Variational Continual Learning
Jiajun He, Johannes Vallikivi, Nineli Lashkarashvili

Introduction
Continual learning aims to learn incrementally when data arrive in a possibly non-i.i.d. way whereby tasks may change over time, without revisiting all previous data.

Methods
Bayes Rule gives a well-defined way to perform Continual Learning:

\[ p(\theta|\mathcal{D}_T) \propto p(\theta|\mathcal{D}_{T-1}) p(\mathcal{D}_T|\theta) \]

The intractability of posteriors is tackled by Variational Inference by

\[ \log p(\mathcal{D}) = \mathbb{E}_{q(\theta)} [\log p(\mathcal{D}|\theta)] - KL[q(\theta)||p(\theta)] + KL[q(\theta)||p(\theta|\mathcal{D})] \]

Extension: Alternative Loss

- IWA-E Bound provides a tighter bound of marginal likelihood than ELBO.
- GVCL Bound uniforms VCL and EWCL, another continual learning algorithm.

Extension: VCL-GAN

VCL on GAN is tricky:
- It is difficult to balance the Bayesian-generator and discriminator;
- GAN loss is not a well-defined likelihood.

A successful VCL-GAN involves 3 key components:
1. Wasserstein GAN Loss as the "negative log-likelihood" term
2. Generalized-VCL Bound
3. Bayesian body with task-specific non-Bayesian heads

Ablation study:
- Standard GAN loss + (2) + (3)
- (1) + VCL ELBO + (3)
- (1) + (2) + Bayesian head

Conclusions
- VCL is a universal continual learning framework for both discriminative models and generative models;
- A more representative coreset tends to improve knowledge retention – Generative Replay combined with coreset provides a consistently better memory mechanism;
- Both IWA-E Bound and GVCL Bound present superior results to the original VCL;
- VCL was successfully applied to GAN. WGAN loss, GVCL Bound and non-Bayesian heads are key to fully functional VCL-GAN.

References