Motivation

We are interested in learning features that keep stationary while learning novel classes incrementally.

Practical advantages:
- Features can be used interchangeably in time,
- Visual search systems avoid re-computing features in the gallery when updating the model.

Along a similar vein, [Yantao et al. CVPR2020] introduces feature back-compatibility.

How to fix the Classifier?

Classifier values are taken from the coordinate vertices of Regular Polytopes. [Pernici et al. arXiv 2019]
- High dimensional Platonic Solids
- In 2D, regular n-sided polygon

Fixed and learnable classifiers have shown to achieve the same classification accuracy [Hoffer et al. ICLR2018]

Class Pre-allocation

Pre-allocation of the output nodes of future unseen classes allows to see negative samples since the beginning of learning.
- The space of unseen classes is not occupied by the seen ones.

As no prior assumption about the semantic similarity between future classes can be made, the natural assumption is to consider the d-Simplex fixed classifier
- all classes are nearest to all others (i.e., same cosine distance).

Experiments

MNIST (LeNet++ architecture, 2D feature dimension)

References

[Yantao et al. CVPR2020] Yantao Shen, Yuanjun Xiong, Wei Xie, and Stefano Soatto - Towards Backward-Compatible Representation Learning, CVPR 2020
[Hoffer et al. ICLR2018] Eldad Hoffer, Hay Hubara, Daniel Soudry - Fix your classifier: the marginal value of training the last weight layer - ICLR 2018

Class incremental Learning and Feature Representation

Class incremental Learning: training a single model on a sequence of disjoint classification problems without forgetting how to solve the previous ones.

Even assuming no catastrophic forgetting, internal feature representation changes as new classes are incorporated into the learning model.

We propose a pre-allocated fixed classifier (i.e., not undergoing learning). This keeps the features in a constant specific spatial configuration as novel classes are incorporated into the learning model.

Class Pre-allocation: training a single model on a sequence of disjoint classification problems without forgetting how to solve the previous ones.