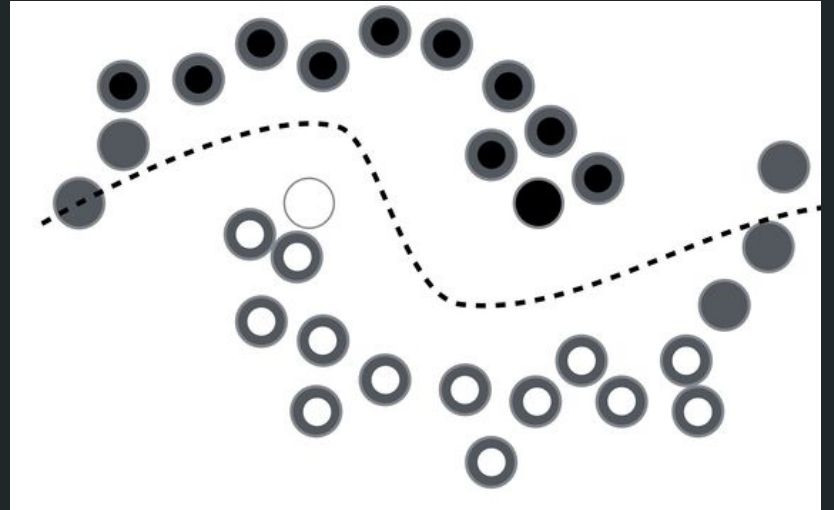


# Semi-Supervised Learning

---

Asmita, Natalie, Shikhar

# Introduction



# Motivation

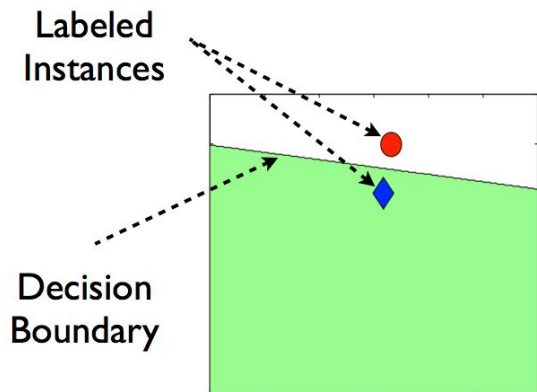
- Labelled data is **expensive** and **difficult to acquire** or generate
  - 100,000 images could cost you at least \$30,000 on Amazon Mechanical Turk
  - Manual labelling is tedious and time consuming
- **Small labelled datasets** might not be enough for developing **generalizable models**
- Unlabelled data is relatively **inexpensive**
- Semi-supervised learning is closer to **human way of learning**

# Semi-supervised learning

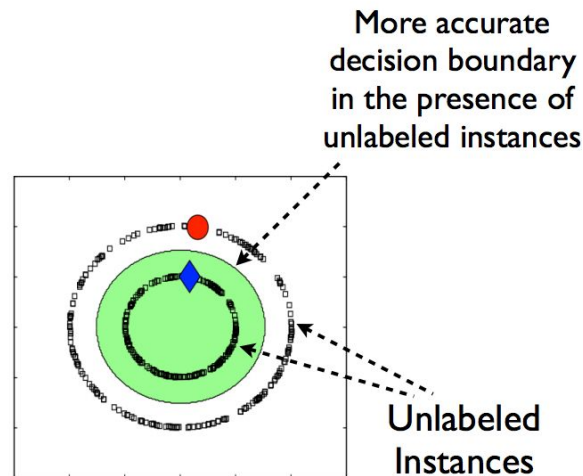
Class of techniques which uses unlabelled data along with labelled data for training

## Benefits

- Improves model performance by **more precise decision boundary** (**generalizable**) compared to just using labelled data
- Less money spent on labelling

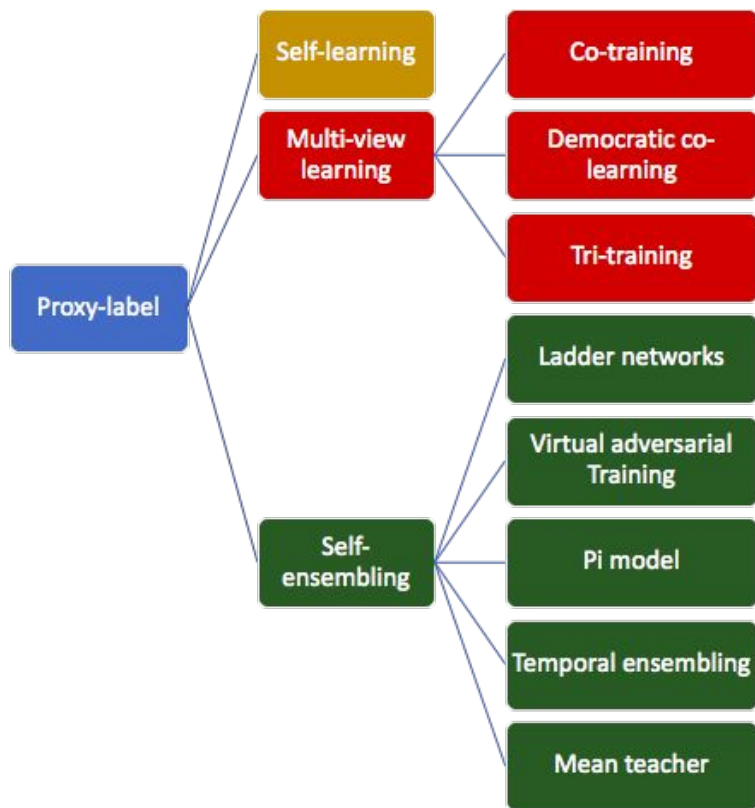


Supervised (2 labelled observations)



Semi-supervised (using unlabelled observations also)

# Types of semi-supervised learning (Proxy labels)



## Self-learning

Uses model's own predictions as proxy labels

## Multi-view learning

Uses the predictions of models trained with **different views of the data**

## Self-ensembling

Ensembles **variations of a model's own predictions** and uses these as feedback for learning

# How it works

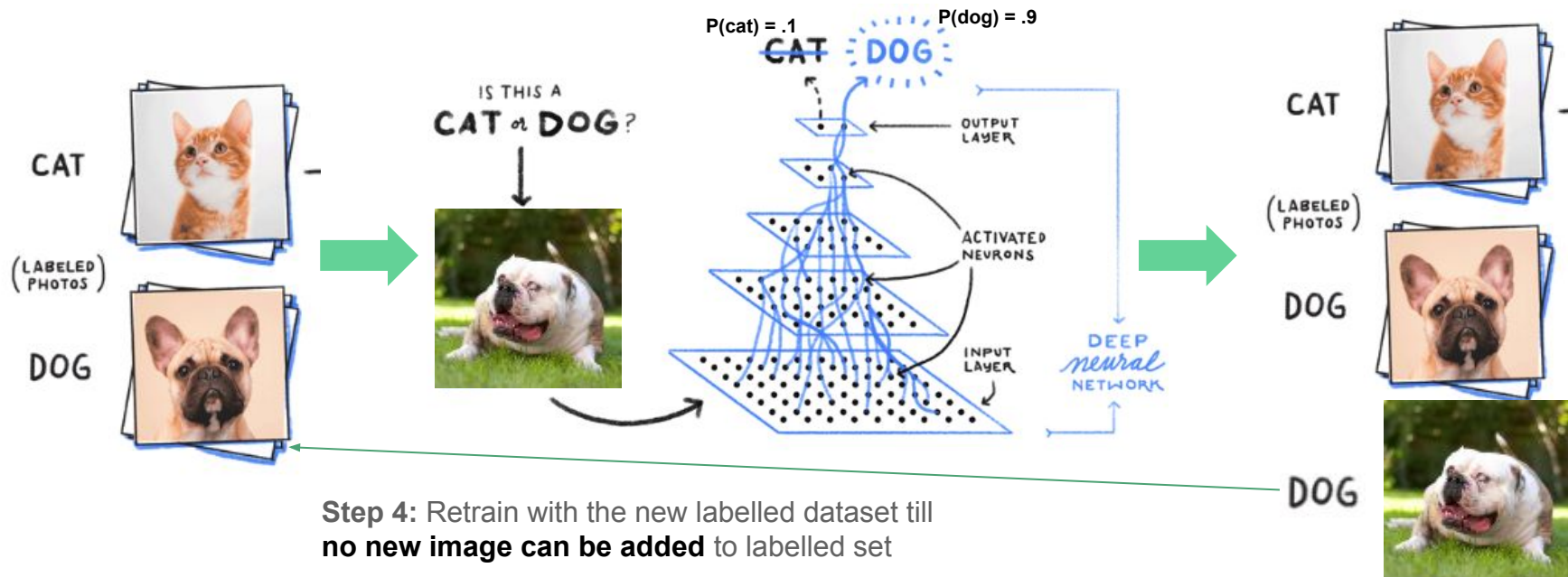
---

# Self-learning

**Step 1:** Train your model on labelled data

**Step 2:** Predict on unlabelled data using the trained model

**Step 3:** Add images from unlabelled data to labelled data **where model is highly confident ( $P > \text{threshold}$ )** with prediction as true label

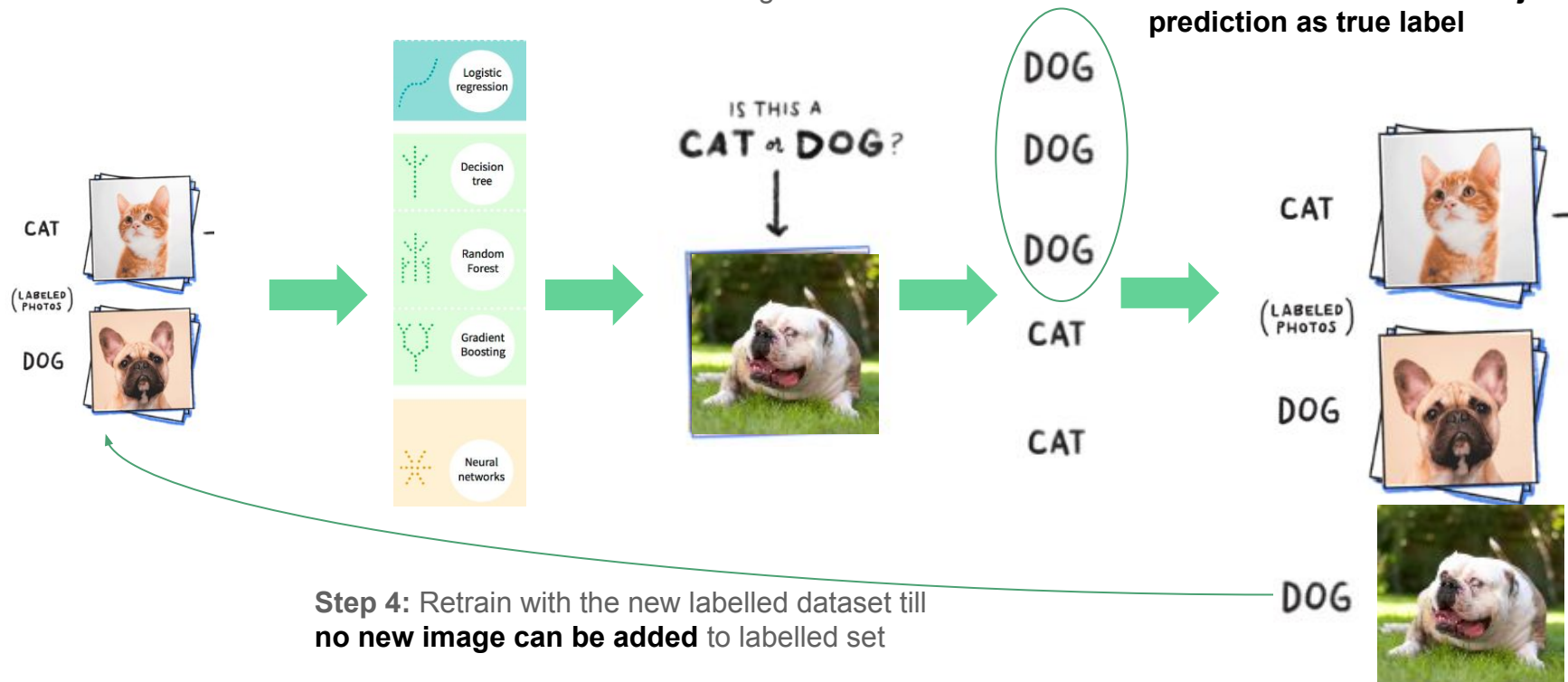


# Multi-view Learning: Democratic co-learning

**Step 1:** Train different models on the labelled dataset

**Step 2:** Predict on unlabelled data using the trained model

**Step 3:** Add images from unlabelled data to labelled data with **majority prediction as true label**



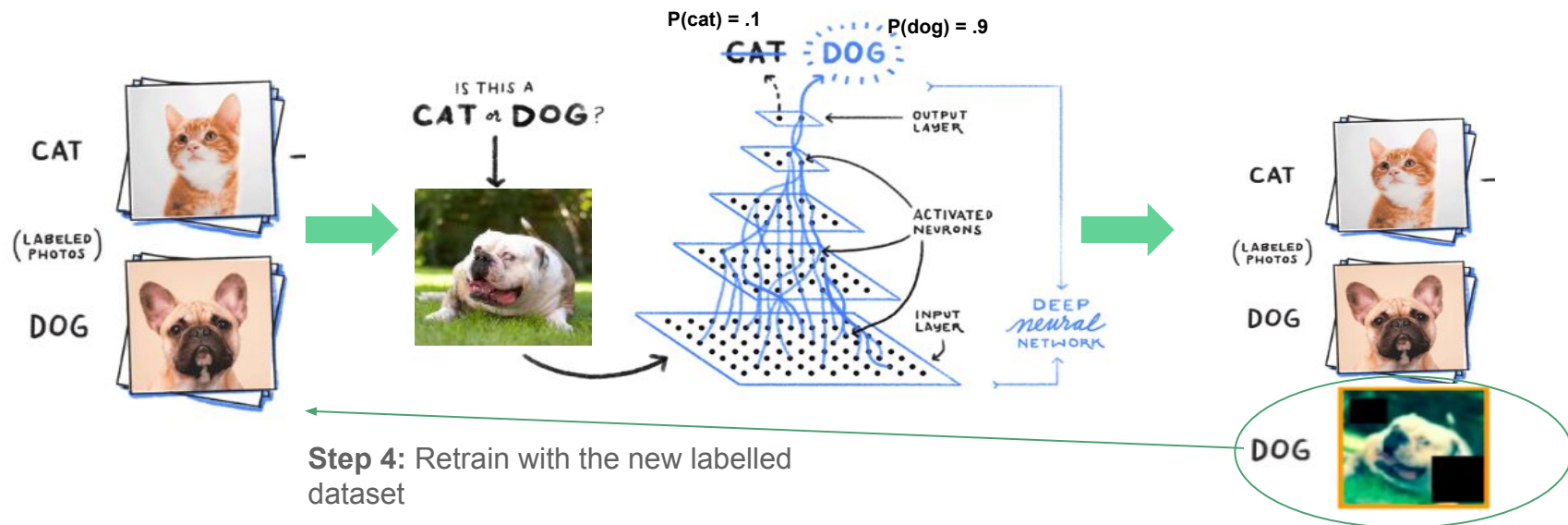


# Self-ensembling: Ladder networks

**Step 1:** Train your model on labelled data

**Step 2:** Predict on unlabelled data

**Step 3:** Add **perturbed version** of images from unlabelled data to labelled data with **prediction as true label**



The model learns to develop **features that are invariant to noise** and predictive of the labels on the labelled training data

# Demo

---

# Demo – Comparing supervised vs. semi-supervised

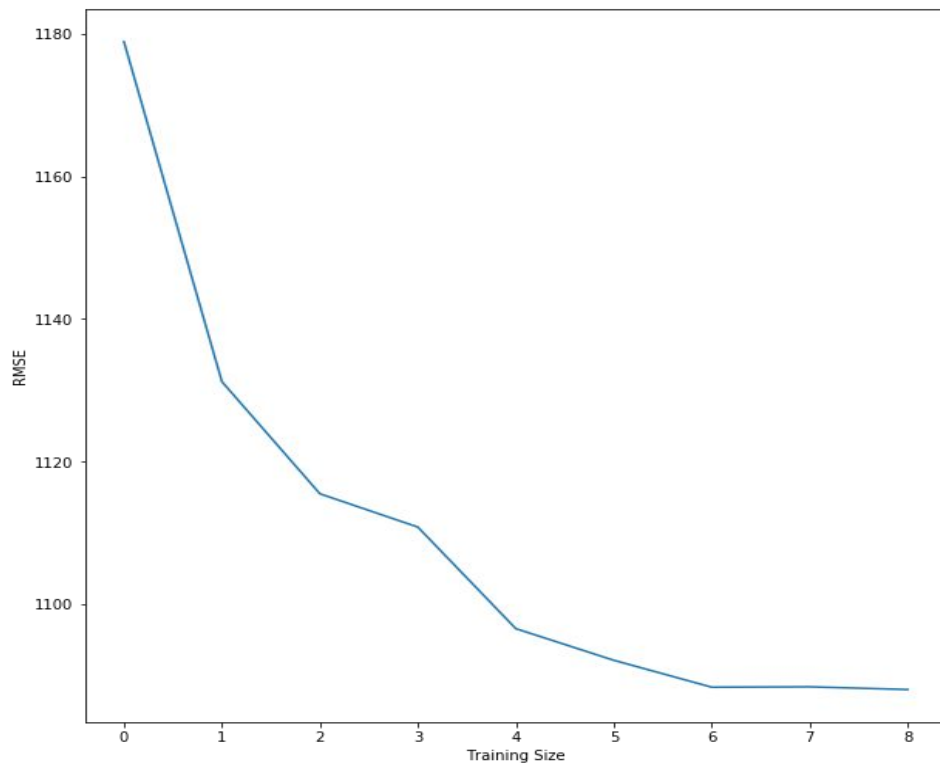
---

RandomForestRegressor	CV-8 MSE: 1193.0077 (+/- 117500.2586)
XGBRegressor	CV-8 MSE: 1083.2088 (+/- 122498.1182)
GradientBoostingRegressor	CV-8 MSE: 1085.1444 (+/- 124154.7107)
ExtraTreesRegressor	CV-8 MSE: 1212.6371 (+/- 98181.5808)
ElasticNet	CV-8 MSE: 1259.1340 (+/- 174781.8006)
KNeighborsRegressor	CV-8 MSE: 1230.1431 (+/- 172002.5180)
Ridge	CV-8 MSE: 1206.2309 (+/- 153934.3730)
BayesianRidge	CV-8 MSE: 1206.3157 (+/- 153576.2349)
PseudoLabeler	CV-8 MSE: 1082.6737 (+/- 128550.9096)

Here the `sample_rate` is 0.3.

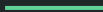
So each time we take 30% of the unlabeled data and put it through the training set

# Demo – Comparing supervised vs. semi-supervised



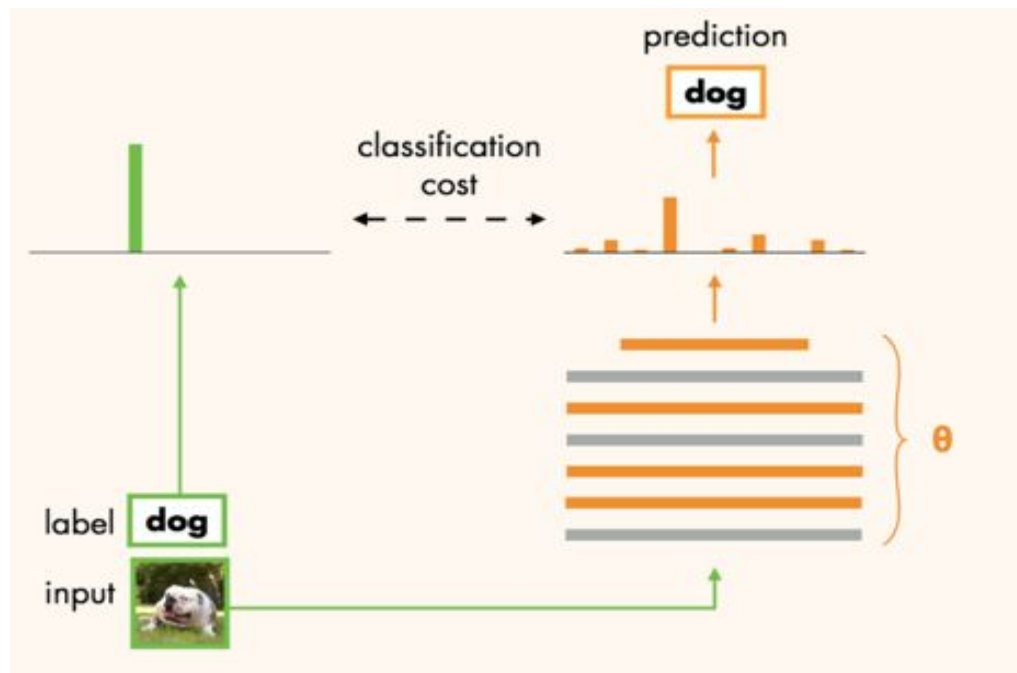
# Applications: The mean teacher model

<https://github.com/CuriousAI/mean-teacher>



# The mean teacher model

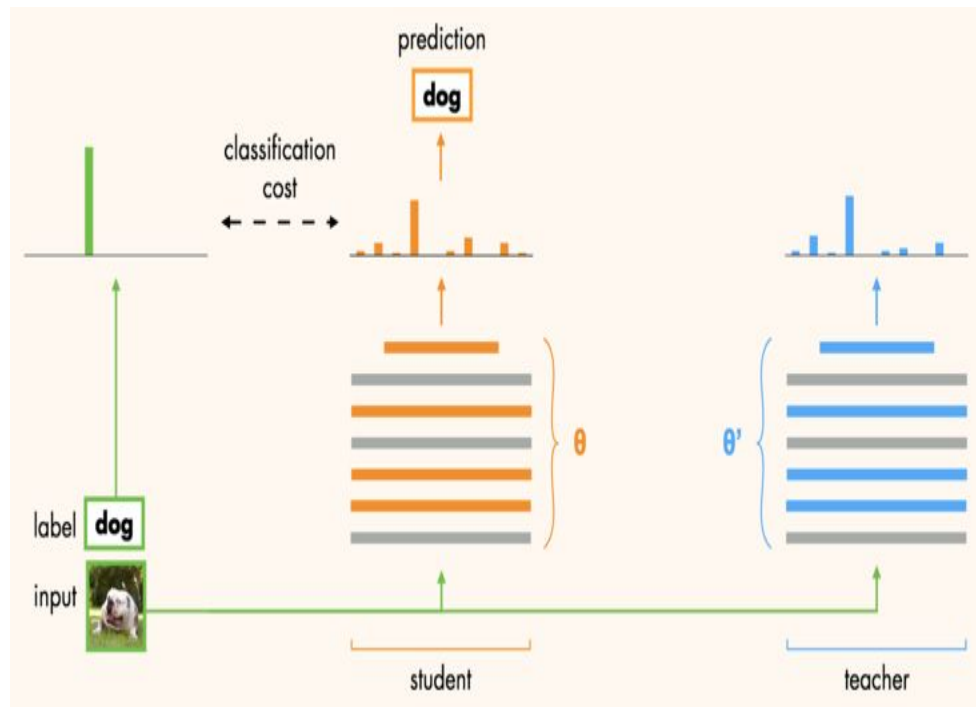
Step 1. Take a supervised model



# The mean teacher model

Step 1. Take a supervised model

Step 2. Make a copy of it

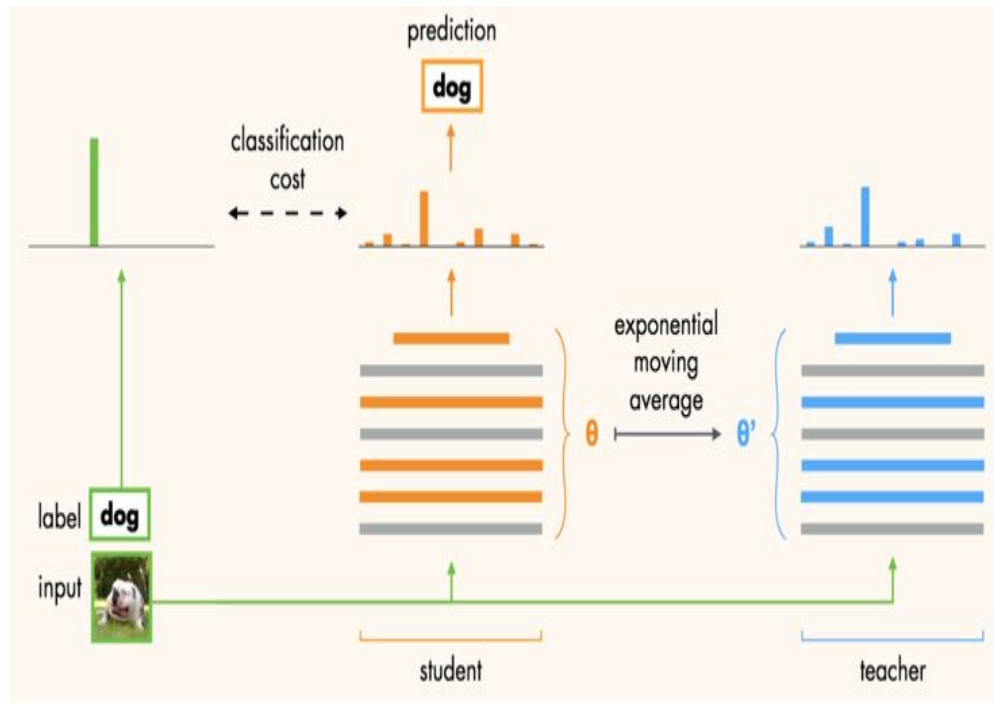


# The mean teacher model

Step 1. Take a supervised model - student

Step 2. Make a copy of it - teacher

Step 3. Update weights of the teacher after  
training each step





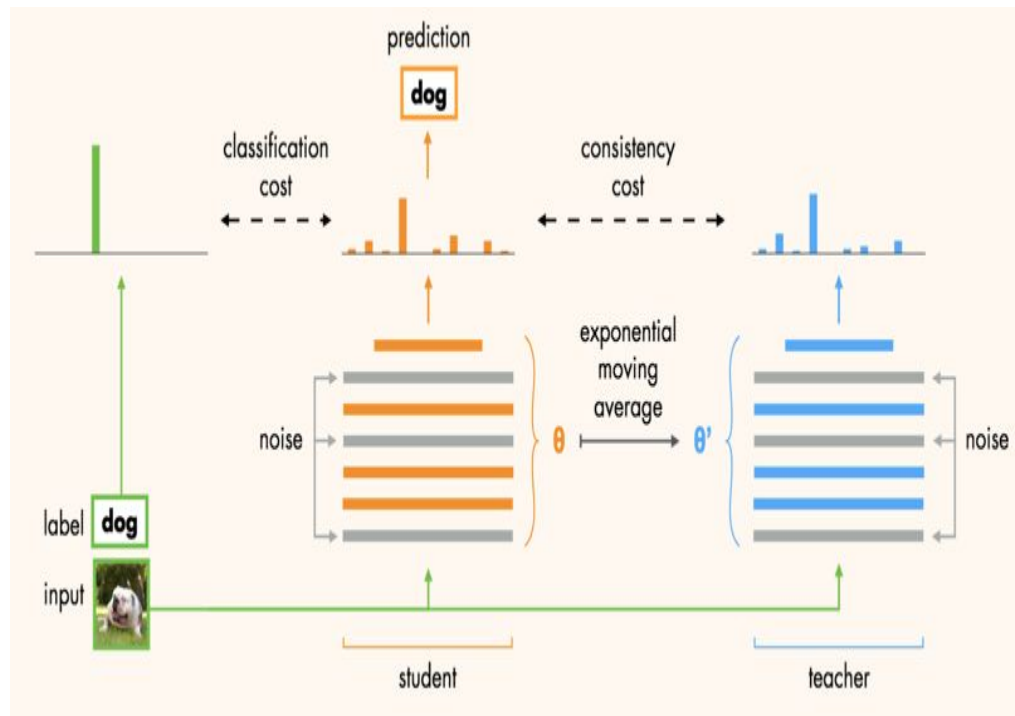
# The mean teacher model

Step 1. Take a supervised model - student

Step 2. Make a copy of it - teacher

Step 3. Update weights of the teacher after  
training each step

Step 4. Add some cost and noise to teacher



# The mean teacher model

Step 1. Take a supervised model - student

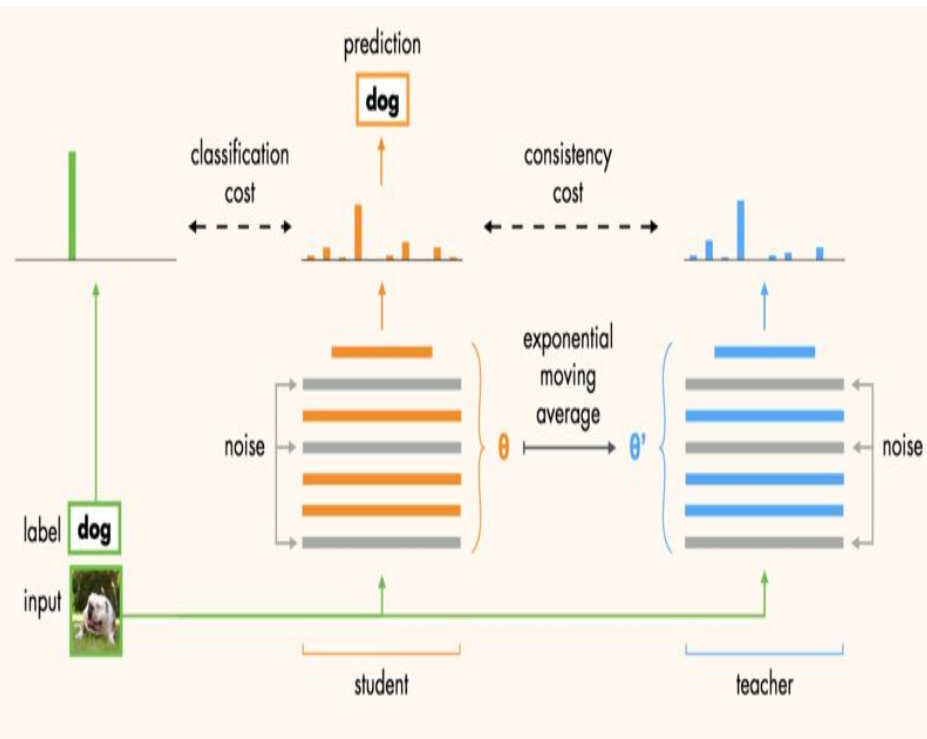
Step 2. Make a copy of it - teacher

Step 3. Update weights of the teacher after  
training each step

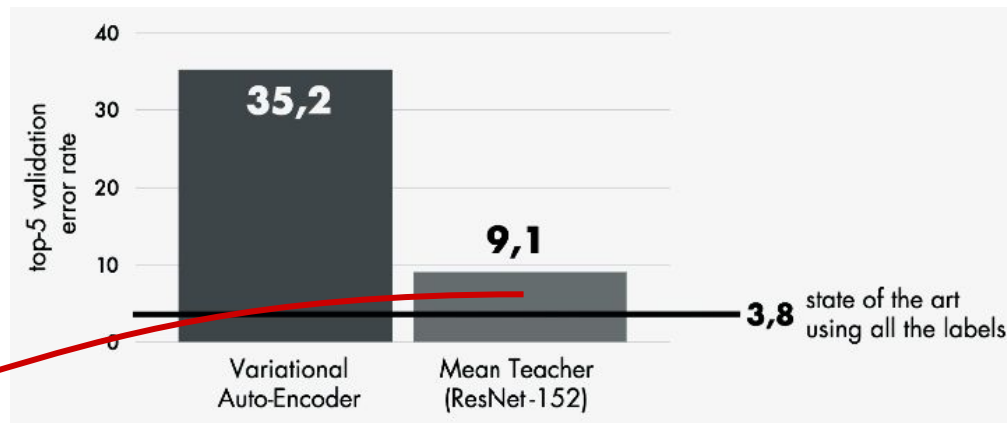
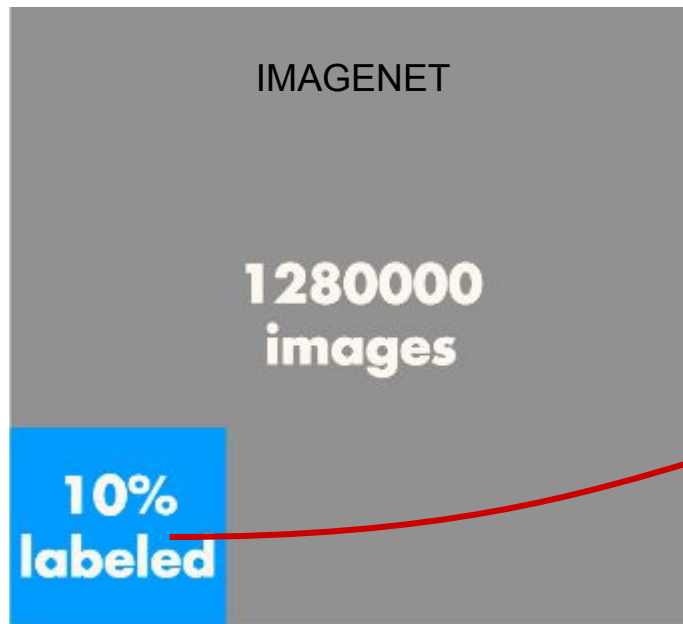
Step 4. Add some cost and noise to teacher

Step 5. Start using teacher as unlabelled dataset

For semi-supervised learning



# Mean teachers are better role models



Source: <https://github.com/CuriousAI/mean-teacher>

# Issues in Real Application

- Sensitive to amount of labeled and unlabeled data
  - More parameters to tune
- Performance can degrade substantially when unlabeled data contains a different distribution of classes than the labeled data
  - Say you are trying to train a model to distinguish between ten different faces, but you only have a few images for each of these ten faces.
- Realistically Small Validation Sets

# Resources

- <https://github.com/CuriousAI/mean-teacher>
- <http://runder.io/semi-supervised/>
- <https://www.analyticsvidhya.com/blog/2017/09/pseudo-labelling-semi-supervised-learning-technique/>
- [https://en.wikipedia.org/wiki/Semi-supervised\\_learning](https://en.wikipedia.org/wiki/Semi-supervised_learning)
- <https://www.kdnuggets.com/2018/01/value-semi-supervised-machine-learning.html>
- <https://www.google.co.in/about/stories/machine-learning-qa/>
- <https://www.robots.ox.ac.uk/~vgg/rg/papers/multimodallearning.pdf>
- <https://arxiv.org/pdf/1804.09170.pdf>