Aim: Implement state of the art deep SSL methods in an easy to use framework.

SSL applications are not benefiting from the deep learning revolution as:

- **Problem**
  - Why
    - Most research has been restricted to simple computer vision tasks.
    - Many deep SSL methods are based on latent-variable models, which tend to under-perform compared to transfer learning due to approximate inference.
    - State of the art algorithms require large amount of data.

- **What**: Semi-supervised learning (SSL) tackles problems with labelled \( X_{labeled} \) and unlabelled points \( X_{unlabeled} \). Typically \( L \ll U \). The assumption is that decision boundaries pass through low density regions.

- **Why**: Important real-life scenarios, as labels are usually expensive and cumbersome to obtain, but state of the art algorithms require large amount of data.

Semi Supervised Learning Framework

- **Introduction**
  - What: Semi-supervised learning (SSL) tackles problems with labelled \( (X_{C}, Y_{C}) \) and unlabelled points \( X_{T} \). Typically \( L \ll U \). The assumption is that decision boundaries pass through low density regions.
  - Why: Important real-life scenarios, as labels are usually expensive and cumbersome to obtain, but state of the art algorithms require large amount of data.
  - Problem: SSL applications are not benefiting from the deep learning revolution as:
    - Most research has been restricted to simple computer vision tasks.
    - Many deep SSL methods are based on latent-variable models, which tend to under-perform compared to transfer learning due to approximate inference.
    - Real-world data often has missing features or comprises of unevenly spaced time series, making them hard to use by deep learning.
  - Aim: Implement state of the art deep SSL methods in an easy to use framework.

- **Neural Processes Family**
  - **A CNP is a neutral model \( \psi_r \) inspired by conditional stochastic processes, which predict the target posterior at \( X_T \) conditioned on context points \( (X_C, Y_C) \):**

\[
p(Y_T|X_T, X_C, Y_C) \approx p(Y_T|X_T, T_C) \prod_i q_t(y_i|x_l, T_C)
\]

  - **\( \psi_r \) denotes any commutative operation, thereby enforcing permutation invariance in the context \( (X_C, Y_C) \)**
  - **The fixed dimensional representation \( r_C \) assumption (Eq. 1a) gains in scalability:**
    - The latent representation \( r_C \) assumption (Eq. 1a) gains in scalability: \( O(C+T) \) but makes the model non consistent. i.e. auto-regressive prediction give different results than predicting at once.
    - The factorisation assumption (Eq. 1b) enforces permutation invariance in \( X_T \).
  - **CNP require training on a distribution over functions.**
  - **This framework has been extended to have latent variable \( [3] \) and use target-dependent attention of the context \( [4] \).**

- **Initial Experiments**
  - **Fig. 3** shows that the target function is nearly completely contained in the region of uncertainty of the model. Despite this, the model seems to underfit as it does not pass through all context points and always breaks in the extrapolation setting.

- **Fig. 4** shows how attention improves model compared to Fig. 3. Although the model has a much better fit, it is still unable to extrapolate, which is one of issues we want to solve by incorporating an inductive bias which enforces equivariance.

Fig. 1: Label Spreading [2] decision boundaries and confidence, when predicting in a 2-dimensional latent space of a variational auto encoder for 100 labels of MNIST.

In order to help and encourage future work in SSL we have implemented a general framework with many recent deep SSL methods [1].

**References**


Fig. 2: (Left) Computational graph of the general Neural Process framework. Right) Zoom in on the \( \psi \) module, incorporating the Conditional Neural Process [1] (top center), the latent Neural Process [3] (top right) and the Attentive Neural Process [4] (bottom).

Fig. 3: Approximations of Gaussian Processes (GP) using Conditional Neural Processes (CNP). During training, samples from a GP with a fixed kernel were generated and a random number of points were used as target and context. The dotted black line shows the target sample from the GP, while the black dots are context points. The blue region depicts the mean and standard deviation predicted by the CNP. The red dotted line separates the interpolation and extrapolation setting.

Fig. 4: Attentive Neural Processes (ANP) version of Fig. 3. Note that ANP have latent variables and can thus sample different mean functions at once (the various blue lines).