Conditional Neural Progress

Chen Zhao

ICML 2018, `https://arxiv.org/pdf/1807.01613.pdf`

1 Introduction

Deep neural networks excel at function approximation, yet they are typically trained from scratch for each new function. On the other hand, Bayesian methods, such as Gaussian Processes (GPs), exploit prior knowledge to quickly infer the shape of a new function at test time. Yet GPs are computationally expensive, and it can be hard to design appropriate priors. In this paper we propose a family of neural models, Conditional Neural Processes (CNPs), that combine the benefits of both. CNPs are inspired by the flexibility of stochastic processes such as GPs, but are structured as neural networks and trained via gradient descent. CNPs make accurate predictions after observing only a handful of training data points, yet scale to complex functions and large datasets. We demonstrate the performance and versatility of the approach on a range of canonical machine learning tasks, including regression, classification and image completion.

In this work we propose a family of models that represent solutions to the supervised problem, and an end-to-end training approach to learning them, that combine neural networks with features reminiscent of Gaussian Processes. We call this family of models Conditional Neural Processes (CNPs), as an allusion to the fact that they define conditional distributions over functions given a set of observations. The dependence of a CNP on the observations is parametrized by a neural network that is invariant under permutations of its inputs. We focus on architectures that scale as $O(n+m)$ at test time, where n, m are the number of observations and targets, respectively. In its most basic form a CNP embeds each observation, aggregates these embeddings into a further embedding of fixed dimension with a symmetric aggregator, and conditions the function g on the aggregate embedding; see Figure 1 for a schematic representation. CNPs are trained by sampling a random dataset and following a gradient step to maximize the conditional likelihood of a random subset of targets given a random observation set. This encourages CNPs to perform well across a variety of settings, i.e. $n \ll m$ or $n \gg m$.

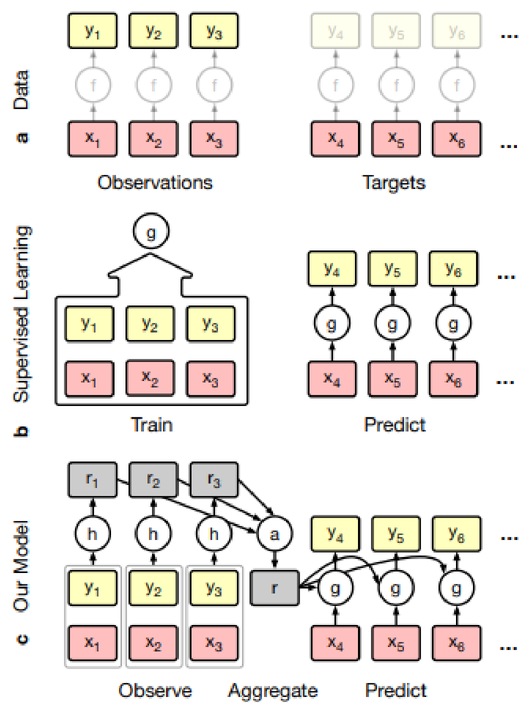

Figure 1. **Conditional Neural Process.** a) Data description b) Training regime of conventional supervised deep learning models c) Our model.