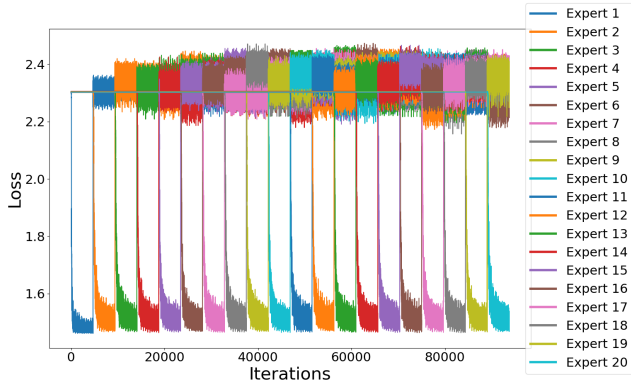# Supplementary Material

Table 1. Hyperparameter search for the regularization-based techniques
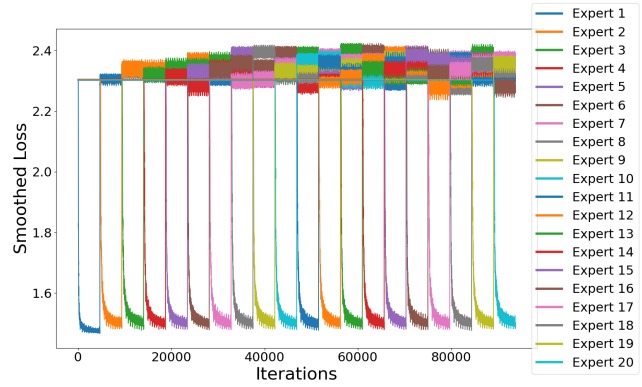
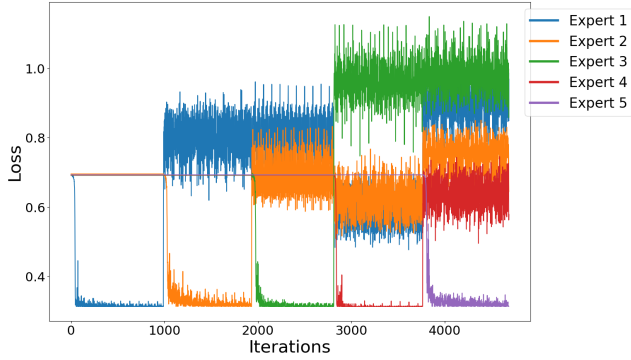| Name	Permuted MNIST	Split MNIST	Split CIFAR-100 (20)
EWC (λ: memory strength)	$\{10, 20, 50, 100\}$	$\{10^4, 10^5, 10^6, 10^7\}$	$\{10^4, 10^5, 10^6, 10^7\}$
SI (c: dimensionless strength)	$\{0.01, 0.1, 1, 10\}$	$\{0.01, 0.1, 1, 100\}$	$\{0.01, 0.1, 1, 10\}$
RWALK (λ: regularization term)	$\{0.01, 0.1, 1, 100\}$	$\{0.01, 0.1, 1, 10\}$	$\{10^1, 10^2, 10^3, 10^4\}$

Table 2. Hyperparameter settings used for A-GEM

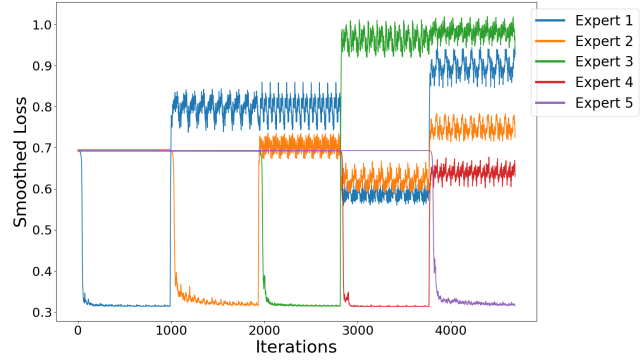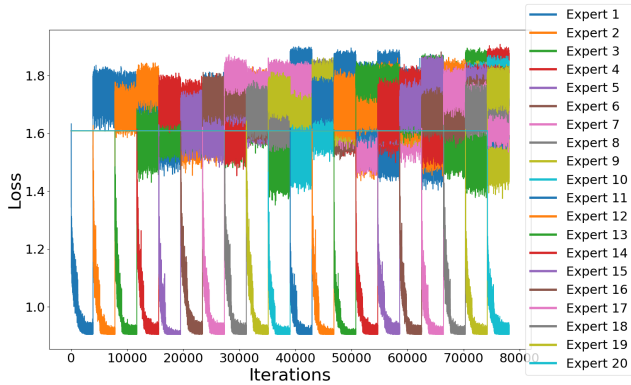| A-GEM	Permuted MNIST	Split MNIST	Split CIFAR-100 (20)
Episodic memory size	256	256	512
Episodic batch size	256	256	1300

(a) Loss, Permuted MNIST
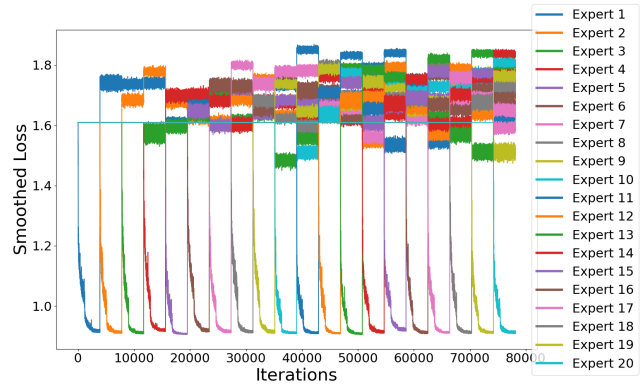
(b) Smoothed loss, Permuted MNIST

(c) Loss, Split MNIST

(d) Smoothed loss, Split MNIST

(e) Loss, Split CIFAR-100 (20)

(f) Smoothed loss, Split CIFAR-100 (20)

Figure 1. The value of loss function of different task expert networks (left) and its smoothed versions (right) during training for various data sets. Note that the experts are added sequentially as new tasks arrive. In all plots there is a clearly visible loss increase when a new task arrives