Machine Learning in Healthcare

#C23 Semi-supervised learning

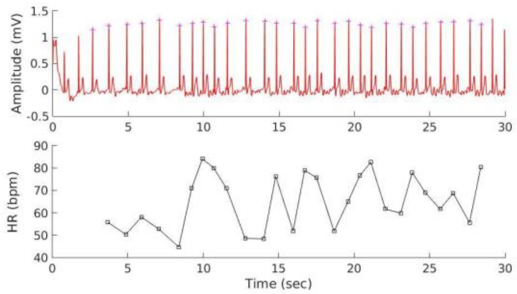
Technion-IIT, Haifa, Israel

Assist. Prof. Joachim Behar
Biomedical Engineering Faculty
Technion-IIT

AIMLab.

Supervised learning in the age of “big data”

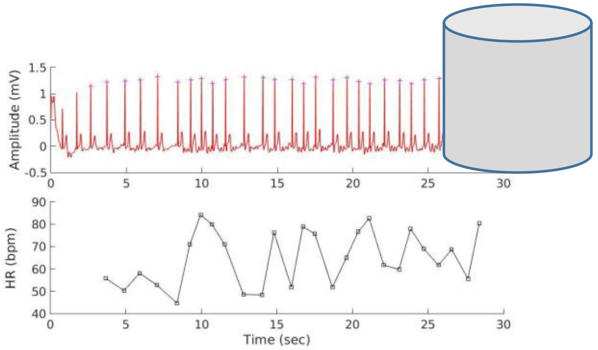
- In supervised learning we learn from labeled examples.
- Having a lot of examples is an opportunity to build a powerful model.
- But with great power comes…
- … a lot of labels to produce!
- Labeling examples is often expensive or even infeasible (and it can be fairly boring!).
  - E.g.:
    - Label 20 Holters ECG: 20 min x 20 = 400 min.
    - Label 2M Holters ECG: 2M x 20 = …



Supervised learning in the age of “big data”

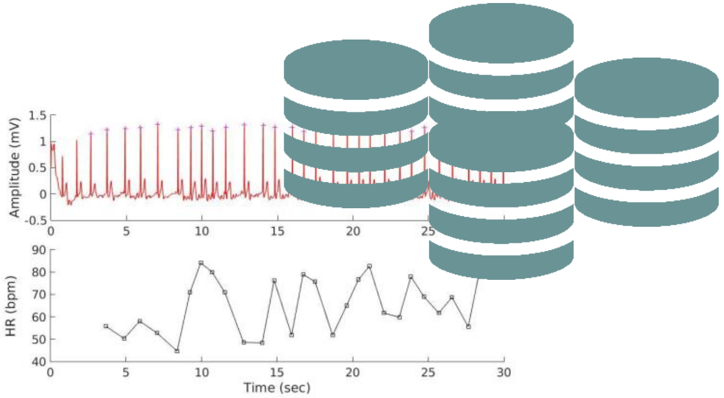
- How can we harness unlabeled examples to improve our classifier?
- In **semi-supervised learning**, we exploit both labeled and unlabeled examples.

Supervised



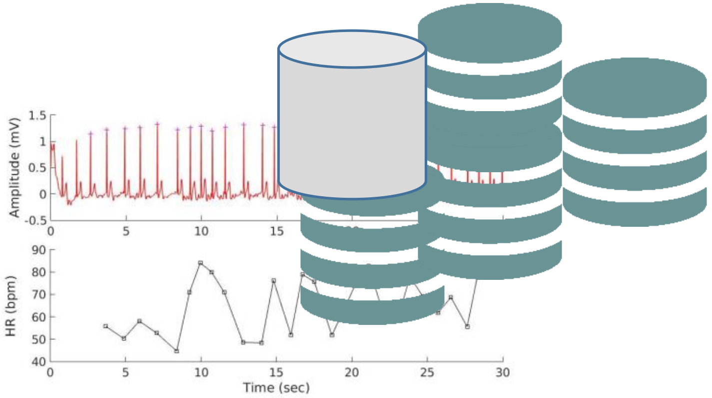
☑ Labels.

Unsupervised



☒ Labels.

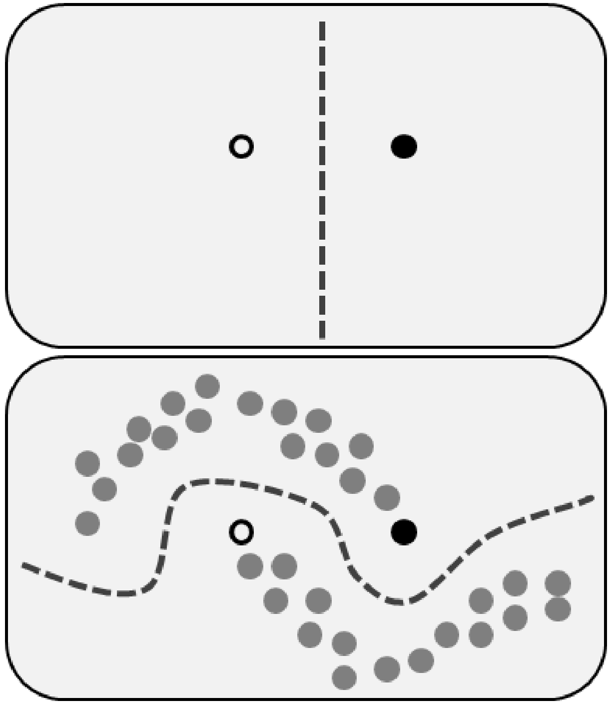
Semi-supervised



☑ ☒ Labels.

Supervised learning in the age of “big data”

- We will look at three flavors of semi-supervised learning
  - Self-training.
  - Generative model.
  - Anomaly detection.



https://en.wikipedia.org/wiki/Semi-supervised_learning#/media/File:Example_of_unlabeled_data_in_semisupervised_learning.png

Self-training, notations

- Labeled data: $D_l = \left\{x_l^{(i)}, y_l^{(i)}\right\}, i \in [\![1, n_l]\!]$
- Unlabeled data: $D_u = \left\{x_u^{(i)}, y_u^{(i)}\right\}, i \in [\![1, n_u]\!]$
- Test data: $D_t = \left\{x_t^{(i)}, y_t^{(i)}\right\}, i \in [\![1, n_t]\!]$
- Typically $n_l \ll n_u$ we want to make use of a lot of unlabeled data to improve the classifier performance.

Self-training

- Base option:
  - Train f from D_l
  - Predict for $x \in D_u$
  - Add $(x, f(x))$ to labeled data.
  - Repeat.
- Alternative flavors:
  - Select the k most confident $\left(x_u^{(i)}, f\left(x_u^{(i)}\right)\right)$ to labeled data.
  - Add all $\left(x_u^{(i)}, f\left(x_u^{(i)}\right)\right)$ to the labeled data but weight.

Self-training

- This is called self-training.
- Some advantages:
  - Very simple.
  - Wrapper like method- can use with any classifier.
- Disadvantage:
  - Early mistakes can reinforce themselves.

AIMLab.

Self-training - example

Require: Labeled images $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ and unlabeled images $\{\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_m\}$.

1: Learn teacher model θ_* which minimizes the cross entropy loss on labeled images

$$\frac{1}{n}\sum_{i=1}^{n} \ell(y_i, f^{noised}(x_i, \theta))$$

2: Use an unnoised teacher model to generate soft or hard pseudo labels for unlabeled images

$$\tilde{y}_i = f(\tilde{x}_i, \theta_*), \forall i = 1, \cdots, m$$

3: Learn student model θ'_* which minimizes the cross entropy loss on labeled images and unlabeled images with noise added to the student model

$$\frac{1}{n}\sum_{i=1}^{n} \ell(y_i, f^{noised}(x_i, \theta')) + \frac{1}{m}\sum_{i=1}^{m} \ell(\tilde{y}_i, f^{noised}(\tilde{x}_i, \theta'))$$

4: Iterative training: Use the student as a teacher and go back to step 2.

Algorithm 1: Noisy Student method

Self-training with Noisy Student improves ImageNet classification

Qizhe Xie*[1], Eduard Hovy[2], Minh-Thang Luong[1], Quoc V. Le[1]
[1]Google Research, Brain Team, [2]Carnegie Mellon University
{qizhex, thangluong, qvl}@google.com, hovy@cmu.edu

iv:1911.04252v1 [cs.LG] 11 Nov 2019

Abstract

We present a simple self-training method that achieves 87.4% top-1 accuracy on ImageNet, which is 1.0% better than the state-of-the-art model that requires 3.5B weakly labeled Instagram images. On robustness test sets, it improves ImageNet-A top-1 accuracy from 16.6% to 74.2%, reduces ImageNet-C mean corruption error from 45.7 to 31.2, and reduces ImageNet-P mean flip rate from 27.8 to 16.1.

To achieve this result, we first train an EfficientNet model on labeled ImageNet images and use it as a teacher to generate pseudo labels on 300M unlabeled images. We then train a larger EfficientNet as a student model on the combination of labeled and pseudo labeled images. We iterate this process by putting back the student as the teacher. During the generation of the pseudo labels, the teacher is not noised so that the pseudo labels are as good as possible. But during the learning of the student, we inject noise such as data augmentation, dropout, stochastic depth to the student so that the noised student is forced to learn harder from the pseudo labels.

1. Introduction

Deep learning has shown remarkable successes in image

beled images, 2) use the teacher to generate pseudo labels on unlabeled images, and 3) train a student model on the combination of labeled images and pseudo labeled images. Finally, we iterate the algorithm a few times by treating the student as a teacher to generate new pseudo labels and train a new student.

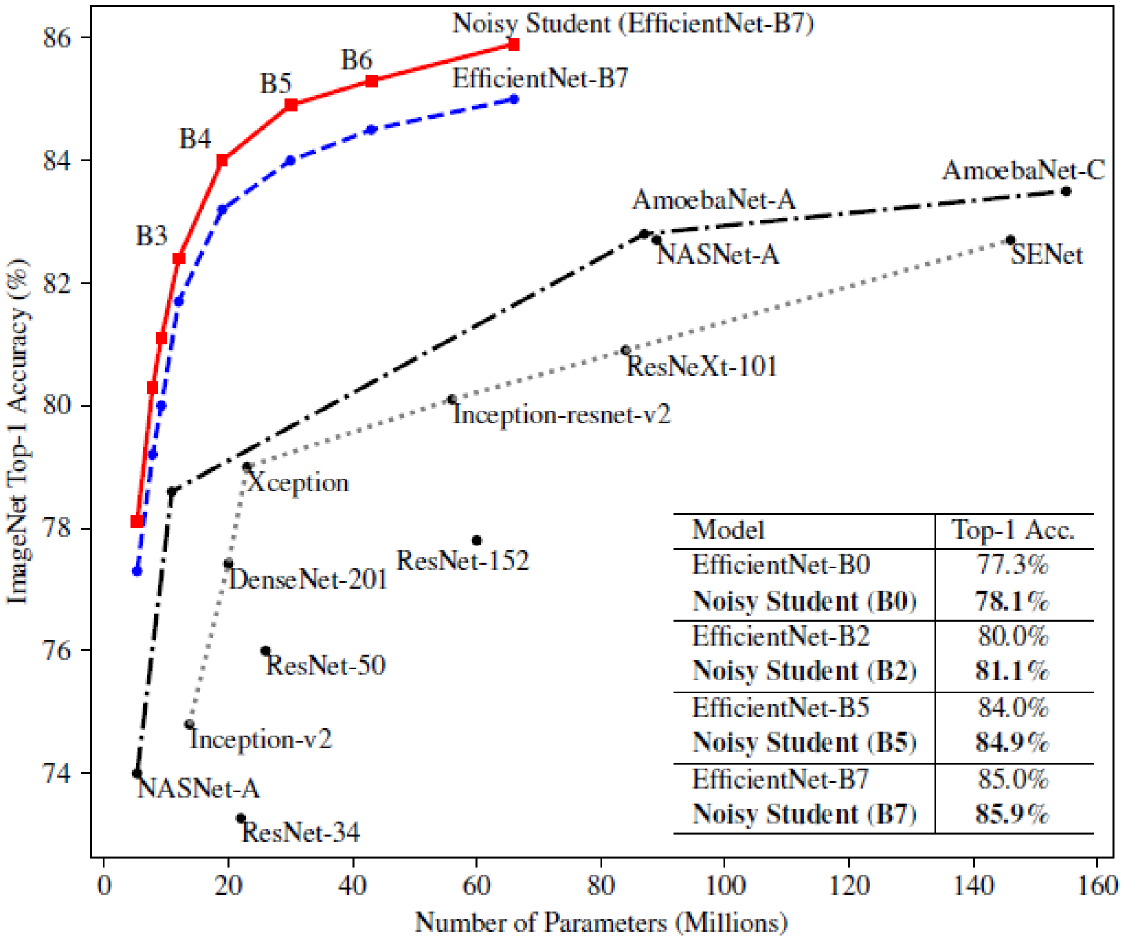
Our experiments show that an important element for this simple method to work well at scale is that the student model should be noised during its training while the teacher should not be noised during the generation of pseudo labels. This way, the pseudo labels are as good as possible, and the noised student is forced to learn harder from the pseudo labels. To noise the student, we use dropout [63], data augmentation [14] and stochastic depth [29] during its training. We call the method self-training with Noisy Student to emphasize the role that noise plays in the method and results. To achieve strong results on ImageNet, the student model also needs to be large, typically larger than common vision models, so that it can leverage a large number of unlabeled images.

Using self-training with Noisy Student, together with 300M unlabeled images, we improve EfficientNet's [69] ImageNet top-1 accuracy to 87.4%. This accuracy is 1.0% better than the previous state-of-the-art ImageNet accuracy which requires 3.5B weakly labeled Instagram images. Not only our method improves standard ImageNet accuracy, it

Xie, Qizhe, et al. "Self-training with Noisy Student improves ImageNet classification." arXiv preprint arXiv:1911.04252 (2019).

Self-training - example

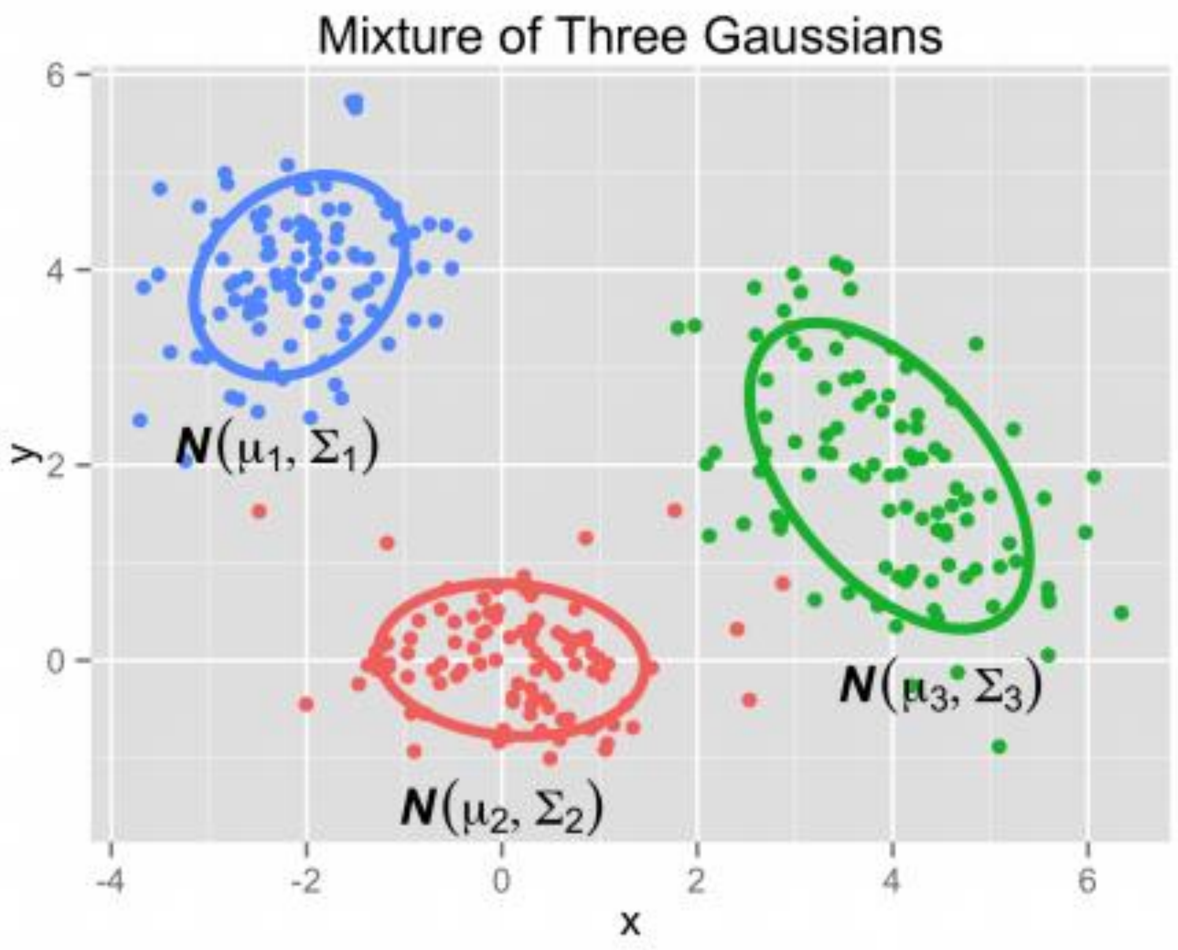
Method	# Params	Extra Data	Top-1 Acc.	Top-5 Acc.
ResNet-50 [23]	26M	-	76.0%	93.0%
ResNet-152 [23]	60M	-	77.8%	93.8%
DenseNet-264 [28]	34M	-	77.9%	93.9%
Inception-v3 [67]	24M	-	78.8%	94.4%
Xception [11]	23M	-	79.0%	94.5%
Inception-v4 [65]	48M	-	80.0%	95.0%
Inception-resnet-v2 [65]	56M	-	80.1%	95.1%
ResNeXt-101 [75]	84M	-	80.9%	95.6%
PolyNet [83]	92M	-	81.3%	95.8%
SENet [27]	146M	-	82.7%	96.2%
NASNet-A [86]	89M	-	82.7%	96.2%
AmoebaNet-A [54]	87M	-	82.8%	96.1%
PNASNet [39]	86M	-	82.9%	96.2%
AmoebaNet-C [13]	155M	-	83.5%	96.5%
GPipe [30]	557M	-	84.3%	97.0%
EfficientNet-B7 [69]	66M	-	85.0%	97.2%
EfficientNet-L2 [69]	480M	-	85.5%	97.5%
ResNet-50 Billion-scale [76]	26M	3.5B images labeled with tags	81.2%	96.0%
ResNeXt-101 Billion-scale [76]	193M		84.8%	-
ResNeXt-101 WSL [44]	829M		85.4%	97.6%
FixRes ResNeXt-101 WSL [71]	829M		86.4%	98.0%
Noisy Student (L2)	480M	300M unlabeled images	**87.4%**	**98.2%**



Xie, Qizhe, et al. "Self-training with Noisy Student improves ImageNet classification." arXiv preprint arXiv:1911.04252 (2019).

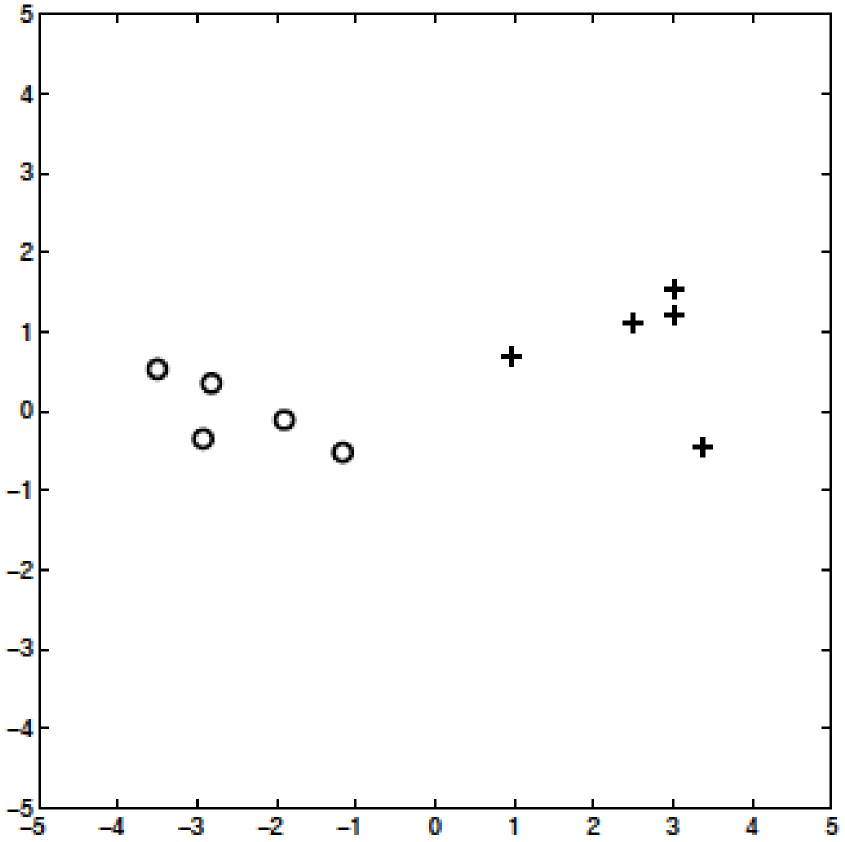
Generative model

- A generative model describes how data is generated, in terms of a probabilistic model.



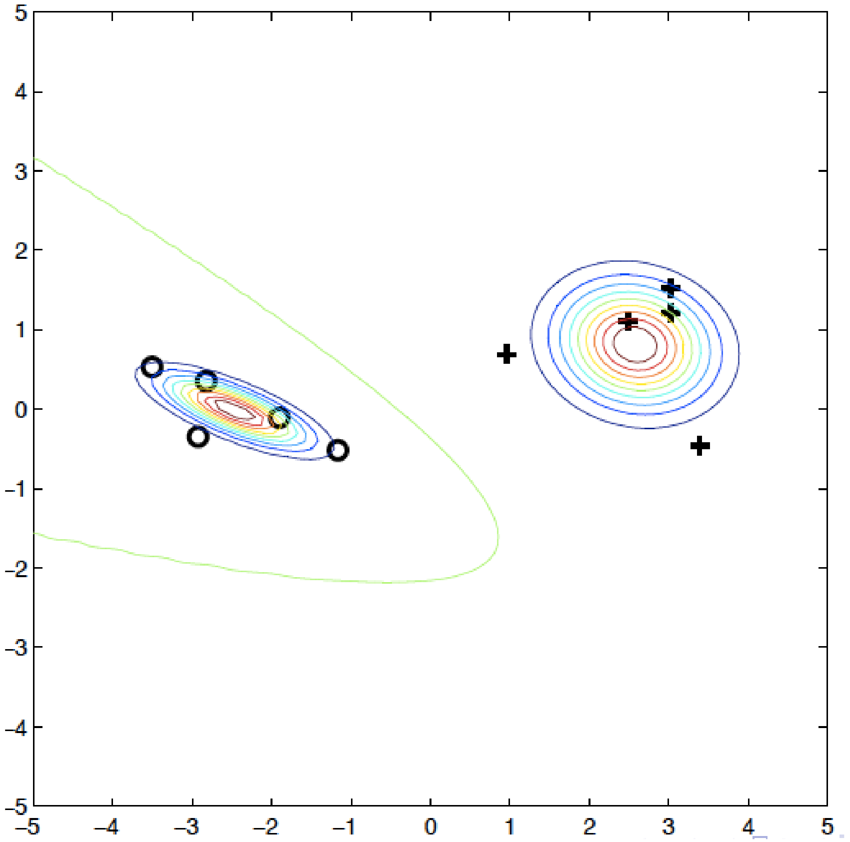
Generative model

- Given labeled data, assume each class has a Gaussian distribution.



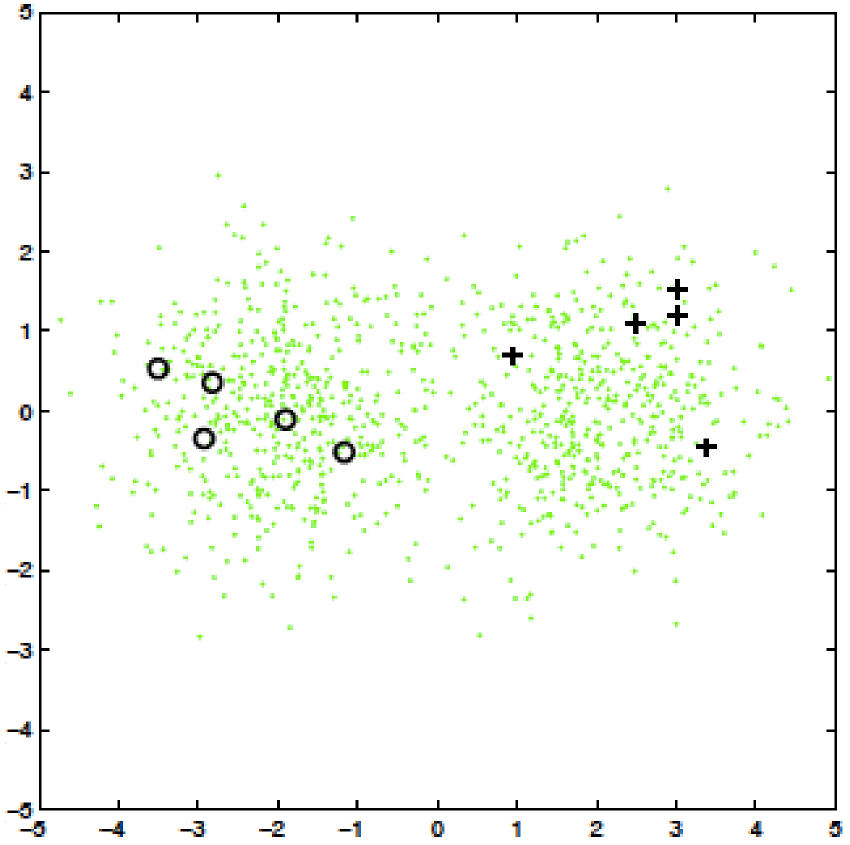
Generative model

- Given labeled data, assume each class has a Gaussian distribution.
- The most likely model and its decision boundary.



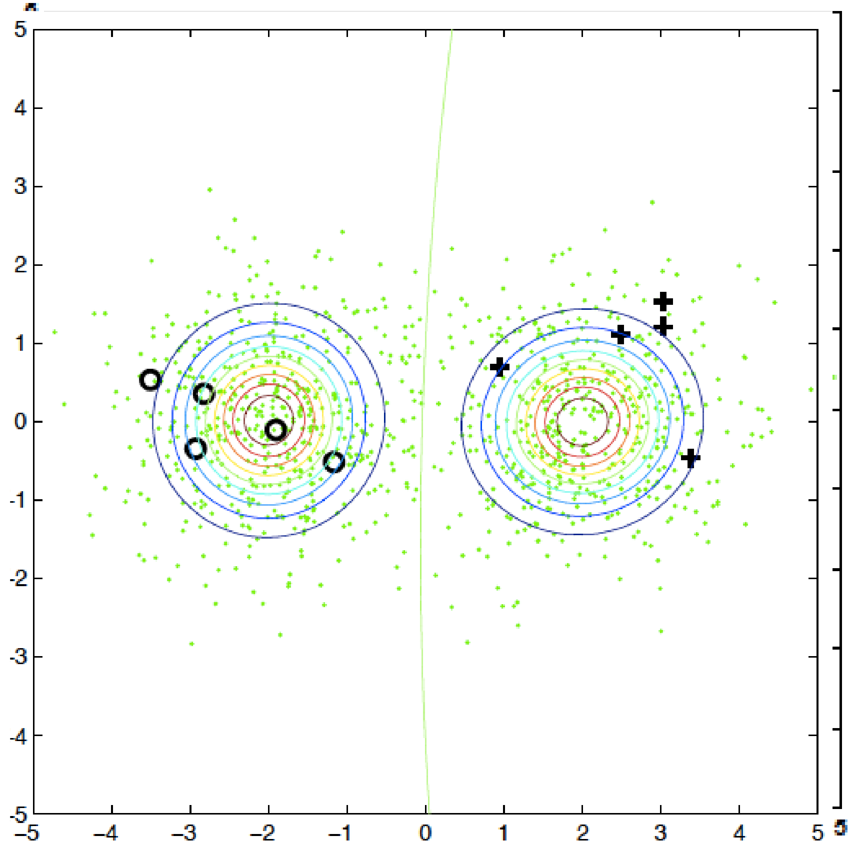
Generative model

- Given labeled data, assume each class has a Gaussian distribution.
- The most likely model and its decision boundary.
- Add unlabeled data.



Generative model

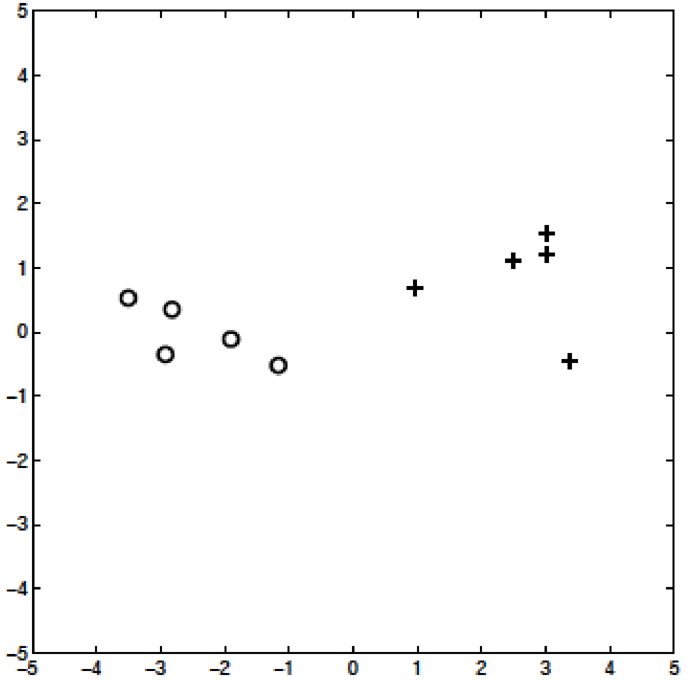
- Given labeled data, assume each class has a Gaussian distribution.
- The most likely model and its decision boundary.
- Add unlabeled data.
- The most likely model and decision boundary change.



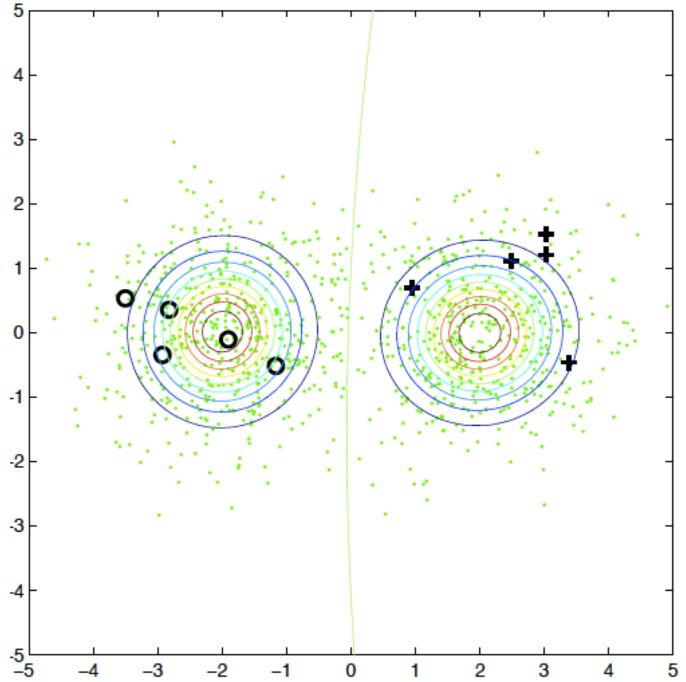
Generative model

- Decision boundaries are different because they maximize different quantities.

$P(X_l, Y_l|\omega)$



$P(X_l, Y_l, X_u|\omega)$



Generative model

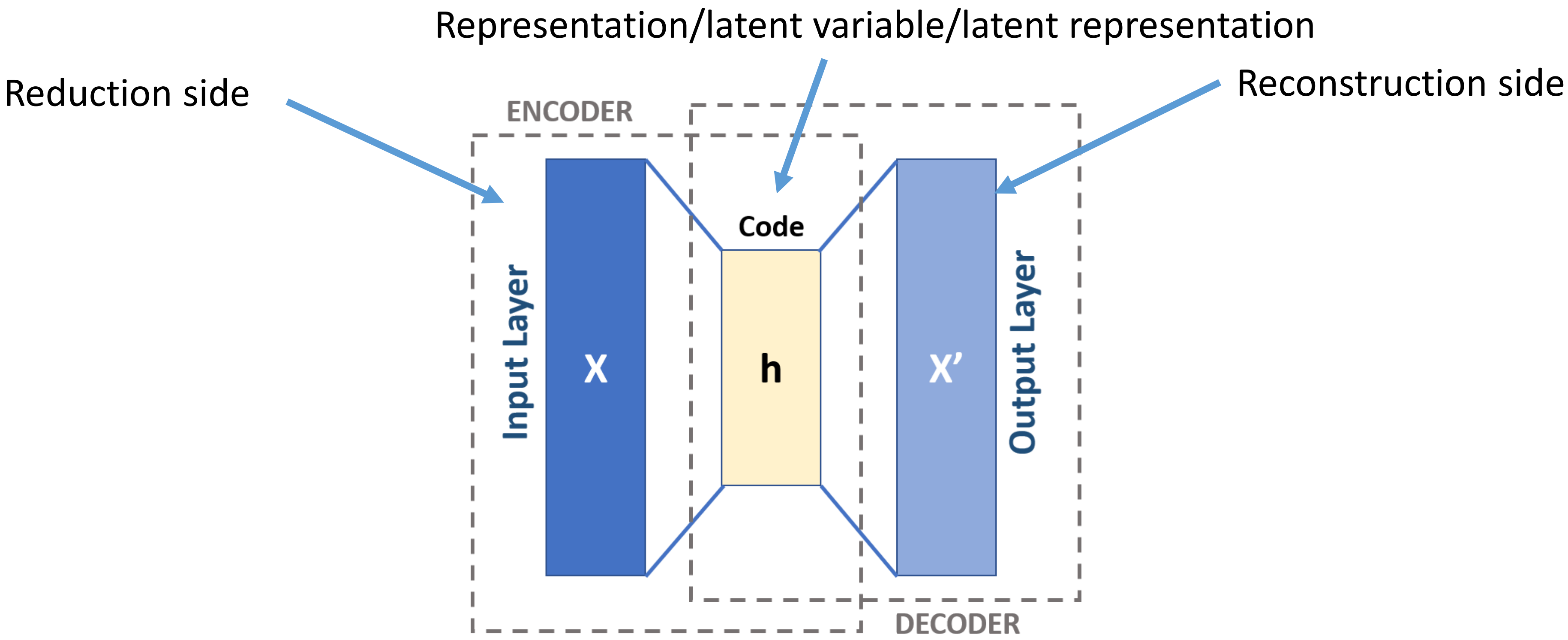
- This can offer a clear probabilistic framework.
- If the underlying probabilities model is close to correct then it can work very well.
- However, if the generative model is not suited then unlabeled data will not help.

Anomaly detection

- Identifying unexpected items or events in a dataset which differs from normal:
  - Often applied an unlabeled data.
  - Assumes that anomalies happen rarely.
  - Features of anomalies differ from the normal instances significantly.
- To illustrate this idea we will first introduce the notion of autoencoders.

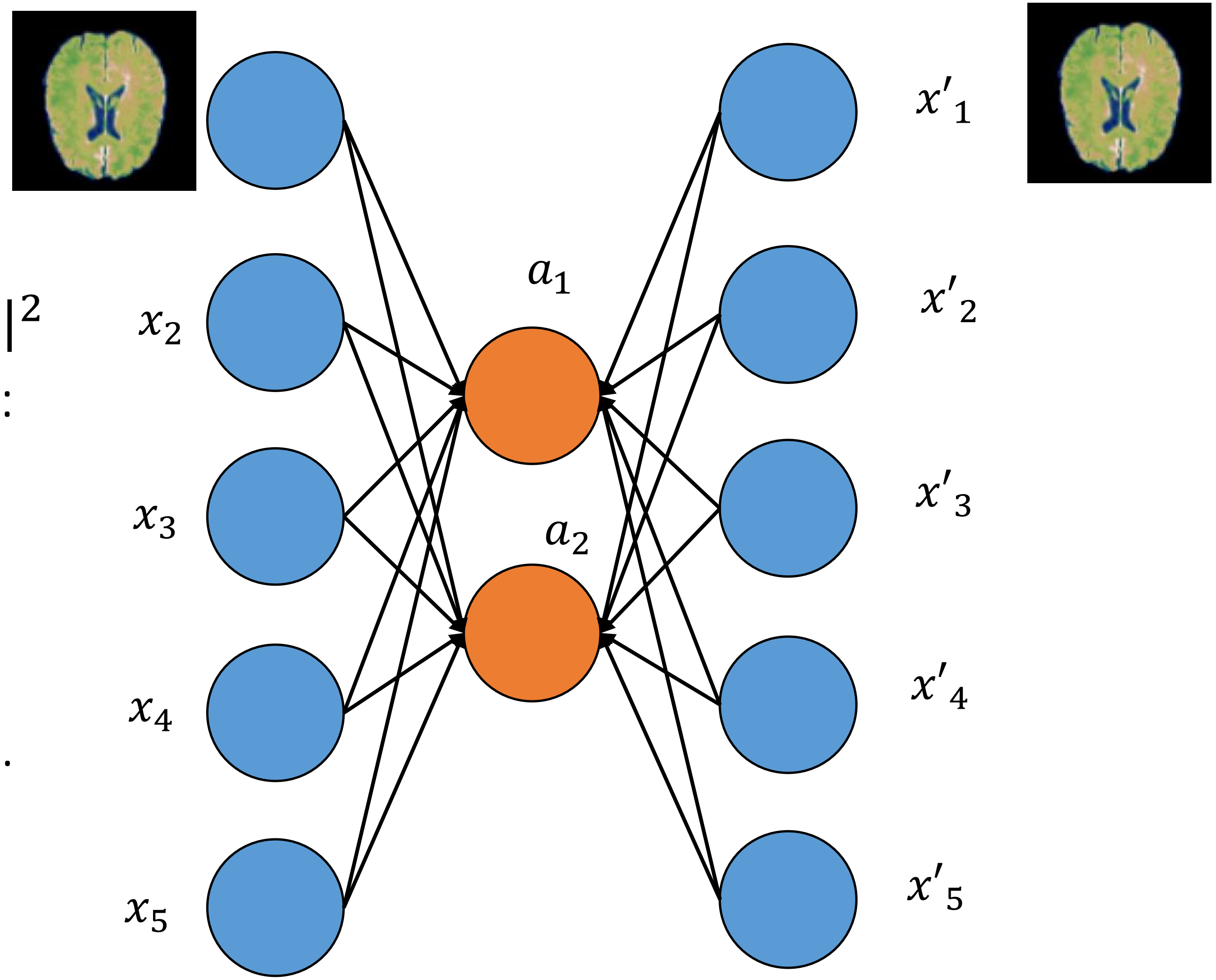
Anomaly detection

- **Autoencoders:** type of ANN used to learn a representation (i.e. encoding) for a set of data in an unsupervised manner.



Anomaly detection

- $\phi: \chi \to \mathcal{F}$
- $\varphi: \mathcal{F} \to \chi$
- $(\phi, \varphi) = argmin_{\phi,\varphi} \|X - (\varphi \cdot \phi)X\|^2$
- In the simple case a one hidden layer:
  - $x \in \mathbb{R}^d = \chi$
  - $h \in \mathbb{R}^d = \mathcal{F}$.
  - $h = g(\mathbf{W}\boldsymbol{x} + \boldsymbol{b})$
  - The image $\boldsymbol{h}$ is referred as latent variables or latent representation.



Anomaly detection

- How can we use Autoencoders for the purpose of anomaly detection?

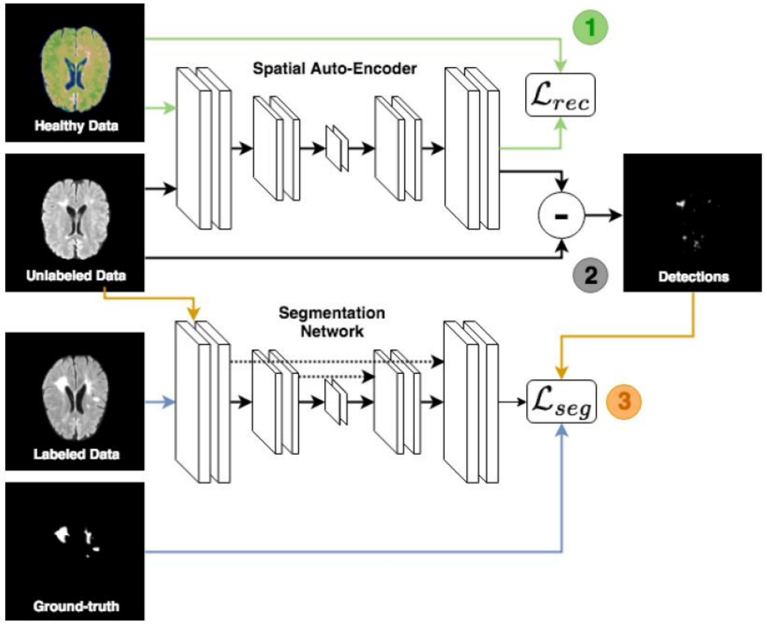


Figure 1: The proposed framework at a glance. Step 1: Training of a spatial AE on healthy data; Step 2: Inference on unlabeled data to obtain delineations; Step 3: Training of a supervised model from both labeled data with ground-truth and unlabeled data with UAD delineations.

Fusing Unsupervised and Supervised Deep Learning for White Matter Lesion Segmentation

Christoph Baur[1] C.BAUR@TUM.DE
Benedikt Wiestler[3]
Shadi Albarqouni[1]
Nassir Navab[1,2]
[1] *Computer Aided Medical Procedures (CAMP), TU Munich, Germany*
[2] *Whiting School of Engineering, Johns Hopkins University, Baltimore, United States*
[3] *Department of Diagnostic and Interventional Neuroradiology, Klinikum rechts der Isar, TU Munich, Germany*

Abstract

Unsupervised Deep Learning for Medical Image Analysis is increasingly gaining attention, since it relieves from the need for annotating training data. Recently, deep generative models and representation learning have lead to new, exciting ways for unsupervised detection and delineation of biomarkers in medical images, such as lesions in brain MR. Yet, Supervised Deep Learning methods usually still perform better in these tasks, due to an optimization for explicit objectives. We aim to combine the advantages of both worlds into a novel framework for learning from both labeled & unlabeled data, and validate our method on the challenging task of White Matter lesion segmentation in brain MR images. The proposed framework relies on modeling normality with deep representation learning for Unsupervised Anomaly Detection, which in turn provides optimization targets for training a supervised segmentation model from unlabeled data. In our experiments we successfully use the method in a Semi-supervised setting for tackling domain shift, a well known problem in MR image analysis, showing dramatically improved generalization. Additionally, our experiments reveal that in a completely Unsupervised setting, the proposed pipeline even outperforms the Deep Learning driven anomaly detection that provides the optimization targets.

Keywords: Deep Learning, Anomaly Detection, Unsupervised, Semi-Supervised, Supervised, White Matter Lesion Segmentation, Multiple Sclerosis

Baur, Christoph, et al. "Fusing Unsupervised and Supervised Deep Learning for White Matter Lesion Segmentation." International Conference on Medical Imaging with Deep Learning. 2019.

Take home

- Semi-supervised learning has become (and is-expending) a very active field of research due to the large number of unlabeled data.
- There are different ways to make use of these unlabeled data. We covered three:
  - Self-training.
  - Generative model.
  - Anomaly detection.

References

[1] COMP 551 – Applied Machine Learning Lecture 18: Semi-supervised learning. Joelle Pineau. https://www.cs.mcgill.ca/~jpineau/comp551/

[2] Xie, Qizhe, et al. "Self-training with Noisy Student improves ImageNet classification." arXiv preprint arXiv:1911.04252 (2019).

[3] Baur, Christoph, et al. "Fusing Unsupervised and Supervised Deep Learning for White Matter Lesion Segmentation." International Conference on Medical Imaging with Deep Learning. 2019.