Semi-Supervised Learning

AI Shot #4 @ VISUM

Tiago Freitas
Data Scientist at nilg.ai
tiago.freitas@nilg.ai

01

The Learning Spectrum

AI shot

02

Embedding Domain Knowledge

AI shot

03

Diving into an AI use case

04

Semi-Supervised Learning

AI shot

01

What is Semi-Supervised Learning?

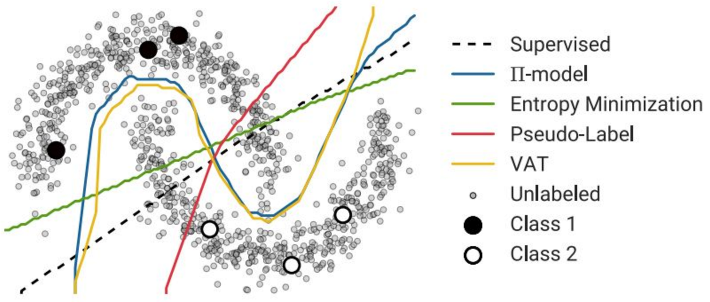
Semi-Supervised Learning

What is it and when to use it?

- Semi-Supervised learning tries to make use of **unlabeled data to learn its inherent structure** and empower a supervised model

- Useful for when you have a **big amounts of unlabeled data** from the **same distribution** of the labeled dataset
    - Not all unlabeled data is good for your models
    - If the distribution between labeled and unlabeled dataset is not similar, performance increments are smaller

- We need to be careful with unbalanced datasets



Oliver, Avital, Augustus Odena, Colin A. Raffel, Ekin Dogus Cubuk, and Ian Goodfellow. "Realistic evaluation of deep semi-supervised learning algorithms." In *Advances in neural information processing systems*, pp. 3235-3246. 2018.

Types of Semi-Supervised Learning

Semi-Supervised Algorithms can be classified in:

- Entropy Minimization:
  - Maximize confidence in unlabeled predictions
- Consistency Regularization
  - Guarantee that the model is consistent in unlabeled sample augmentations
- Hybrid Models
  - Combination of the previous classes with regularization
  - Generally the most powerful models combine all of the classes

nilg.ai

Oliver, Avital, Augustus Odena, Colin A. Raffel, Ekin Dogus Cubuk, and Ian Goodfellow. "Realistic evaluation of deep semi-supervised learning algorithms." In *Advances in neural information processing systems*, pp. 3235-3246. 2018.

Good Evaluation of SSL algorithms

- Compare with baselines:
  - Supervised
  - Transfer Learning

- Keep the same model architecture when comparing different algorithms

- Do not use unrealistic large validation sets

- If possible try to evaluate different % of ratios of labeled and unlabeled samples

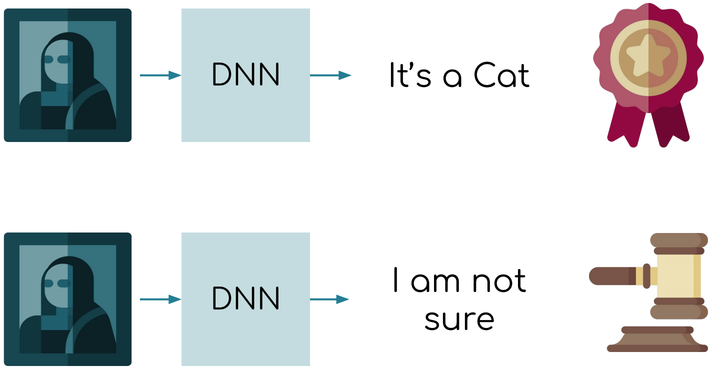
Oliver, Avital, Augustus Odena, Colin A. Raffel, Ekin Dogus Cubuk, and Ian Goodfellow. "Realistic evaluation of deep semi-supervised learning algorithms." In *Advances in neural information processing systems*, pp. 3235-3246. 2018.

02

Semi-Supervised Learning Algorithms

Entropy Minimization

- Loss term is added to loss function that encourages the model to **make “confident” (low-entropy) predictions** for all unlabeled examples, regardless of their class
- Tries to discourage the decision boundary from passing near data points



$$-\sum_{k=1}^{K} f_\theta(x)_k \log f_\theta(x)_k$$

Grandvalet, Yves and Bengio, Yoshua. Semi-supervised learning by entropy minimization. In Advances in Neural Information Processing Systems, 2005.

Pseudo-Labelling

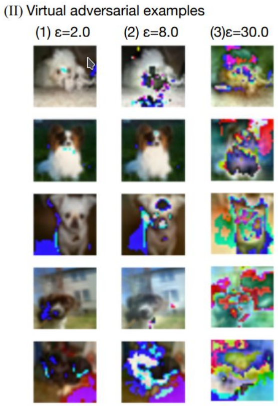
- Create Pseudo-Labels for unlabeled samples with a confidence above a certain threshold
- Targets above a certain threshold is used for training in the next epochs as if they were supervised labels
- Similar to Entropy Minimization as it enforces confident predictions for the unlabeled dataset, the difference is that it only uses the examples with low entropy (confident examples)

Lee, Dong-Hyun. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In ICML Workshop on Challenges in Representation Learning, 2013.

Virtual Adversarial Training (VAT)

- Train the model with adversarial examples on the unlabeled dataset
- Tiny perturbations on the image that lead to misclassification
- Enforces the model to be robust against adversarial perturbations
- Can be used in combination with Entropy Minimization to improve the results

(a) SVHN

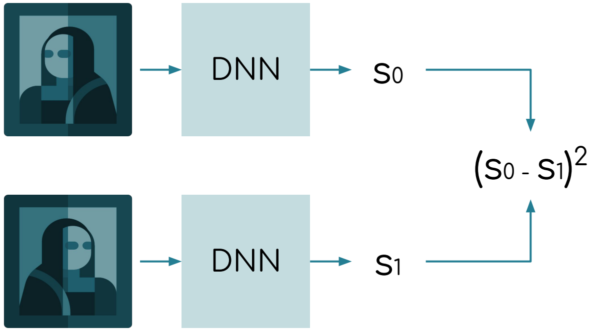


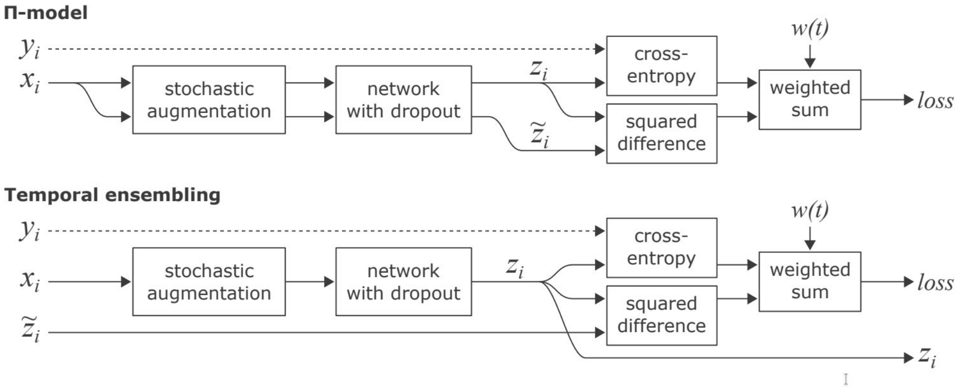
(b) CIFAR-10

Miyato, Takeru, Maeda, Shin-ichi, Koyama, Masanori, and Ishii, Shin. Virtual adversarial training: a regularization method for supervised and semi-supervised learning.

Pi-Model/Temporal Ensembling

- Augmentations on the unlabeled dataset are used to minimize the squared difference across samples
- Regularizes network by making it invariant to those augmentations
- Temporal ensembling of predictions is more stable than the simple mean squared difference of the current predictions

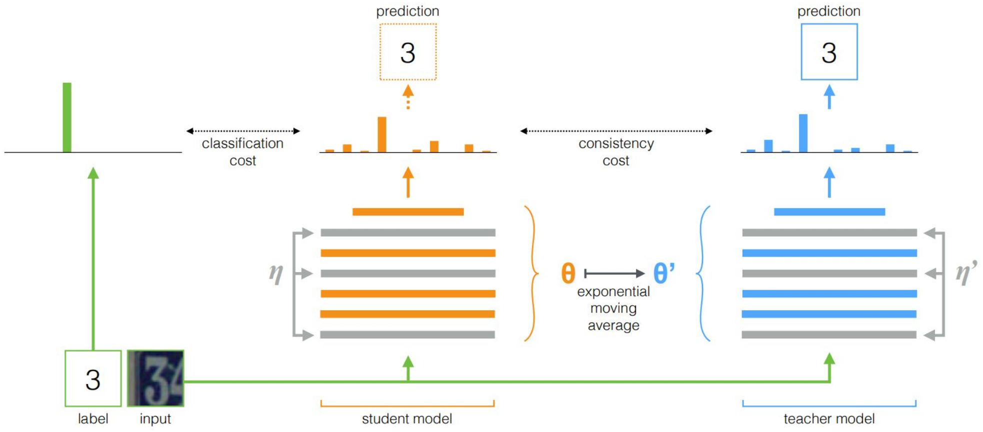




Laine, Samuli and Aila, Timo. Temporal ensembling for semi-supervised learning. In Fifth International Conference on Learning Representations, 2017.

Mean Teacher

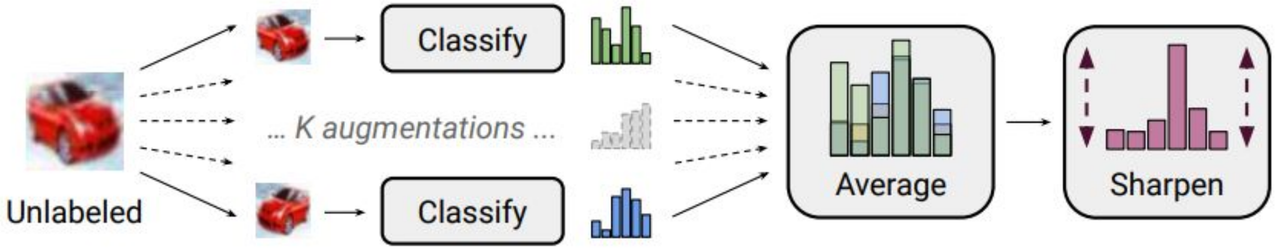
- Improves on the temporal ensembling model by using an exponential moving average of the weights of the model across epochs
- This leads to more stable predictions and models and improved performance



Tarvainen, Antti and Valpola, Harri. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in Neural Information Processing Systems, 2017.

MixMatch

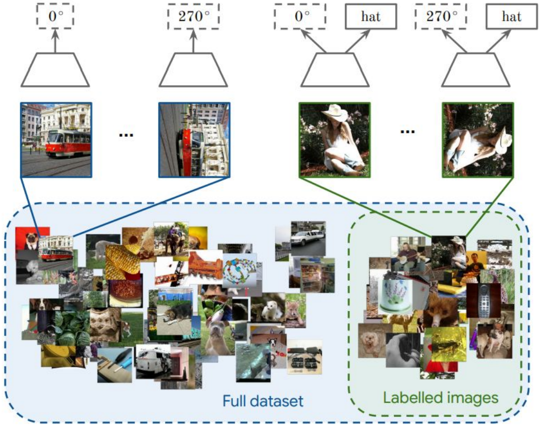
- New training schedule proposed:
  - For the unlabeled generate K number of augmentations
  - Optimizes the model using the mean of the k augmentations on the unlabeled dataset.
  - Apply Mixup augmentation to labeled and unlabeled images (convex combinations of labels and inputs)
  - Temperature Calibration of probabilities
  - Cross Entropy + L_2 norm loss
  - Weight Average is used as in mean teacher paper



Berthelot, David, et al. "Mixmatch: A holistic approach to semi-supervised learning." Advances in Neural Information Processing Systems. 2019.

S4L: Self-supervised semi-supervised learning

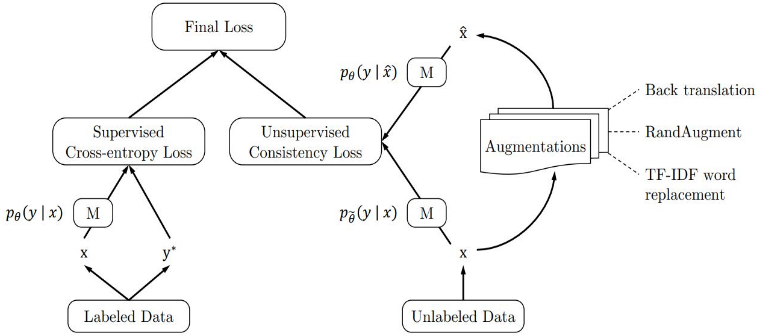
- It uses self-supervised learning pretraining on pretext tasks
- S4L Rotation: rotation prediction
- S4L Exemplar: Triplet Loss with soft margin within augmented unlabeled examples
- Using Self-Sup in simultaneously with semi-sup the results are improved:
  - S4L Rotation
  - Virtual Adv Training
  - Entropy Minimization
  - Model Exponential moving average of the Weights
  - Re-train with Pseudo Labels
  - Fine Tune on 10 % labels



Zhai, Xiaohua, et al. "S4l: Self-supervised semi-supervised learning." Proceedings of the IEEE international conference on computer vision. 2019.

Unsupervised Data Augmentation (UDA)

- KL divergence between original and augmented examples added to the supervised loss
- Better Performance with equal weight between label and unlabeled loss part, but with bigger unlabeled batches.
- AutoAugment + Cutout used for image augmentation
- Train Signal Annealing: If labeled examples are too confident they are removed from the loss. The threshold for this is varied across epochs.
- Unlabeled training only uses confident examples
- Also uses Probability calibration



Xie, Qizhe, et al. "Unsupervised data augmentation for consistency training." *arXiv preprint arXiv:1904.12848* (2019).

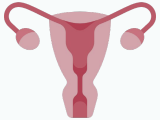
A lot of interesting papers recently

- FixMatch
- ReMixMatch
- Local Label Propagation for SSL
- Big Self-Supervised Models are Strong Semi-Supervised Learners
- Keep looking for the baselines in:
  https://paperswithcode.com/sota/semi-supervised-image-classification-on-2

03

NILG.AI Projects using Semi-Supervised Learning

NILG.AI Projects using SSL



Cervical Cancer

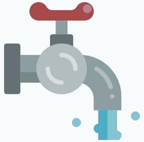
Diagnose patients, explain the reasons for this decision, suggest the right treatment (Computer Vision + Tabular Data)

- Healthcare (Colposcopy) -

Spam detection

Detect Spam in Online Questionnaires (NLP)

- Market Research -



Fraud detection

Detect fraud in water meters

- Utilities (Water supply) -

Thank you!
Questions?