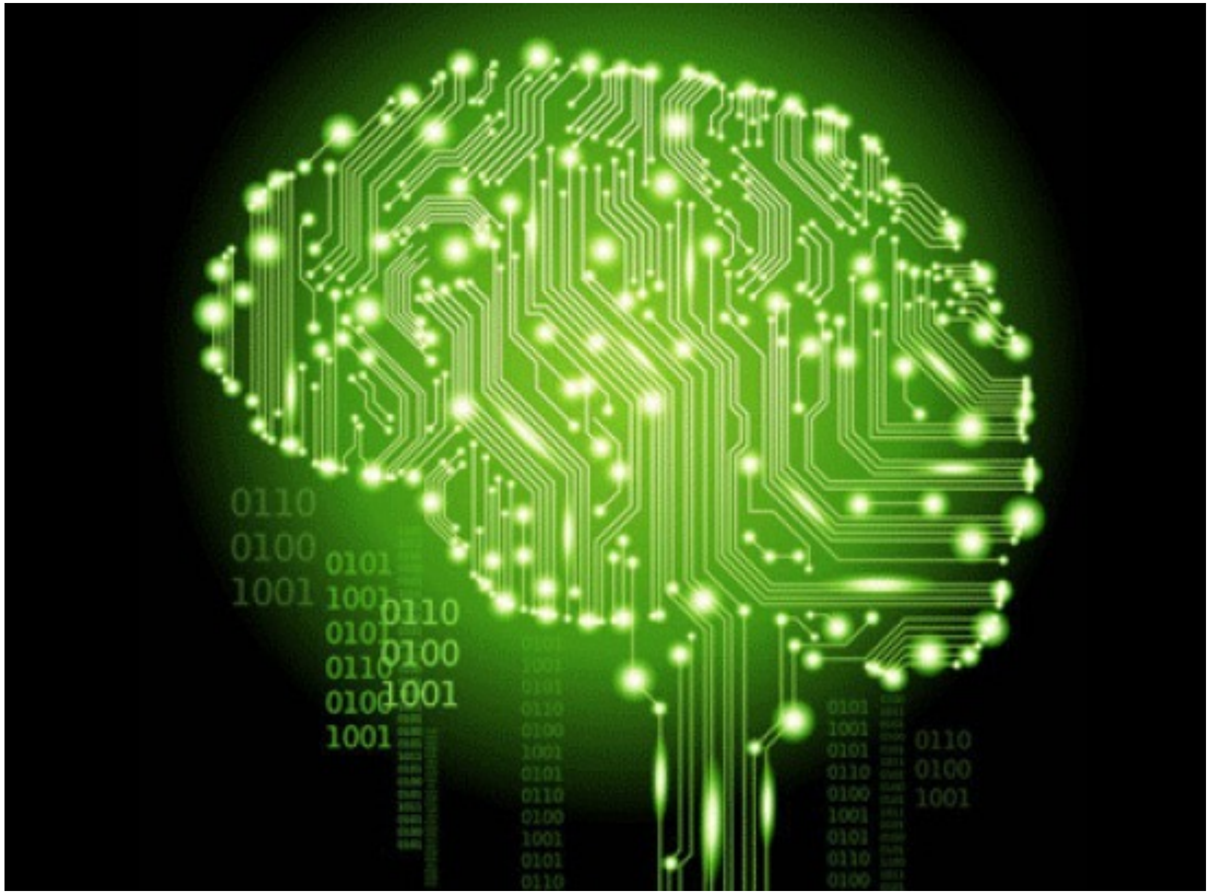
technische universität
dortmund



Uncertainty in Semi-Supervised Learning

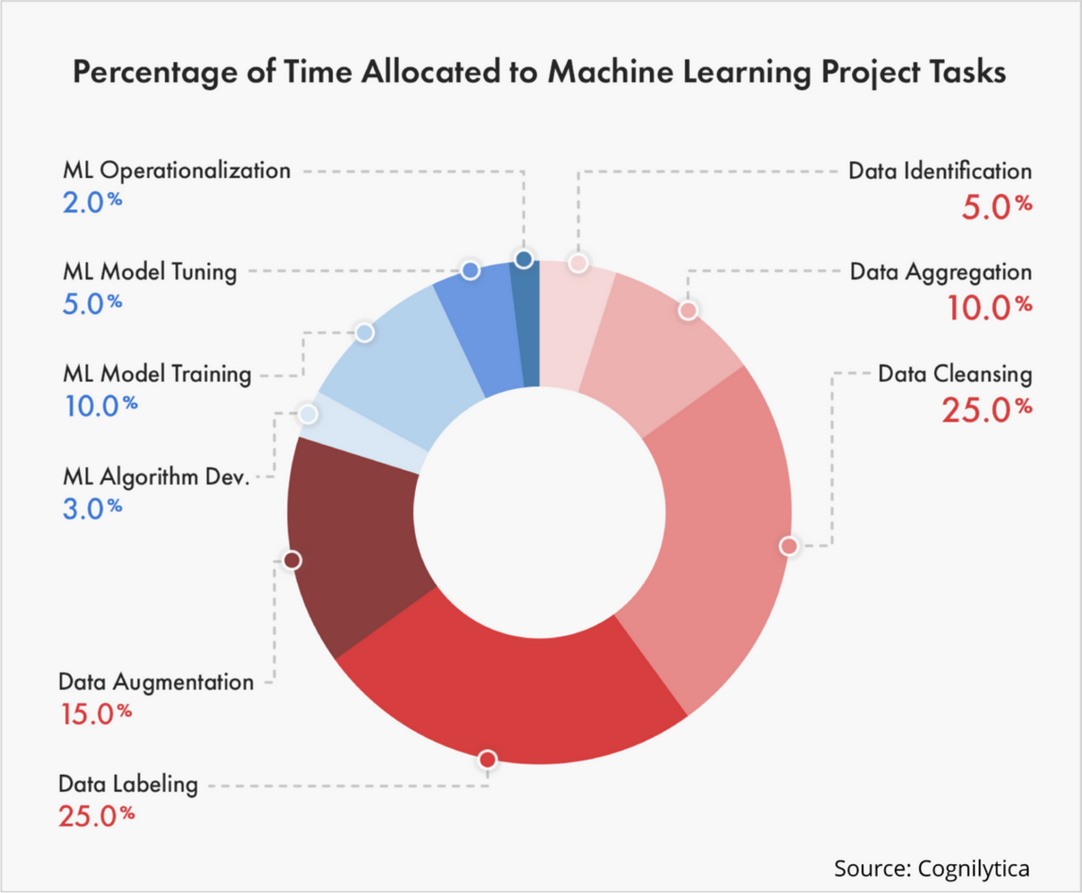
Jonas Dauer,11.07.2022

Table of Contents

technische universität dortmund

Table of Contents

Labeling Data

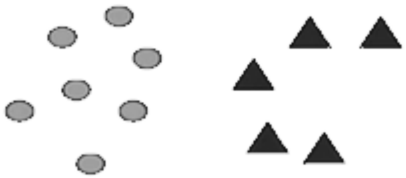


Source: https://www.cloudfactory.com/data-labeling-guide

technische universität
dortmund

Table of Contents

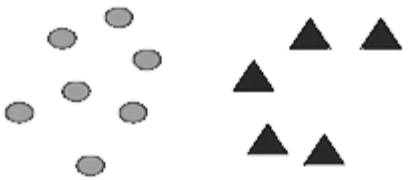
Semi-supervised Learning



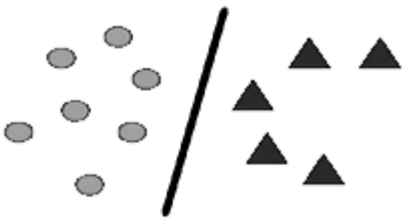
Labeled Data

Source: https://medium.com/dataseries/two-minutes-of-semi-supervised-learning-f0eb62729530

Semi-supervised Learning



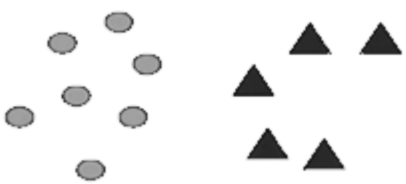
Labeled Data



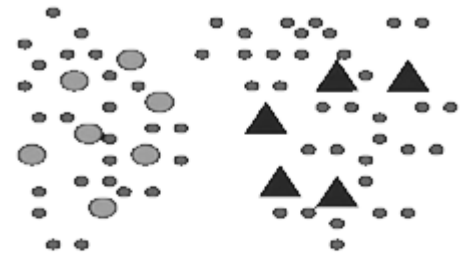
Supervised Learning

Source: https://medium.com/dataseries/two-minutes-of-semi-supervised-learning-f0eb62729530

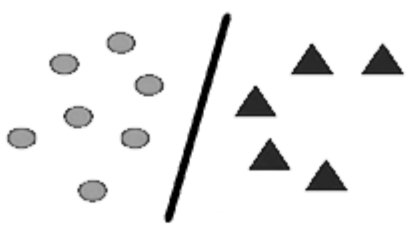
Semi-supervised Learning



Labeled Data



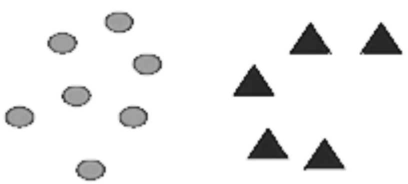
Labeled and
Unlabeled Data



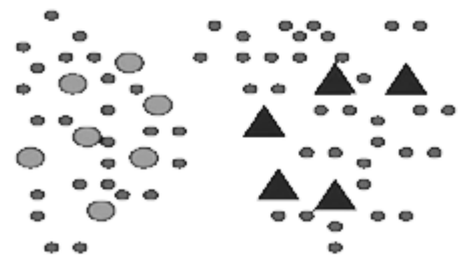
Supervised Learning

Source: https://medium.com/dataseries/two-minutes-of-semi-supervised-learning-f0eb62729530

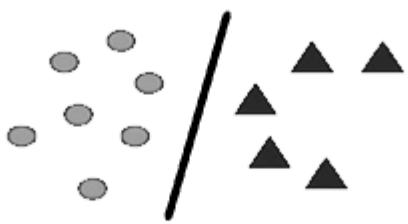
Semi-supervised Learning



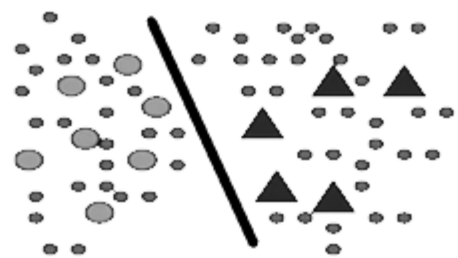
Labeled Data



Labeled and
Unlabeled Data



Supervised Learning



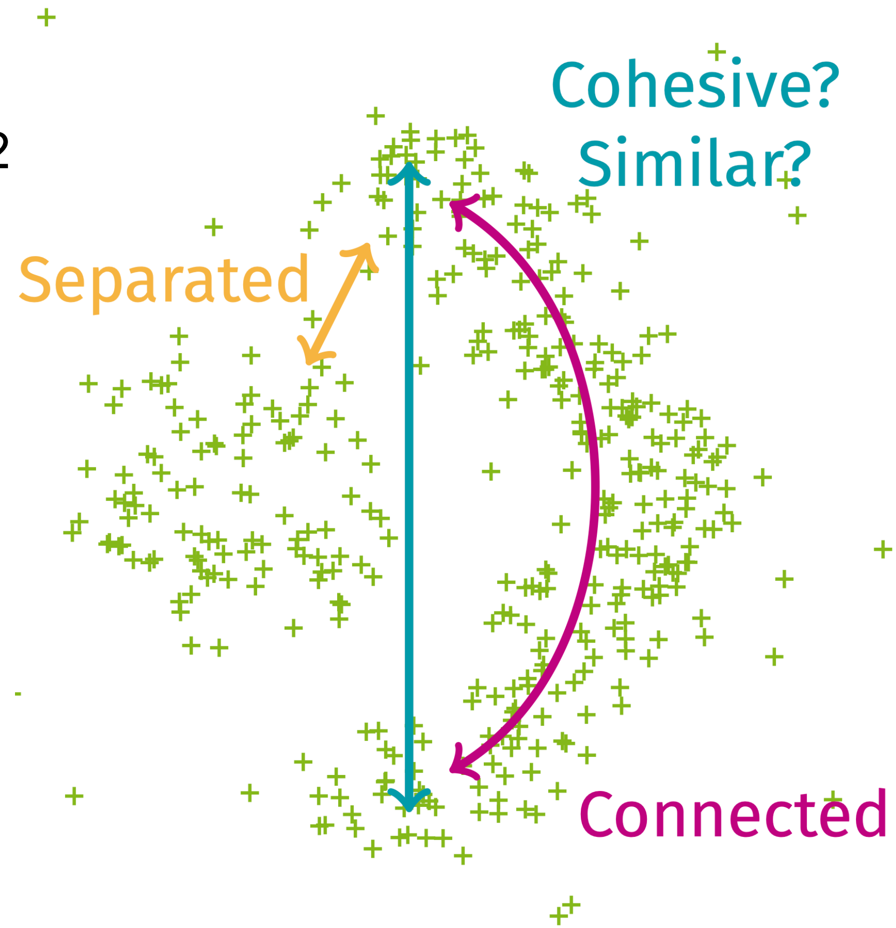
Semi-Supervised
Learning

Source: https://medium.com/dataseries/two-minutes-of-semi-supervised-learning-f0eb62729530

technische universität
dortmund

Smoothness Assumption

If two points x_1 and x_2 are close, then so should be the corresponding outputs y_1 and y_2.

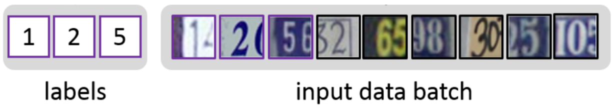


Source: Maschinelles Lernen

Table of Contents

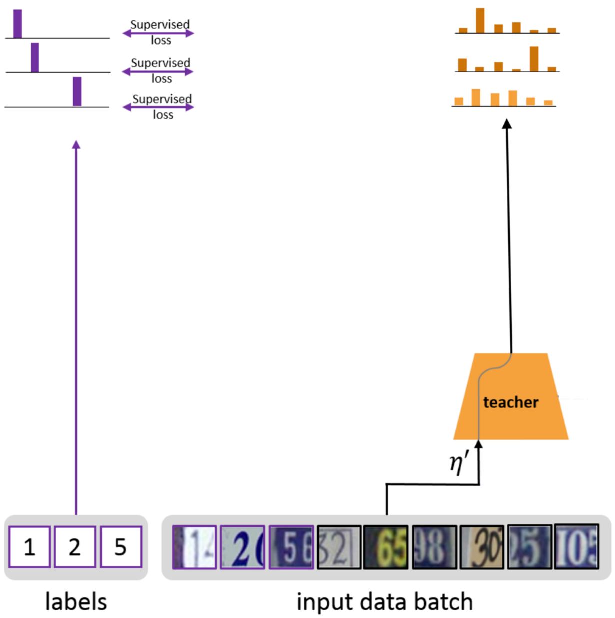
Architecture

- Partly labeled Dataset



Architecture

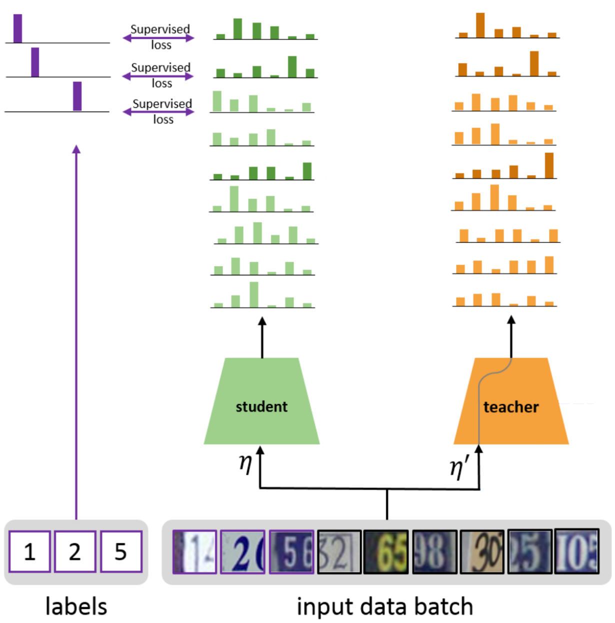
- Partly labeled Dataset
- Supervised teacher network



technische universität
dortmund

Architecture

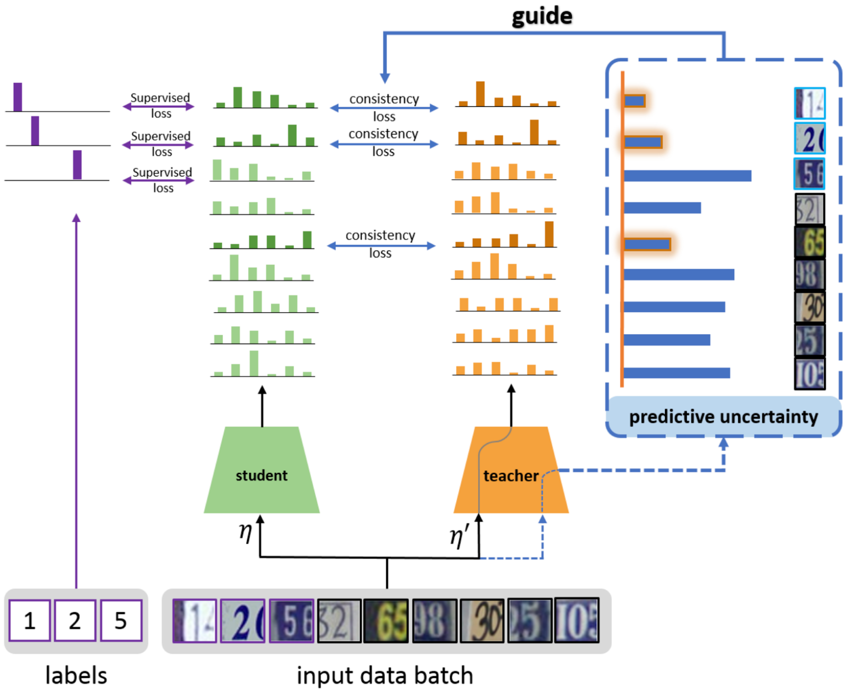
- Partly labeled Dataset
- Supervised teacher network
- „Trained“ student network



technische universität
dortmund

Architecture

- Partly labeled Dataset
- Supervised teacher network
- „Trained“ student network
- Uncertainty measurement



technische universität
dortmund

Architecture

- Partly labeled Dataset
- Supervised teacher network
- „Trained“ student network
- Uncertainty measurement
- Teacher gets knowledge from student

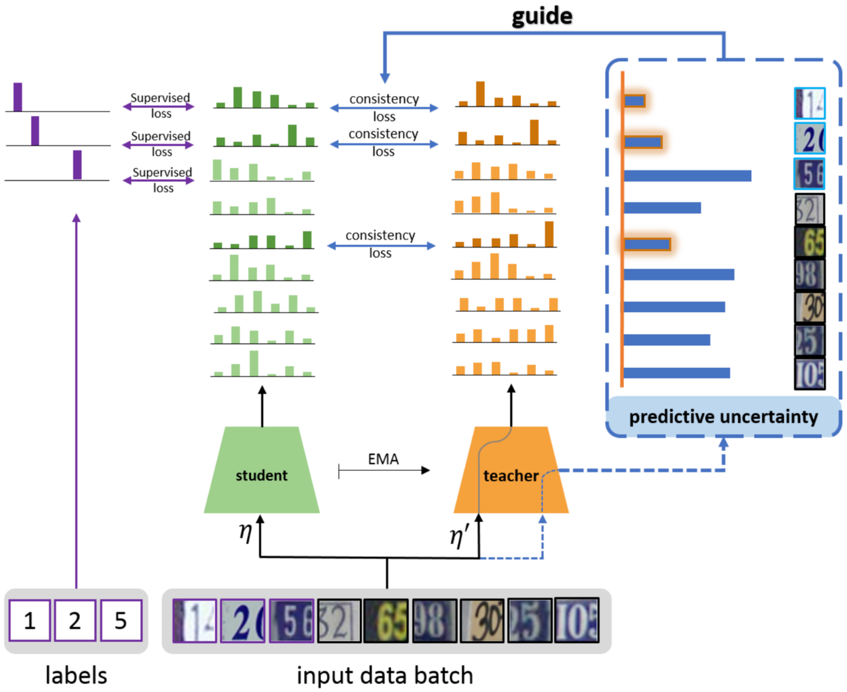
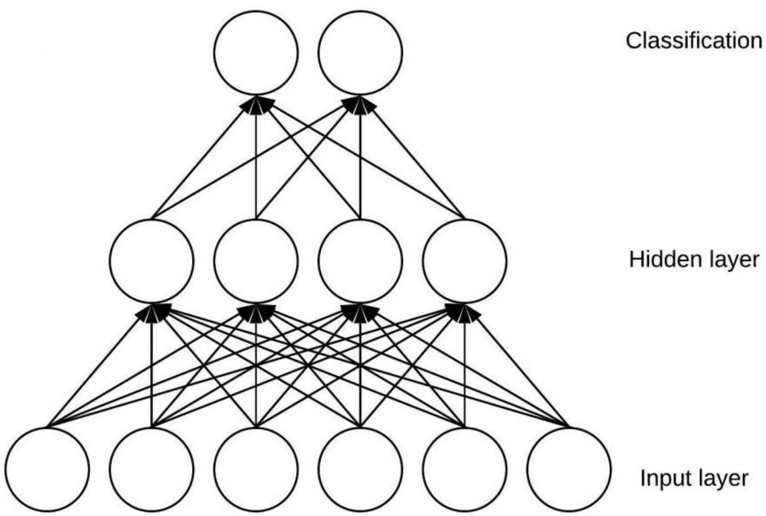
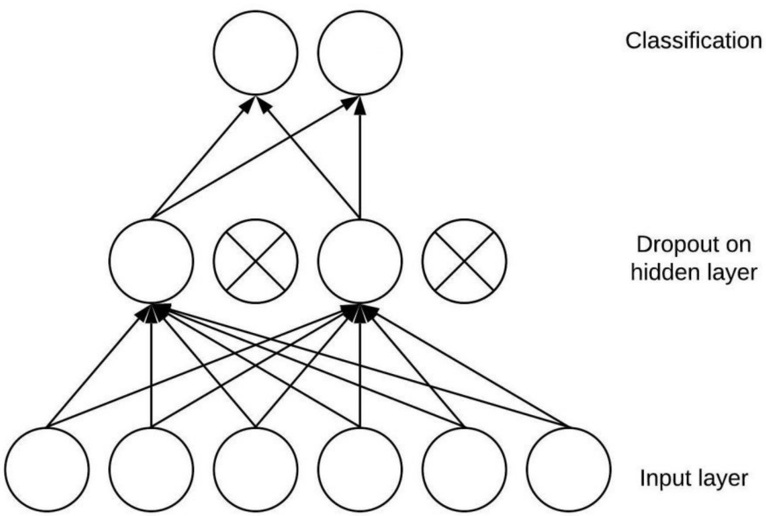


Table of Contents

Dropout



Without Dropout

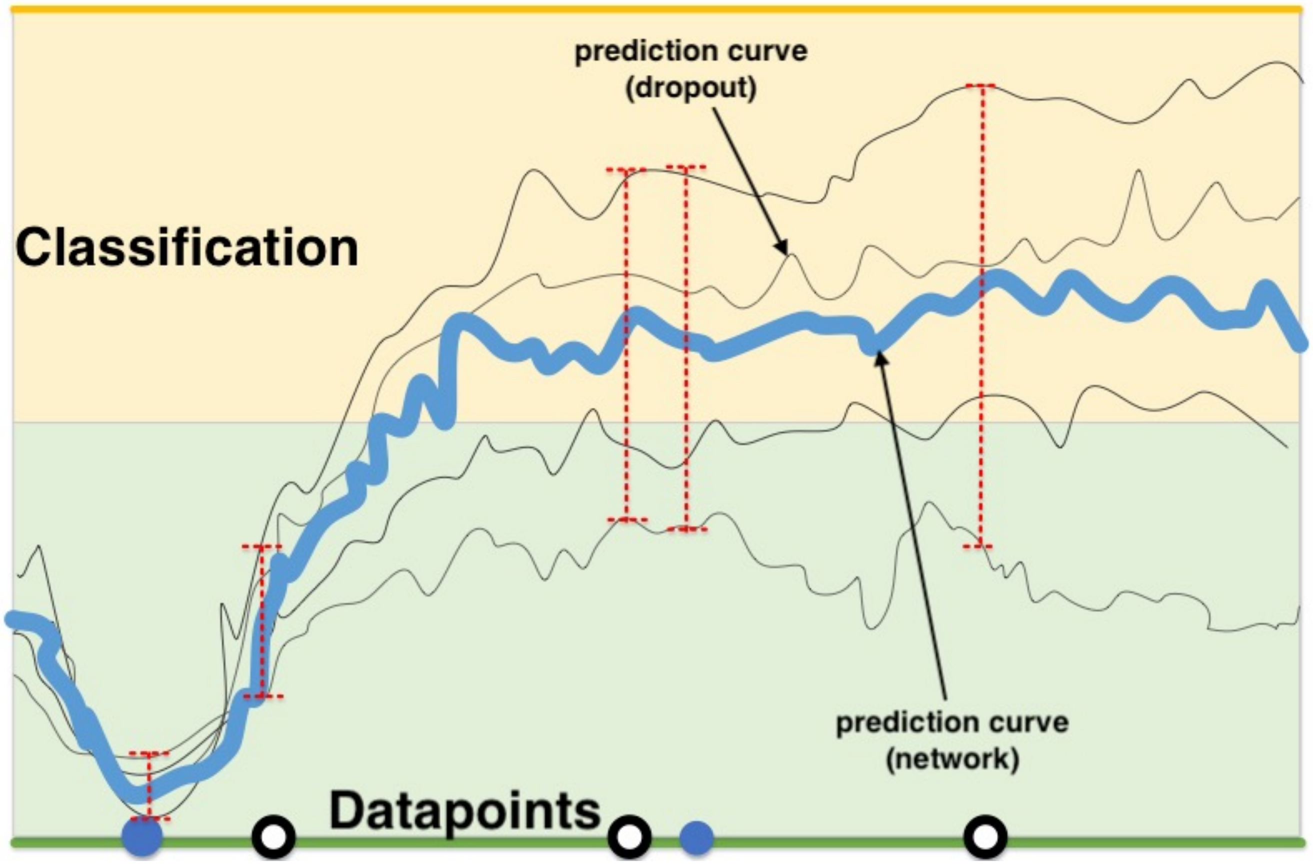


With Dropout

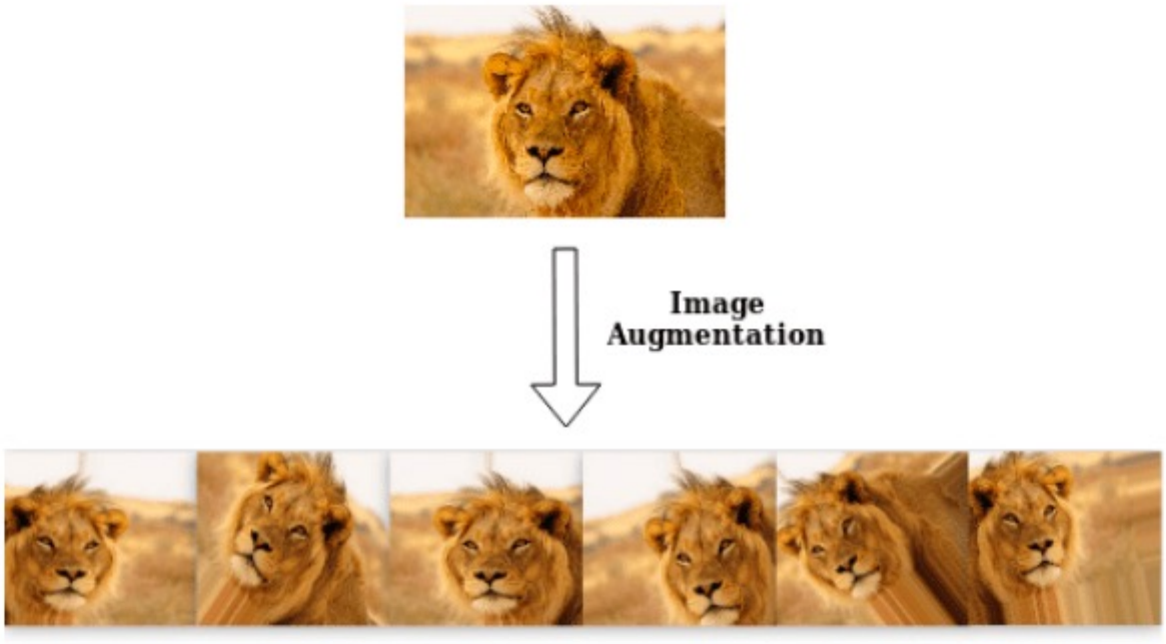
https://wenkangwei.github.io/2020/11/13/DL-DropOut/

technische universität
dortmund

Dropout

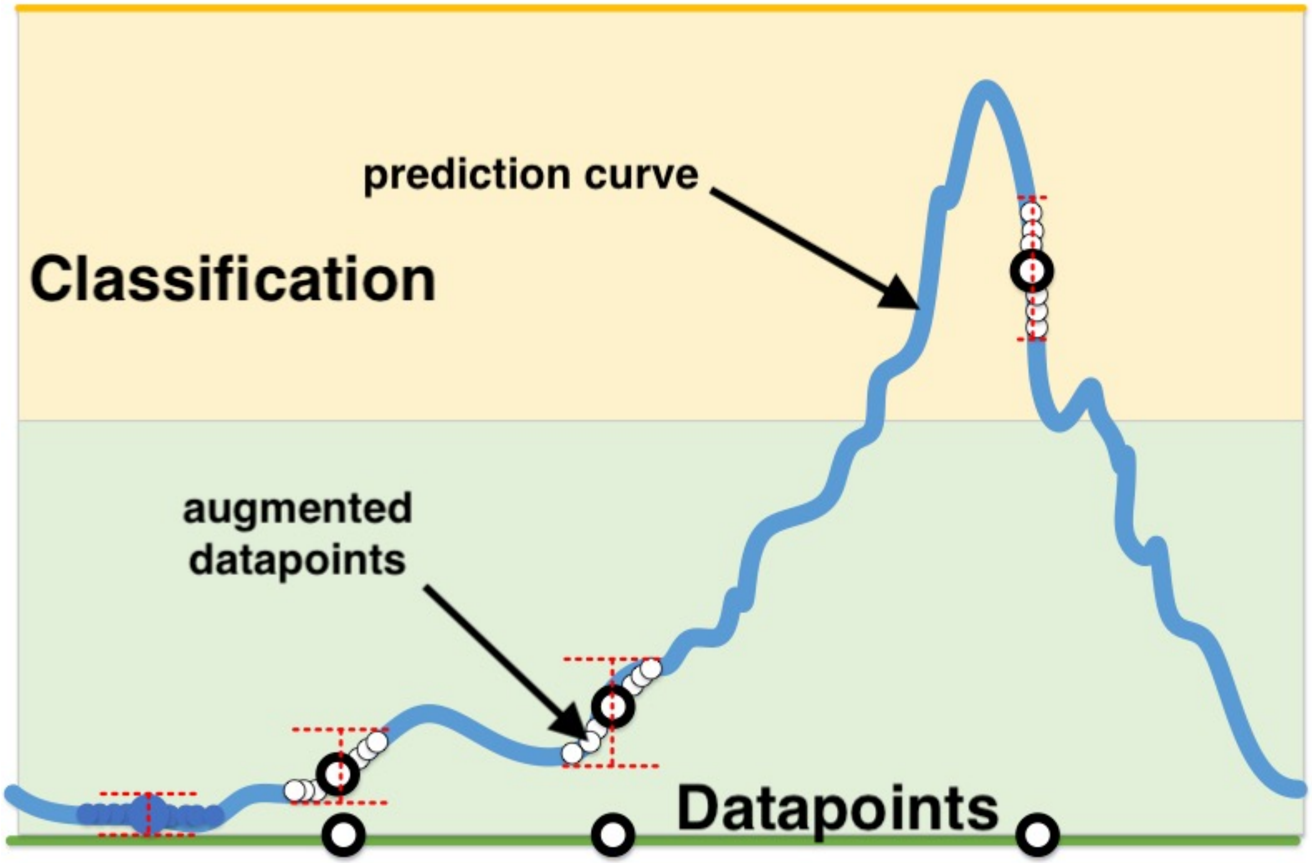


Augmentation



https://towardsdatascience.com/machinex-image-data-augmentation-using-keras-b459ef87cd22

Augmentation



Determine Uncertainty

Procedure:

1. Take batch B
2. Classify with teacher every x_i in B 20 times with dropout and augmentation
3. Calculate Criteria
4. Learn depending on the criteria

Determine Uncertainty

Procedure:

1. Take batch B
2. Classify with teacher every x_i in B 20 times with dropout and augmentation
3. Calculate Criteria
4. Learn depending on the criteria

Criteria:

- Measure variance over T times random samples
- Reflect the probability distribution of different classes
- Continuous scalar as output

Determine Uncertainty

Procedure:

1. Take batch B
2. Classify with teacher every x_i in B 20 $\mathrm{times\ with}$ dropout and augmentation
3. Calculate Criteria
4. Learn depending on the criteria

Criteria:

- Measure variance over T times random samples
- Reflect the probability distribution of different classes
- Continuous scalar as output

e.g. Predictive Variance (PV):

- Variance of multiple soft predictions
- The larger the variance, the higher uncertainty

$$PV = \sum_c Var\left[p\left(y = c \middle| x, \widehat{\Theta}^1, \hat{\eta}^1\right), \dots, p\left(y = c \middle| x, \widehat{\Theta}^T, \hat{\eta}^T\right)\right]$$

Table of Contents

Certainty-driven consistency with filtering

- Idea: Learn only from certain targets

Certainty-driven consistency with filtering

- Idea: Learn only from certain targets

- Procedure:
  1. Compute criteria for data points in input Batch B
  2. Sort data points according to uncertainty values
  3. Chose the top-k certain data points
  4. Filter these randomly dependent on their uncertainty
  5. Train student network with these data points

technische universität
dortmund

Certainty-driven consistency with filtering

- Idea: Learn only from certain targets

- Procedure:
  1. Compute criteria for data points in input Batch B
  2. Sort data points according to uncertainty values
  3. Chose the top-k certain data points
  4. Filter these randomly dependent on their uncertainty
  5. Train student network with these data points

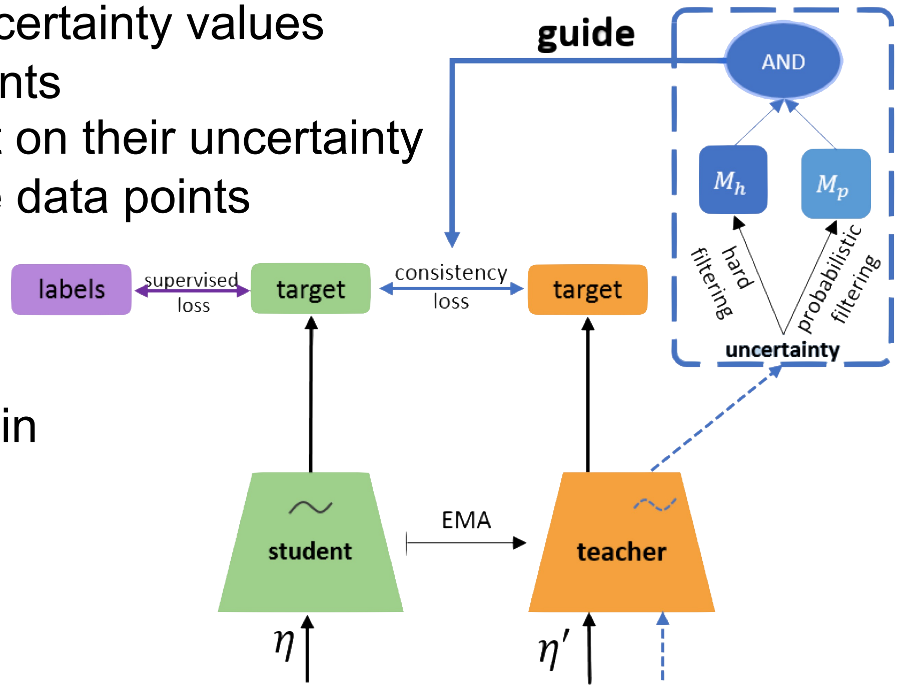
- Increase k with every epoche
  - Consistency loss decreases
  - More data points should be certain

Certainty-driven consistency with filtering

- Idea: Learn only from certain targets

- Procedure:
  1. Compute criteria for data points in input Batch B
  2. Sort data points according to uncertainty values
  3. Chose the top-k certain data points
  4. Filter these randomly dependent on their uncertainty
  5. Train student network with these data points

- Increase k with every epoche
  - Consistency loss decreases
  - More data points should be certain



Temperature CCL

- Idea: Learn more from certain targets and less from uncertain
  - „dark knowledge“ could be helpful
  - e.g. similarity between classes

- Procedure:
  - Use softmax activation with temperature

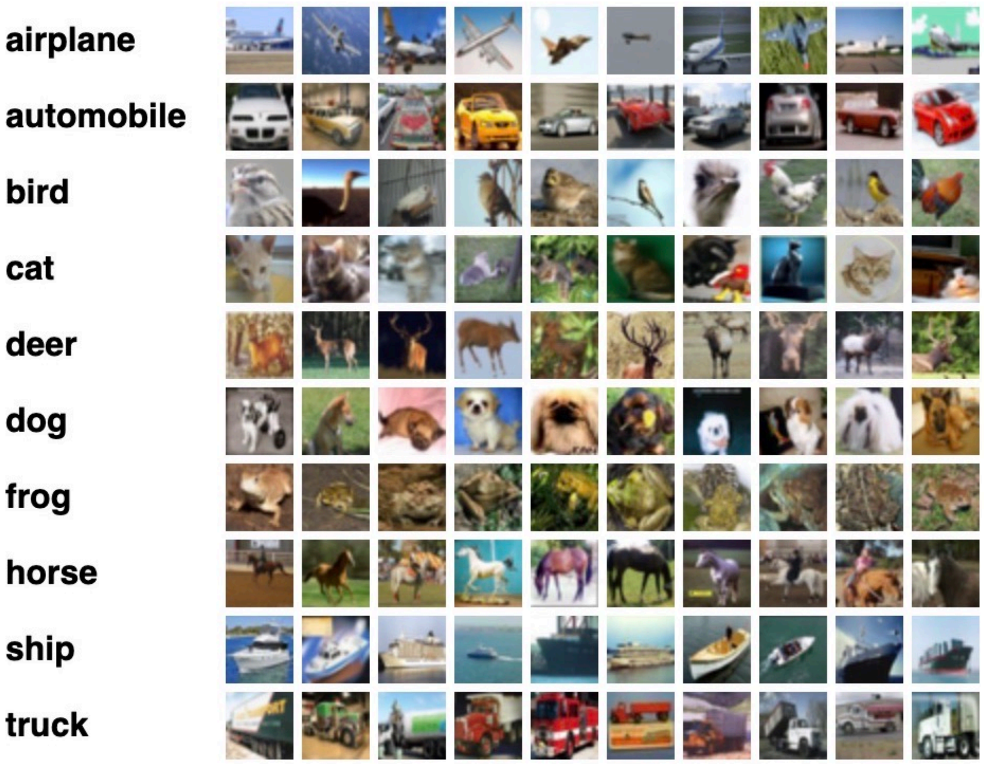
$$q_i = \frac{\exp(z_i/V_i)}{\sum_j \exp(z_j/V_j)}$$

  - V_i depends on certainty of x_i
  - If $V_i = 1$: softmax activation
  - For large V_i: equal distribution

Table of Contents

CIFAR

- 60.000 images
- CIFAR-10 and CIFAR-100



Source: https://www.cs.toronto.edu/~kriz/cifar.html

SVHN

- Street View House Numbers Dataset
- 73.257 images

Source: http://ufldl.stanford.edu/housenumbers/

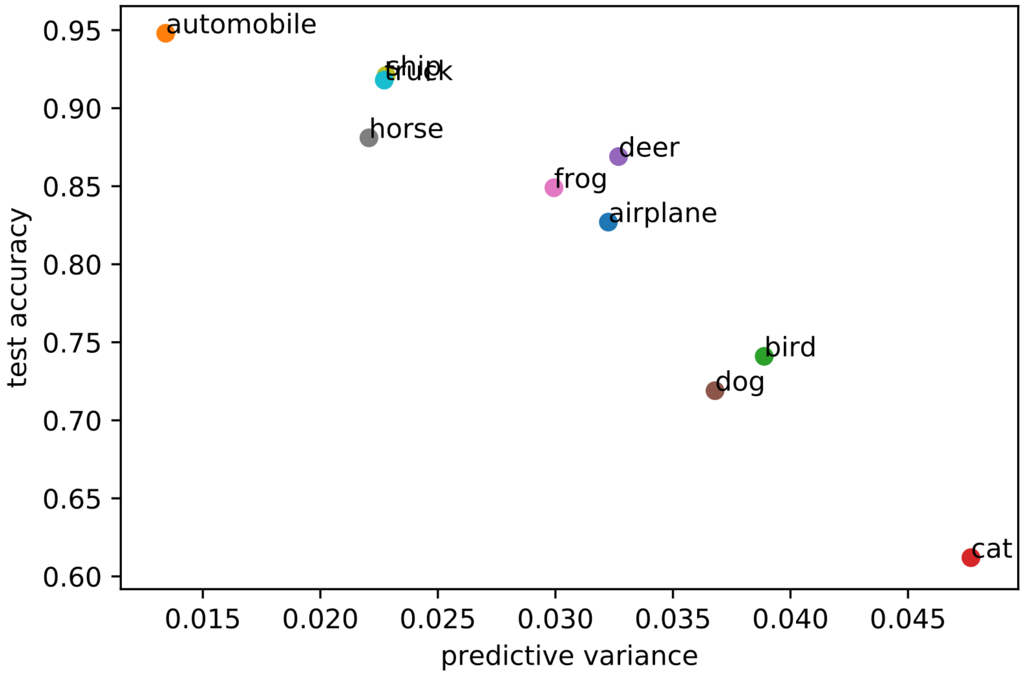
technische universität dortmund

Experiments overview

- Run every method 10 times (average)
- Metric: Error rate (%), ± standard deviation

Model	CIFAR-10			SVHN	CIFAR-100
	1000 labels	2000 labels	4000 labels	1000 labels	10000 labels
Supervised-only	46.43 ± 1.21	33.94 ± 0.73	20.66 ± 0.57	12.32 ± 0.95	44.56 ± 0.30
Π model	–	–	12.36 ± 0.31	4.82 ± 0.17	39.19 ± 0.36
TempEns	–	–	12.16 ± 0.24	4.42 ± 0.16	38.65 ± 0.51
VAT+Ent	–	–	10.55 ± 0.05	3.86 ± 0.11	–
MT	21.55 ± 1.48	15.73 ± 0.31	12.31 ± 0.28	3.95 ± 0.19	37.91 ± 0.37
Π+SNTG	21.23 ± 1.27	14.65 ± 0.31	11.00 ± 0.13	**3.82** ± **0.25**	37.97 ± 0.29
MT+SNTG	–	–	–	3.86 ± 0.27	–
TempEns+SNTG	18.41 ± 0.52	13.64 ± 0.32	10.93 ± 0.14	3.98 ± 0.21	–
MA-DNN	–	–	11.91 ± 0.22	4.21 ± 0.12	**34.51** ± **0.61**
Filtering CCL (**ours**)	**16.99** ± **0.71**	12.57 ± 0.47	**10.63** ± **0.22**	3.86 ± 0.19	34.81 ± 0.52
Temperature CCL (**ours**)	17.26 ± 0.69	**12.45** ± **0.33**	10.73 ± 0.26	3.93 ± 0.21	35.15 ± 0.62

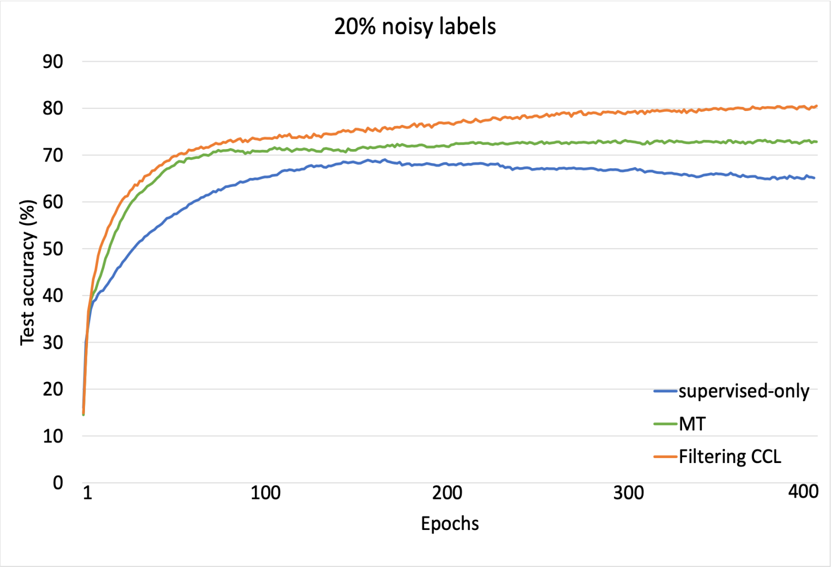
Accuracy vs. PV



- Measurement of PV and accuracy in the CIFAR data set
- Inverse relationship between class accuracy and predictive variance

technische universität
dortmund

Robustness to noisy labels



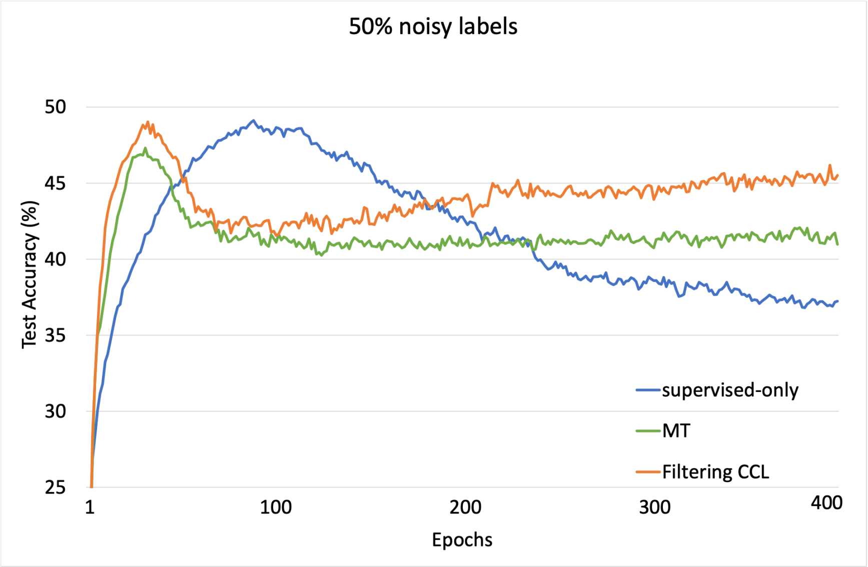


Table of Contents

Sum up

- Semi-supervised Learning
    - Labeling is expensive
    - Learn from partly labeled data set

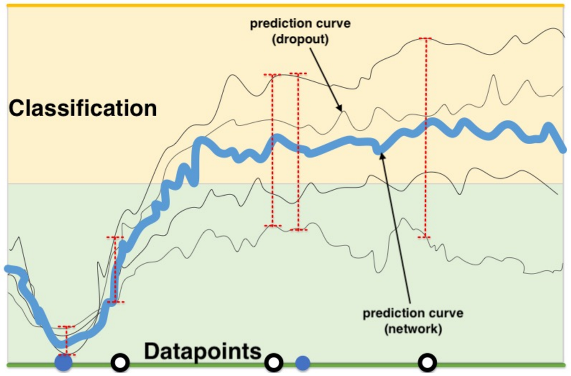
Sum up

- Semi-supervised Learning
  - Labeling is expensive
  - Learn from partly labeled data set

- Only learn from certain data points

Sum up

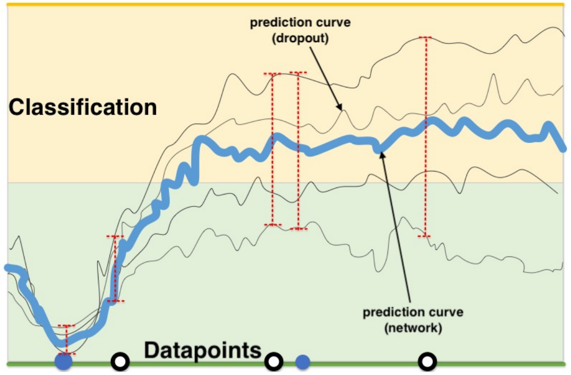
- Semi-supervised Learning
  - Labeling is expensive
  - Learn from partly labeled data set
- Only learn from certain data points
- Determine uncertainty with dropout



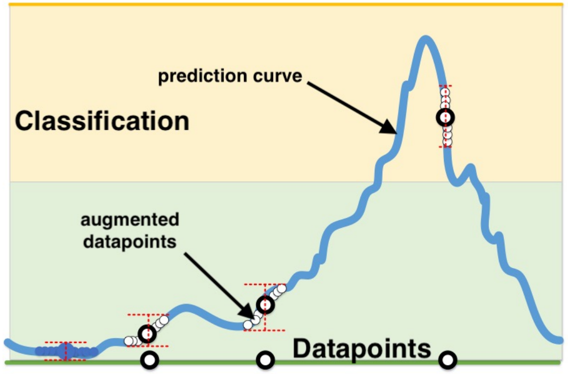
Sum up

- Semi-supervised Learning
  - Labeling is expensive
  - Learn from partly labeled data set

- Only learn from certain data points

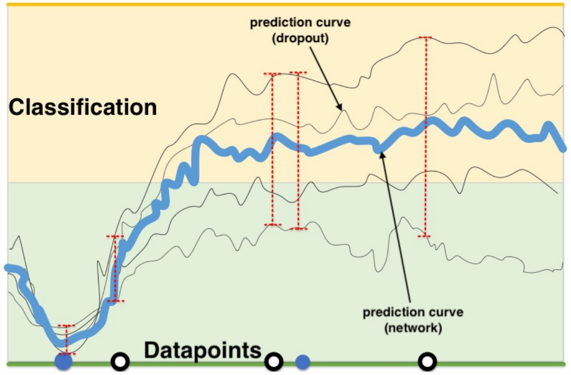


- Determine uncertainty with dropout and augmentation



technische universität dortmund

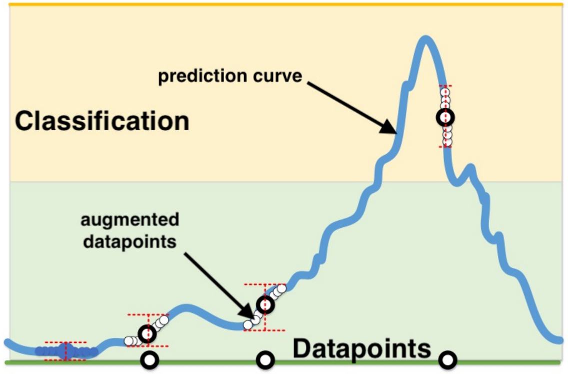
Sum up



- Semi-supervised Learning
  - Labeling is expensive
  - Learn from partly labeled data set

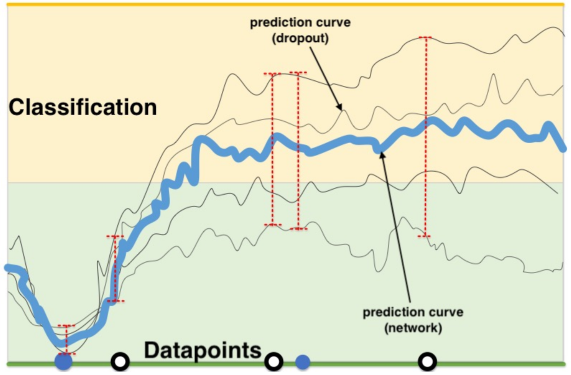
- Only learn from certain data points

- Determine uncertainty with dropout and augmentation
  - Classify one data point multiple times with random dropout and augmentation
  - Compute criteria and get uncertainty



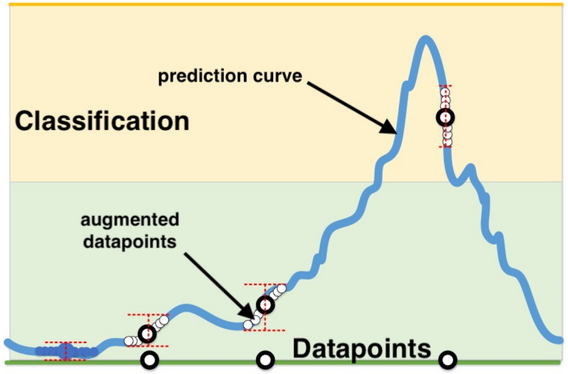
Sum up

- Semi-supervised Learning
  - Labeling is expensive
  - Learn from partly labeled data set



- Only learn from certain data points

- Determine uncertainty with dropout and augmentation
  - Classify one data point multiple times with random dropout and augmentation
  - Compute criteria and get uncertainty



- Use filtering or temperature to learn from certain data points