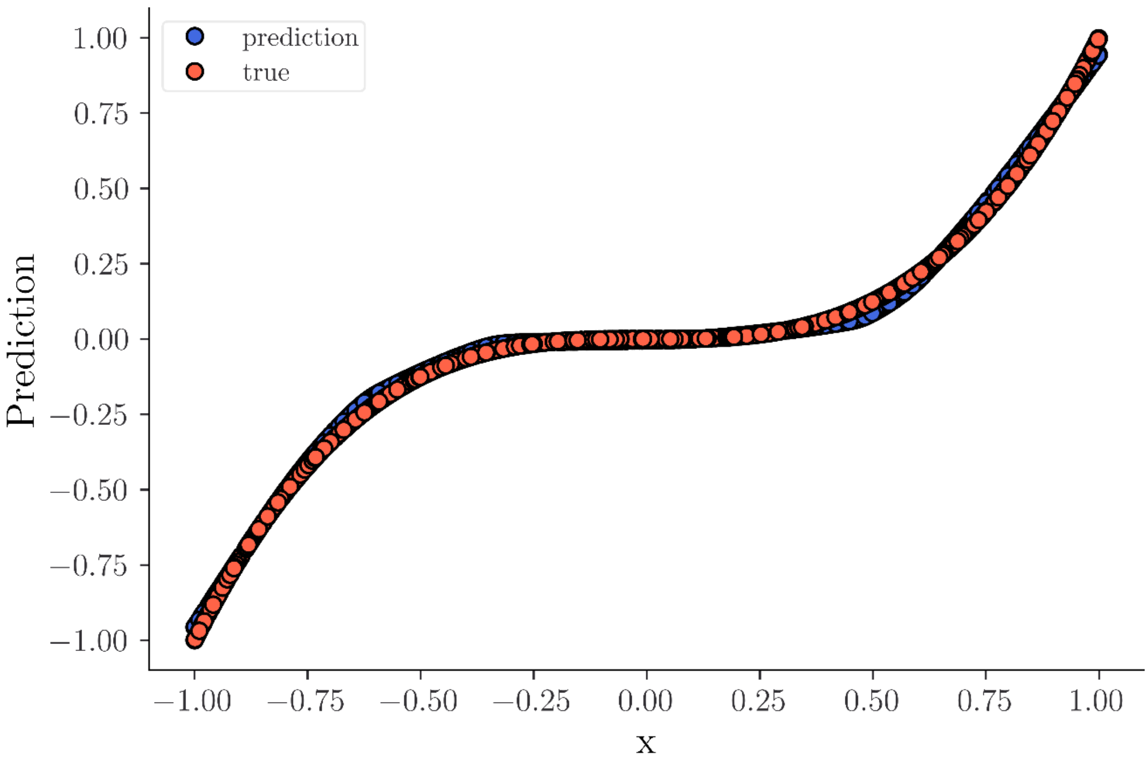


Figure 6 - N = 1000

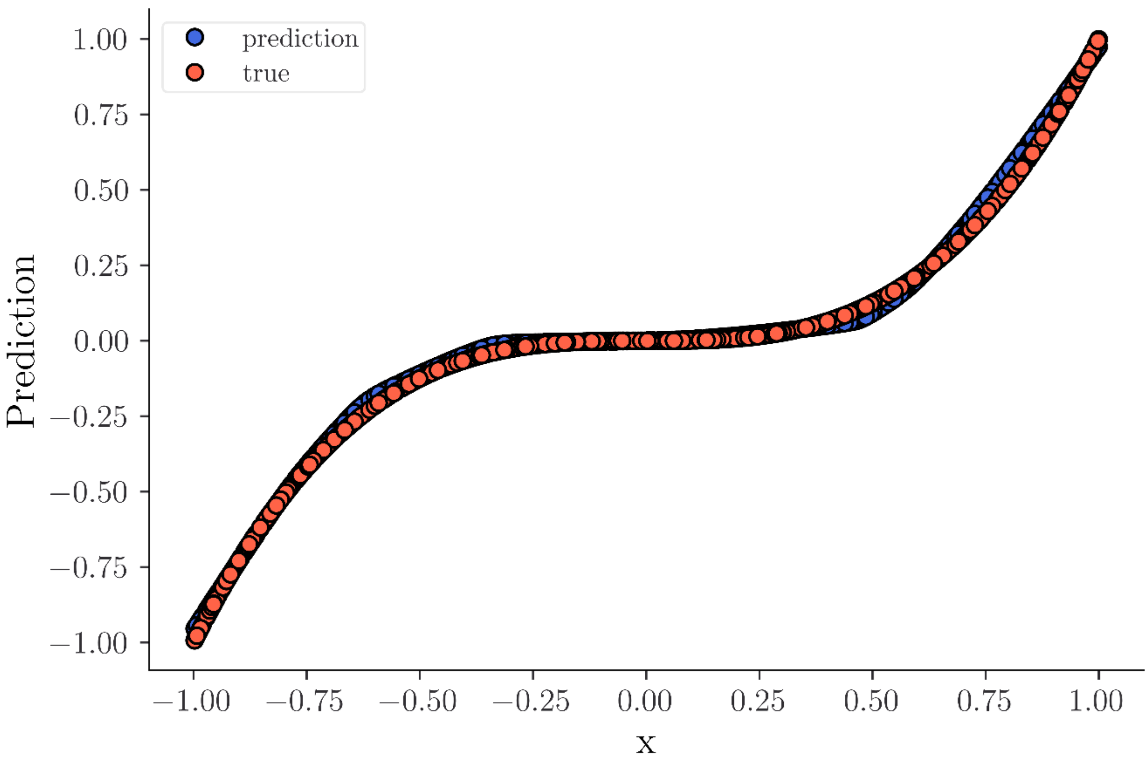


Figure 7 - N = 500

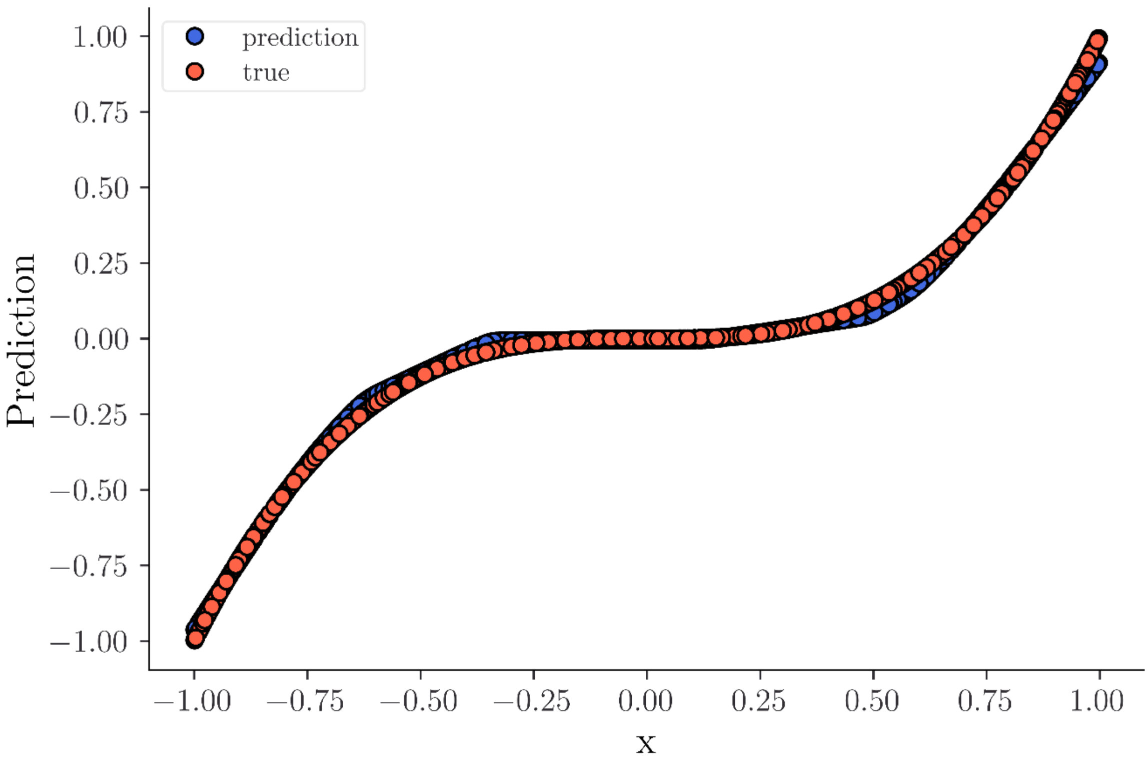


Figure 8 - N = 100

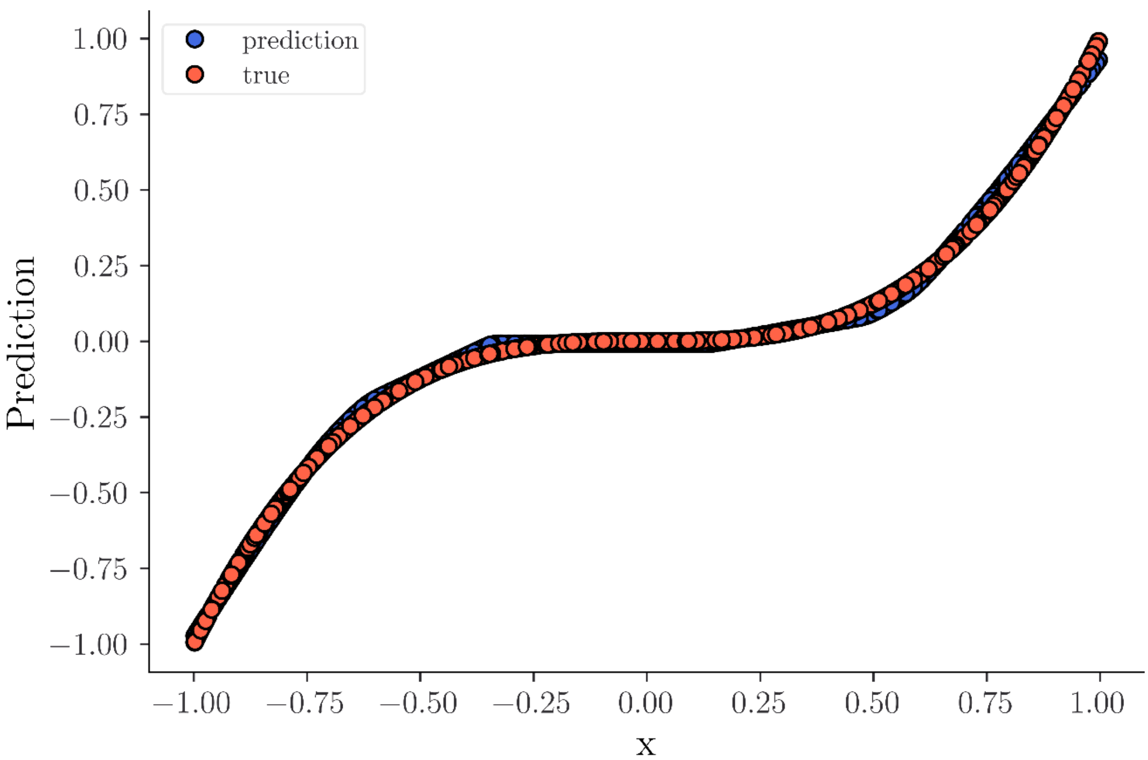


Figure 9 - N = 50

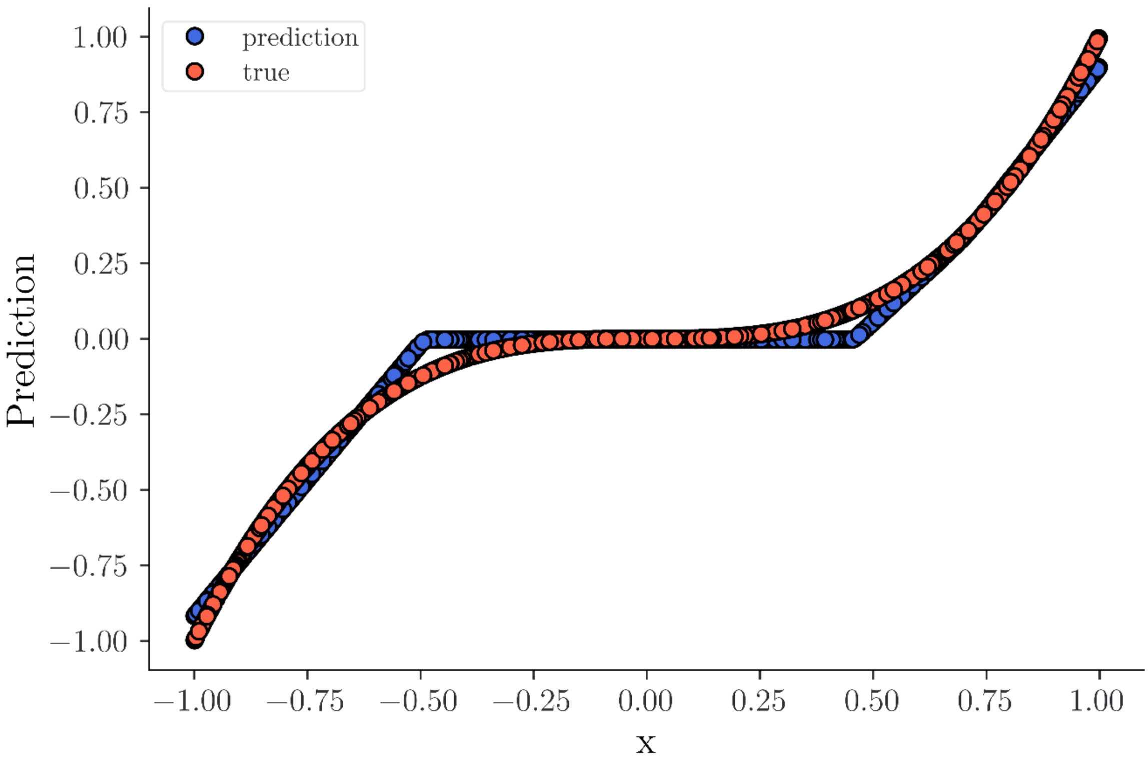


Figure 10 - N = 10