

Semi supervised Learning Challenge

MI203 - Apprentissage automatique

Gianni FRANCHI
ENSTA-Paris

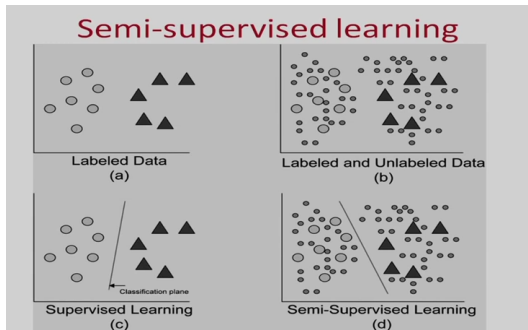


"Lately it seems like nothing but zeroes."



What is semi-supervised learning?

For supervised learning, you have a huge dataset with a lot of annotations. For semi-supervised learning, you have a set with annotation and another one without annotation.



Why do we have these two sets?

Annotations are expensive and consume a lot of time. Hence, semi-supervised learning aims to be able to train a model without a lot of annotation.

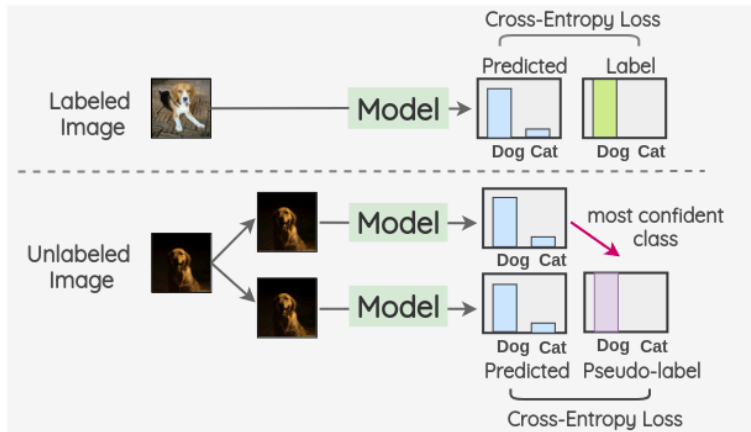
The Challenge

- To solve the challenge, you can explore traditional methods like SVM, random forest.
- You can use more sophisticated methods such as Deep Neural Network. If you do so, please be careful. Training from scratch might not be clever. Maybe try to finetune a Deep Neural Network.
- You can explore semi-supervised learning training procedure. **[Sohn2020]**

Questions

- What results do you get with classical Machine Learning algorithms? (Play with the parameters!)
- What are your conclusions (ML vs DL discussions)
- What results do you get with Fixmatch?

Context [Sohn2020]



CIFAR 10 [Cifar10]

The CIFAR-10 dataset consists of 60000 32×32 images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images.

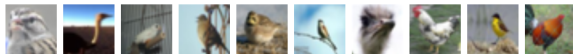
airplane



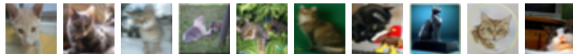
automobile



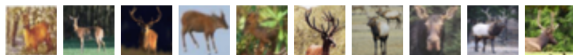
bird



cat



deer



dog



The challenge

- You will select 1000 images from the initial training set; this will be your new training set.
- You will select 2000 images from the initial training set to build the validation set.
- The rest of the data composed an unlabeled set that you can use as you want. But you consider that you don't have the annotation!
- Your goal is to train a model to classify the images such that you have the best accuracy on the test set.

Challenge's link:

<https://www.kaggle.com/t/2168fa6e53d44af6907589d3acaf790a>

Bibliography



[Sohn2020] Sohn, Kihyuk, et al. "Fixmatch: Simplifying semi-supervised learning with consistency and confidence." arXiv preprint arXiv:2001.07685 (2020).



[Cifar10] Learning Multiple Layers of Features from Tiny Images, Alex Krizhevsky, 2009.