Introduction
In this work we study how the learning of modular solutions can allow for effective generalization to both unseen and potentially differently distributed data.

Our main postulate is that the combination of task segmentation, modular learning and memory-based ensembling can give rise to generalization on an exponentially growing number of unseen tasks.

We provide a concrete instantiation of this idea and demonstrate that this system exhibits a number of desirable continual learning properties: robustness to catastrophic forgetting, no negative transfer and increasing levels of positive transfer as more tasks are seen. We show competitive performance against both offline and online methods on standard continual learning benchmarks.

Algorithm Setup
Node-level modular learning algorithm:
Gated Geometric Mixer (GGM)
- Well studied ensemble technique for combining probabilistic forecasts
- Basic building block for Gated Linear Networks (GLNs) [1, 2]

Automatic task segmentation and local ensembling:
- Efficient online Bayesian changepoint detection/task identification/reuse of previous learnt solutions

Experimental Results
We evaluated NCTL on diagnostic and benchmark tasks.

NCTL shows desirable continual learning properties: positive forward/backward transfer, easy interpretability and competitive performance against oracles.

Join the GLN revolution!