

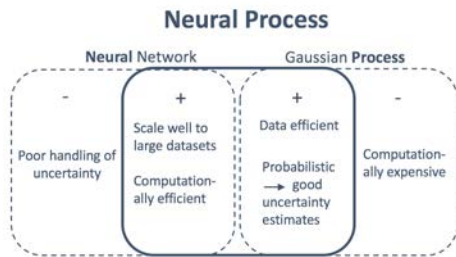
# Neural Processes

Arduin Findeis, Florian Langer, Andrius Ovsianas

Department of Engineering, University of Cambridge

## Motivation

Neural processes (NPs) attempt to combine the best characteristics of Gaussian processes and neural networks.



## Model

The aim of NPs is to do inference for a set of target variables  $y_{m+1:n}$  at location  $x_{m+1:n}$  given some context  $(x_{1:m}, y_{1:m})$ .



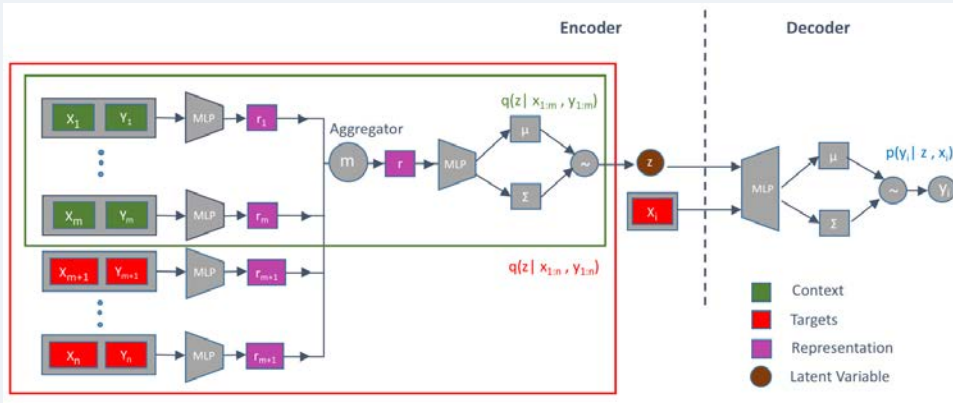
For this purpose the context is passed through an **encoder** which returns a probability distribution over a latent variable  $z$ . In contrast to CNPs [1] this allows to model global uncertainty. Passing a sampled  $z$  and a target point  $x_i$  into the **decoder** one obtains a probability distribution  $p(y_i|x_i, z)$  over the target.

## Training

During training one has to maximise the evidence lower-bound (ELBO) which can be approximated by

$$\log p(y_{m+1:n}|x_{1:n}, y_{1:m}) \geq \mathbb{E}_{q(z|x_{1:m}, y_{1:m})} \left[ \sum_{i=m+1}^n \log p(y_i|z, x_i) + \log \frac{q(z|x_{1:m}, y_{1:m})}{q(z|x_{1:m}, y_{1:m})} \right]. \quad (1)$$

The figure below shows the model and how the terms in equation (1) arise during the training procedure.



## Applications

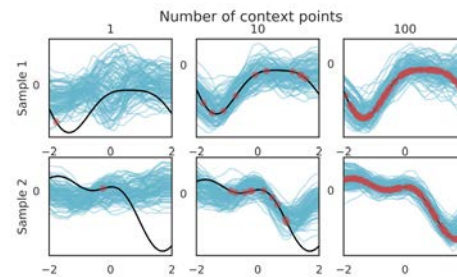
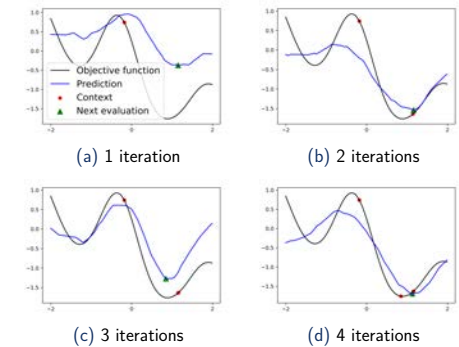


Figure: 1-D function regression with increasing number of context points (shown in red).



Neural Process	Gaussian Process	Random Search
0.27	0.13	1.00

Figure: Optimisation using Thompson sampling: Figures (a) - (d) show the optimisation over 4 iterations. Results in the table are the normalised average number of function evaluations needed to reach the global minimum of 10,000 functions randomly sampled from a GP.

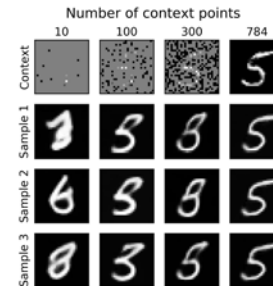


Figure: Pixel-wise image completion can be framed as a 2-D regression task. Given a varying number of context points in the first row, rows two to four show sampled predictions for the MNIST dataset.

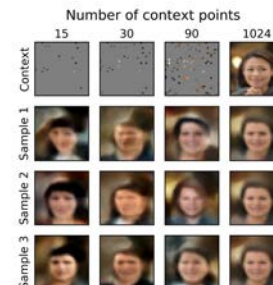


Figure: Image completion task for the CelebA dataset. For a small number of context points the predictions differ significantly. As the number of context points increases the variance across the samples decreases.

## Critique

- Compared to CNPs the variance of the predictions (marginalized over the latent variable) is large even with many context points.
- Thompson sampling experiment results are highly dependent on the chosen kernel parameters.

## References

- [1] Marta Garnelo, Dan Rosenbaum, Chris J Maddison, Tiago Ramalho, David Saxton, Murray Shanahan, Yee Whye Teh, Danilo J Rezende, and SM Eslami. Conditional neural processes. 2018.
- [2] Marta Garnelo, Jonathan Schwarz, Dan Rosenbaum, Fabio Viola, Danilo J Rezende, SM Eslami, and Yee Whye Teh. Neural processes. 2018.