HELMHOLTZ AI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

Uncertainty Quantification in Machine Learning

A Primer

Peter Steinbach, Till Korten, Sebastian Starke, Steve Schmerler

Helmholtz-Zentrum Dresden-Rossendorf / 2024-06-11

Today's Agenda

1. Uncertainties!
2. Predictive Uncertainties
   - MCDropout
   - Uncertainty Calibration
   - DeepEnsembles
3. Use Case: Predictive Uncertainties for Instance Segmentations

Uncertainties!

What is Uncertainty?

Definition (Merriam-Webster today)

uncertainty: not known beyond doubt, not having certain knowledge, not clearly identified or defined, not constant, indefinite, not certain to occur, not reliable

What is Uncertainty?

Definition (Merriam-Webster today)

uncertainty: not known beyond doubt, not having certain knowledge, not clearly identified or defined, not constant, indefinite, not certain to occur, not reliable

Definition (Wikipedia today)

Uncertainty refers to *epistemic* situations involving imperfect or unknown information. It applies to predictions of future events, to physical measurements that are already made, or to the unknown. Uncertainty arises in partially observable or stochastic environments, as well as due to ignorance, indolence, or both.
likely from (Russell & Norvig, 2010)

What is Uncertainty?

Definition (Merriam-Webster today)

uncertainty: not known beyond doubt, not having certain knowledge, not clearly identified or defined, not constant, indefinite, not certain to occur, not reliable

Definition (Wikipedia today)

Uncertainty refers to *epistemic* situations involving imperfect or unknown information. It applies to predictions of future events, to physical measurements that are already made, or to the unknown. Uncertainty arises in partially observable or stochastic environments, as well as due to ignorance, indolence, or both.
likely from (Russell & Norvig, 2010)

...

Uncertainties in ML for Science (Tan *et al.*, 2023)

npj

A.R. Tan et al.

2

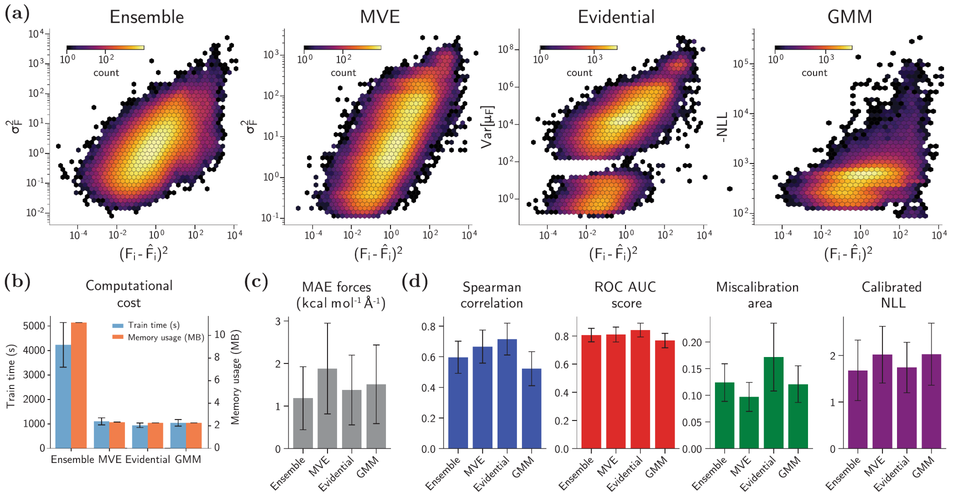


Fig. 1 Comparison of UQ methods on the rMD17 data set. a Hexbin plots showing (predicted) uncertainties versus squared errors of atomic

Uncertainties: a paradigm shift in the ML community

*Let's say I have data and I fit a linear model to predict y from x. …So that would be the predictive take. The inferential take would be: can I say something about which features are significant in this (linear) model? Can I actually give under some assumptions **a confidence interval** for their coefficient in the linear model and so on.*

Uncertainties: a paradigm shift in the ML community

*Let's say I have data and I fit a linear model to predict y from x. …So that would be the predictive take. The inferential take would be: can I say something about which features are significant in this (linear) model? Can I actually give under some assumptions **a confidence interval** for their coefficient in the linear model and so on.*

*In statistics, we think of these as like two very big areas of research. And maybe historically, actually **inference** has been even bigger than **prediction**. And in machine learning, it's exactly the opposite: So prediction dominates, inference is tiny.*

Uncertainties: a paradigm shift in the ML community

*Let's say I have data and I fit a linear model to predict y from x. ...So that would be the predictive take. The inferential take would be: can I say something about which features are significant in this (linear) model? Can I actually give under some assumptions **a confidence interval** for their coefficient in the linear model and so on.*

*In statistics, we think of these as like two very big areas of research. And maybe historically, actually **inference** has been even bigger than **prediction**. And in machine learning, it's exactly the opposite: So prediction dominates, inference is tiny.*

*And historically in machine learning, inference may have been very, very small. And now I think it's grown in a way that people do talk about inference, they use the word **uncertainty quantification**. They don't think about inference and typically in the traditional way statistically. But I think it has somehow emerged as maybe more of a focus.*

Ryan Tibshirani in "The Gradient Podcast", see here for the full interview.

Origins of Uncertainty in ML 1/3

Supervised Machine Learning

- *dataset* $\mathcal{D} = \{(\vec{x}_0, y_0), (\vec{x}_1, y_1), ..., (\vec{x}_N, y_N)\} \subset \mathcal{X} \times \mathcal{Y}$
  $\mathcal{X}$ ...instance space
  $\mathcal{Y}$ ...outcomes associated with instance

Origins of Uncertainty in ML 1/3

Supervised Machine Learning

- *dataset* $\mathcal{D} = \{(\vec{x}_0, y_0), (\vec{x}_1, y_1), ..., (\vec{x}_N, y_N)\} \subset \mathcal{X} \times \mathcal{Y}$
  $\mathcal{X}$ ...instance space
  $\mathcal{Y}$ ...outcomes associated with instance
- each sample of $\mathcal{D}$ is considered an i.i.d. sample

Origins of Uncertainty in ML 1/3

Supervised Machine Learning

- *dataset* $\mathcal{D} = \{(\vec{x}_0, y_0), (\vec{x}_1, y_1), ..., (\vec{x}_N, y_N)\} \subset \mathcal{X} \times \mathcal{Y}$
  $\mathcal{X}$ ...instance space
  $\mathcal{Y}$ ...outcomes associated with instance
- each sample of $\mathcal{D}$ is considered an i.i.d. sample
- given a *hypothesis* space $\mathcal{H}$ (with $h : \mathcal{X} \rightarrow \mathcal{Y}$) and

Origins of Uncertainty in ML 1/3

Supervised Machine Learning

- *dataset* $\mathcal{D} = \{(\vec{x}_0, y_0), (\vec{x}_1, y_1), ..., (\vec{x}_N, y_N)\} \subset \mathcal{X} \times \mathcal{Y}$
  $\mathcal{X}$ ...instance space
  $\mathcal{Y}$ ...outcomes associated with instance
- each sample of $\mathcal{D}$ is considered an i.i.d. sample
- given a *hypothesis* space $\mathcal{H}$ (with $h : \mathcal{X} \rightarrow \mathcal{Y}$) and
- given a *loss function* $l : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$

Origins of Uncertainty in ML 2/3

i.i.d. dataset $\mathcal{D}$
$\{(\vec{x}_0, y_0), ..., (\vec{x}_N, y_N)\}$

hypothesis space
$\mathcal{H}$, $h : \mathcal{X} \rightarrow \mathcal{Y}$

loss function
$l = f(h(x), y))$

Origins of Uncertainty in ML 2/3

i.i.d. dataset $\mathcal{D}$

$$\{(\vec{x}_0, y_0), ..., (\vec{x}_N, y_N)\}$$

hypothesis space

$$\mathcal{H},\ h : \mathcal{X} \rightarrow \mathcal{Y}$$

loss function

$$l = f(h(x), y))$$

Learner Goal

To induce a hypothesis $h^* \in \mathcal{H}$ with low risk $R(h)$:

$$R(h) := \int_{\mathcal{X}\times\mathcal{Y}} l(h(x), y) dP(x, y)$$

Origins of Uncertainty in ML 2/3

i.i.d. dataset $\mathcal{D}$

$\{(\vec{x}_0, y_0), ..., (\vec{x}_N, y_N)\}$

hypothesis space

$\mathcal{H}$, $h : \mathcal{X} \to \mathcal{Y}$

loss function

$l = f(h(x), y))$

Learner Goal

To induce a hypothesis $h^* \in \mathcal{H}$ with low risk $R(h)$:

$$R(h) := \int_{\mathcal{X}\times\mathcal{Y}} l(h(x), y) dP(x, y)$$

Learner Guesses

a good hypothesis h guided by the empirical risk $R_{emp}(h)$:

$$R_{emp} := \frac{1}{N} \sum_{i=1}^{N} l(h(x_i), y_i)$$

Origins of Uncertainty in ML 2/3

i.i.d. dataset $\mathcal{D}$

$$\{(\vec{x}_0, y_0), ..., (\vec{x}_N, y_N)\}$$

hypothesis space

$$\mathcal{H},\ h : \mathcal{X} \rightarrow \mathcal{Y}$$

loss function

$$l = f(h(x), y))$$

Learner Goal

To induce a hypothesis $h^* \in \mathcal{H}$ with low risk $R(h)$:

$$R(h) := \int_{\mathcal{X}\times\mathcal{Y}} l(h(x), y) dP(x, y)$$

Learner Guesses

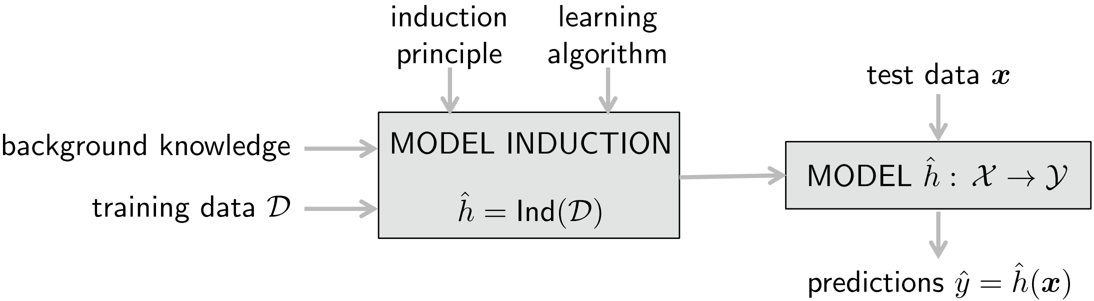
a good hypothesis h guided by the empirical risk $R_{emp}(h)$:

$$R_{emp} := \frac{1}{N} \sum_{i=1}^{N} l(h(x_i), y_i)$$

R_{emp} only estimates R!

- $h^* := \arg\min_{h \in \mathcal{H}} R(h)$ will not coincide with $\hat{h} := \arg\min_{h \in \mathcal{H}} R_{emp}(h)$
- **uncertainy created!** what is h^*? How close is $\hat{h}$ to h^*? What is $R(\hat{h})$?

Wrap-Up Origins of Uncertainty (Hüllermeier & Waegeman, 2021)

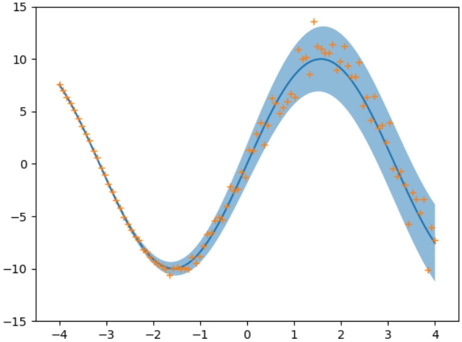


Central Point of Interest: **Predictive Uncertainties**

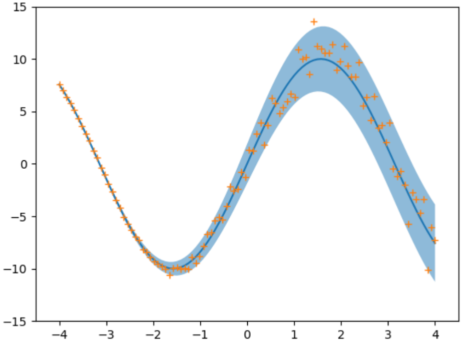
(uncertainty for $y_q = \hat{h}(x_q)$ for a concrete instance $x_q \in \mathcal{X}$)

Predictive Uncertainties

A historic view on Uncertainty sources



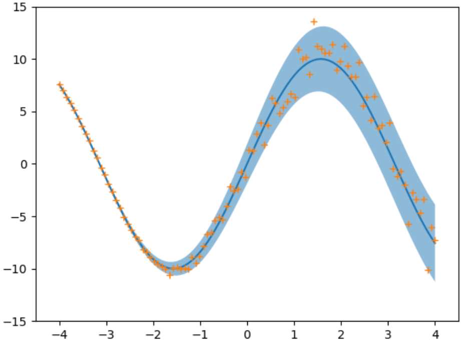
A historic view on Uncertainty sources



- **Aleatoric** or
  Data related **uncertainty**

  (uncertainties from noise in the data, does not decrease with more data)

A historic view on Uncertainty sources



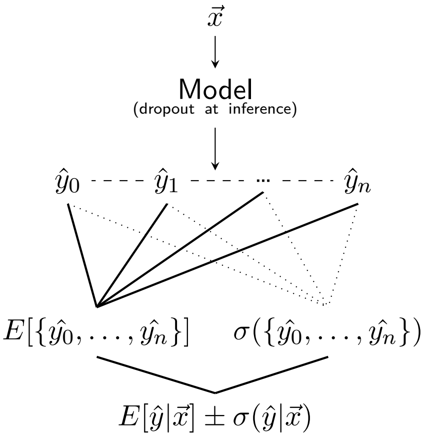
- **Aleatoric** or
  Data related **uncertainty**

  (uncertainties from noise in the data, does not decrease with more data)
- **Epistemic** or
  Model related **uncertainty**

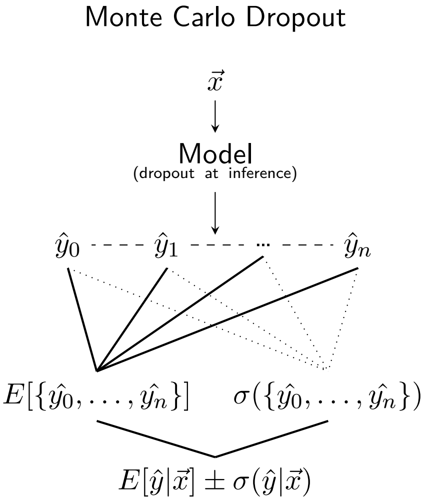
  (uncertainties related to finding the best hypothesis, can be reduced with more data)

Obtaining (epistemic) Uncertainties (Gal & Ghahramani, 2016)

Monte Carlo Dropout



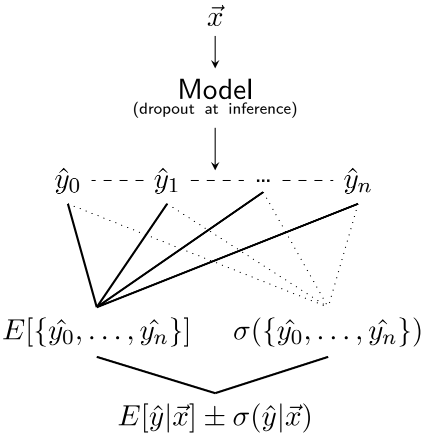
Obtaining (epistemic) Uncertainties (Gal & Ghahramani, 2016)



- model set up including dropout layers
- dropout layers set random portions of weights to 0.
  (implicit regularisation during training)

Obtaining (epistemic) Uncertainties (Gal & Ghahramani, 2016)

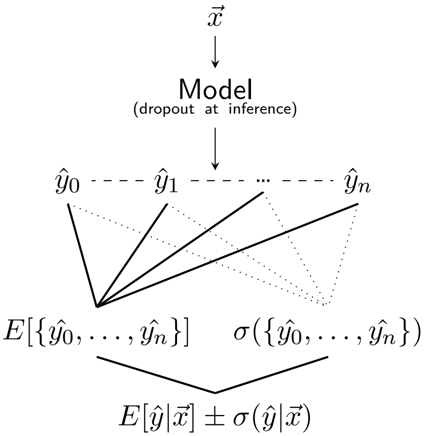
Monte Carlo Dropout



- model set up including dropout layers
- dropout layers set random portions of weights to 0.
  (implicit regularisation during training)
- **trick:** keep dropout enabled during prediction and call `model.predict` multiple times

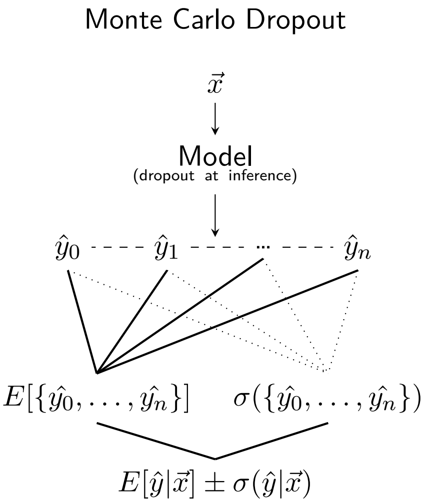
Obtaining (epistemic) Uncertainties (Gal & Ghahramani, 2016)

Monte Carlo Dropout



- model set up including dropout layers
- dropout layers set random portions of weights to 0.
  (implicit regularisation during training)
- **trick:** keep dropout enabled during prediction and call `model.predict` multiple times
- allows to sample the PDF describing $\hat{y}$
  (weights of model disabled randomly)

Obtaining (epistemic) Uncertainties (Gal & Ghahramani, 2016)

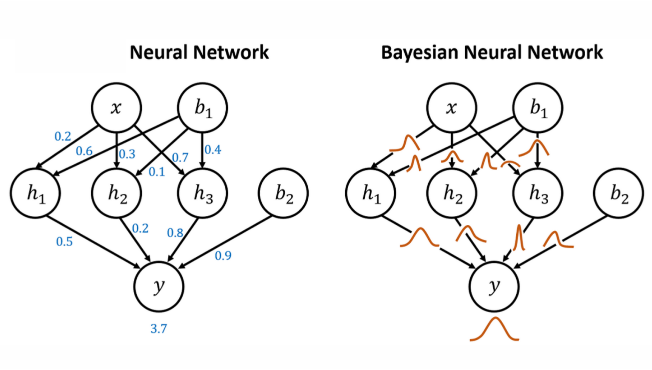


- model set up including dropout layers
- dropout layers set random portions of weights to 0.
  (implicit regularisation during training)
- **trick:** keep dropout enabled during prediction and call `model.predict` multiple times
- allows to sample the PDF describing $\hat{y}$
  (weights of model disabled randomly)
- obtain mean prediction $E[\hat{y}]$ and prediction deviation $\sigma(\hat{y})$ for each input

Try MCDropout for yourself

Please open
`mcdropout_1D_regression_vanilla.ipynb`!
Go through the notebook.

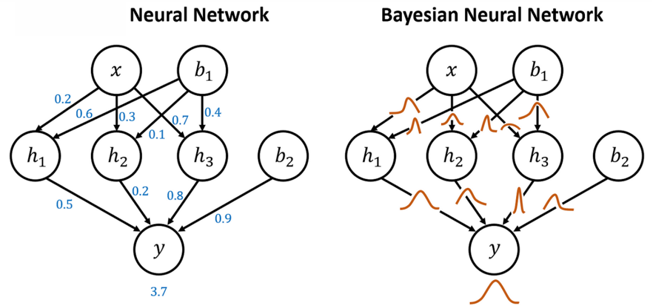
Dropout Recap



from jonascleveland.com

- MCDropout approximates a Bayesian Neural Network (BNNs are computational intensive)

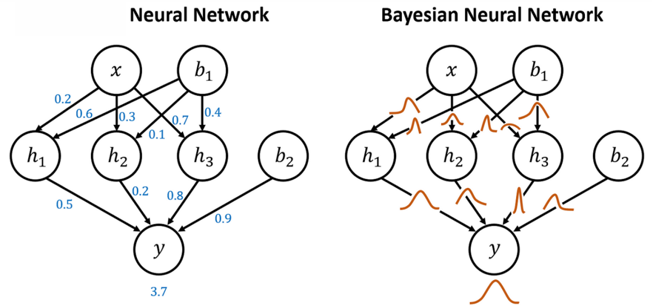
Dropout Recap



from jonascleveland.com

- MCDropout approximates a Bayesian Neural Network
  (BNNs are computational intensive)
- average across possible network configurations
  (enable Dropout layers during inference)

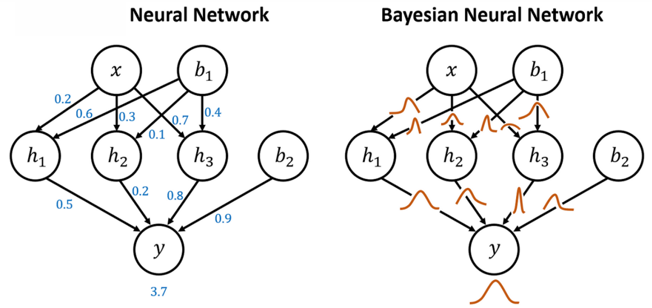
Dropout Recap



from jonascleveland.com

- MCDropout approximates a Bayesian Neural Network
  (BNNs are computational intensive)
- average across possible network configurations
  (enable Dropout layers during inference)
- perform Bayesian Model Averaging
  see (Gawlikowski *et al.*, n.d.) for details

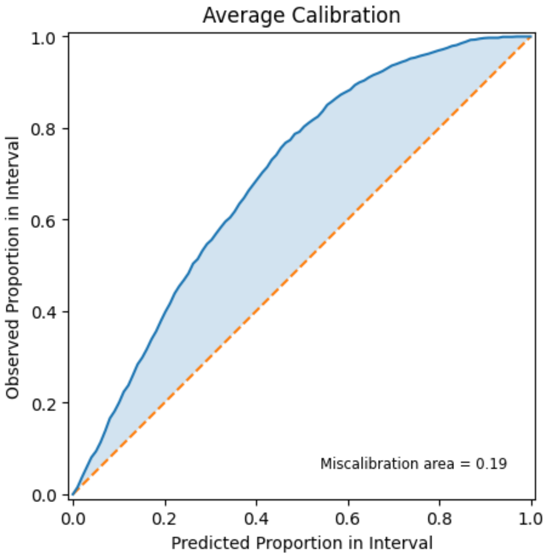
Dropout Recap



from jonascleveland.com

- MCDropout approximates a Bayesian Neural Network
  (BNNs are computational intensive)
- average across possible network configurations
  (enable Dropout layers during inference)
- perform Bayesian Model Averaging
  see (Gawlikowski *et al.*, n.d.) for details
- **epistemic uncertainties**

Diagnosing a UQ method: Calibration Curves



How to judge predictive uncertainties? (Kuleshov *et al.*, 2018)

At Inference, we have

- a label y_{test}
- a prediction $E[y_{test}]$ (by our model)
- an uncertainty for that prediction σy_{test} (from the UQ method)

How to judge predictive uncertainties? (Kuleshov *et al.*, 2018)

At Inference, we have

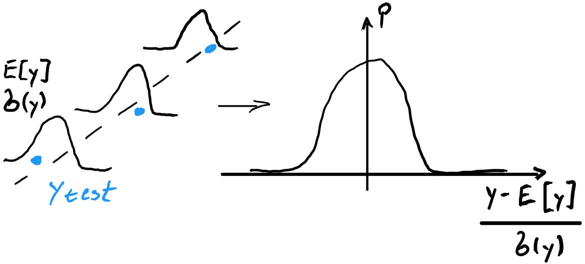
- a label y_{test}
- a prediction $E[y_{test}]$ (by our model)
- an uncertainty for that prediction σy_{test} (from the UQ method)

Assumption

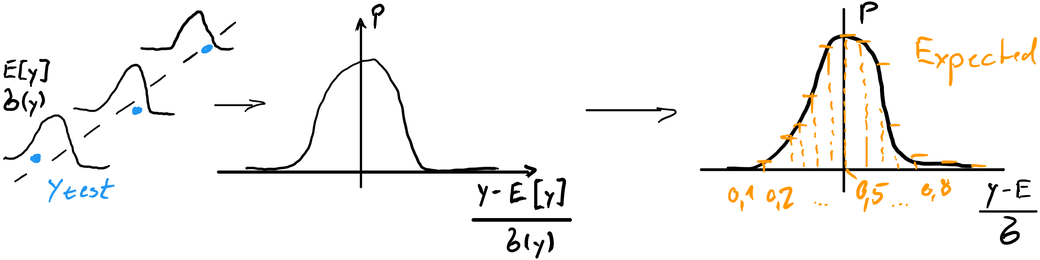
- $E[y_{test}]$ and σy_{test} model a gaussian distribution around y_{test}

Calibration in a nutshell (Kuleshov *et al.*, 2018)

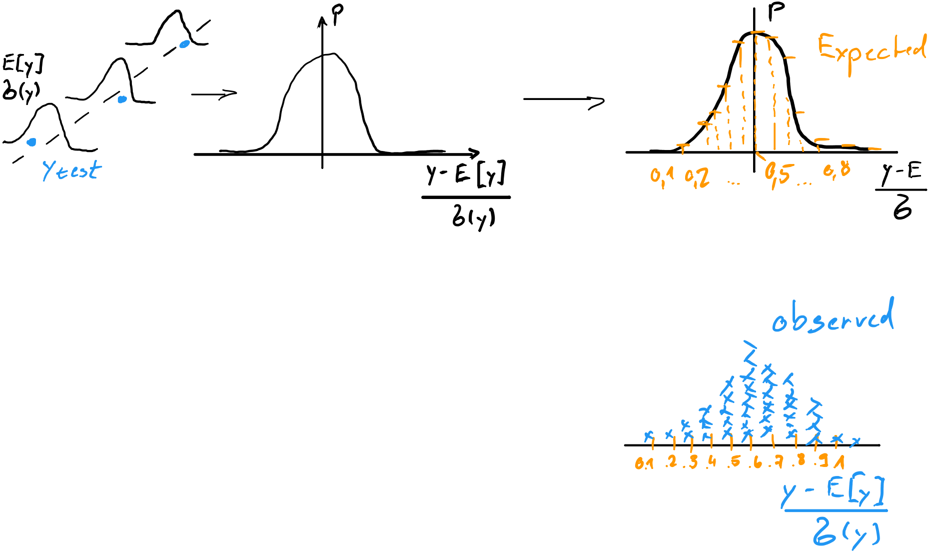
Calibration in a nutshell (Kuleshov *et al.*, 2018)



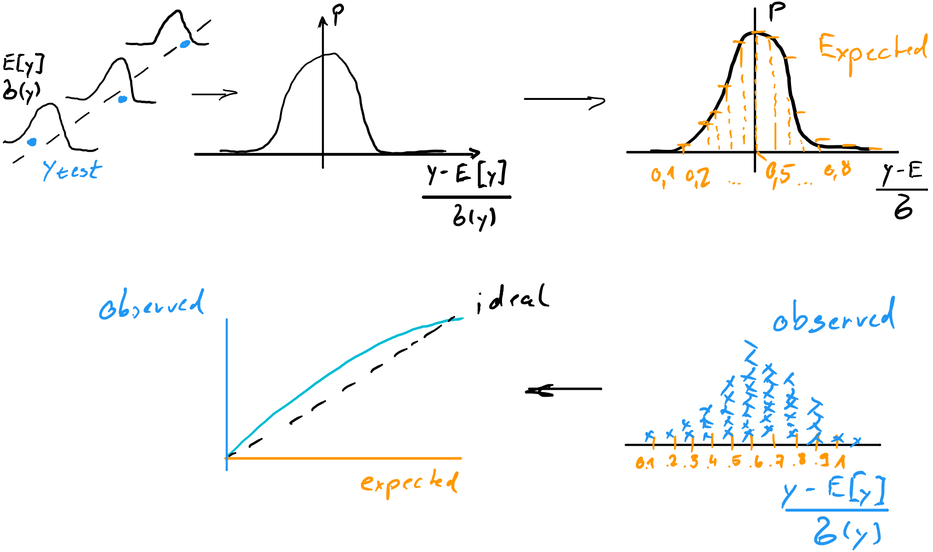
Calibration in a nutshell (Kuleshov *et al.*, 2018)



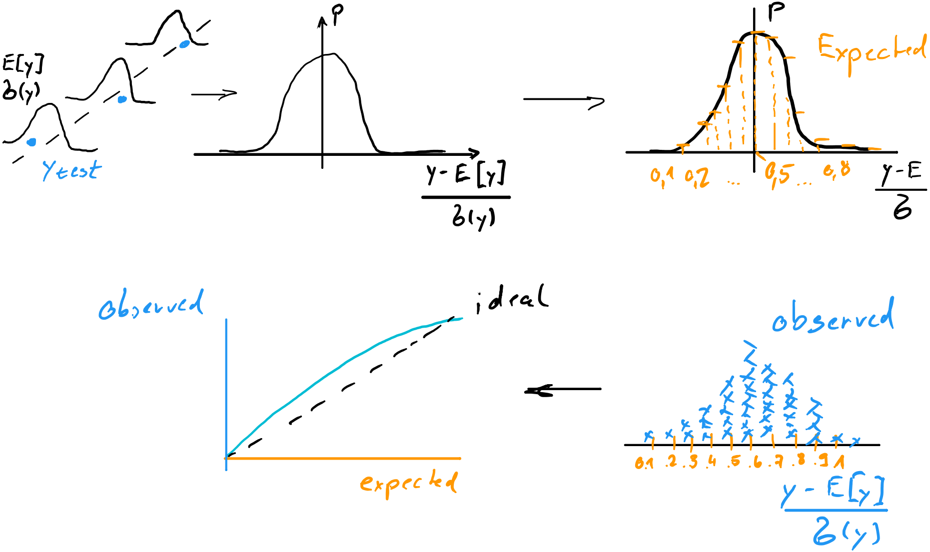
Calibration in a nutshell (Kuleshov *et al.*, 2018)



Calibration in a nutshell (Kuleshov *et al.*, 2018)



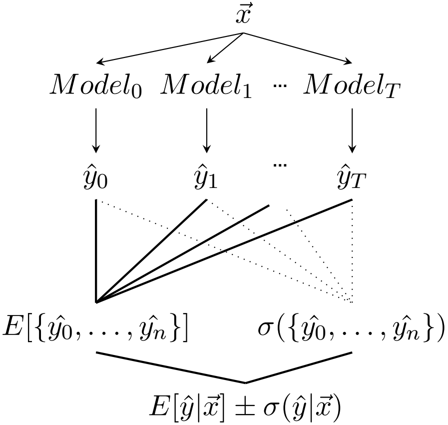
Calibration in a nutshell (Kuleshov *et al.*, 2018)



Note: only sufficient criteria for uncertainty quality (Levi *et al.*, 2020)!

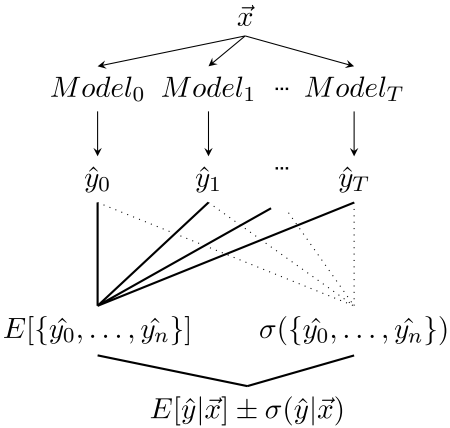
Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)

Deep Ensemble?



Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)

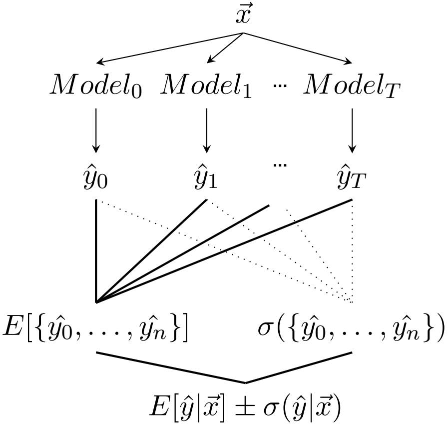
Deep Ensemble?



- use existing model and training loop
- train T models (different seeds)

Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)

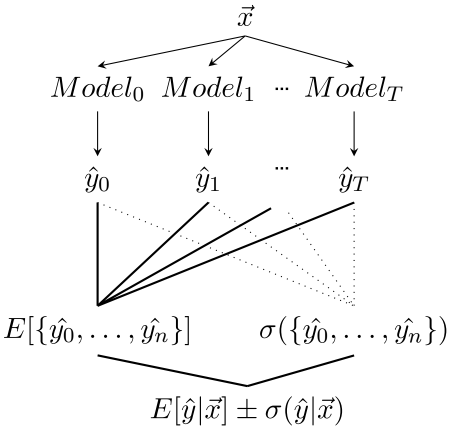
Deep Ensemble?



- use existing model and training loop
- train T models (different seeds)
- **trick:** for each $\vec{x}$ get one prediction from each model

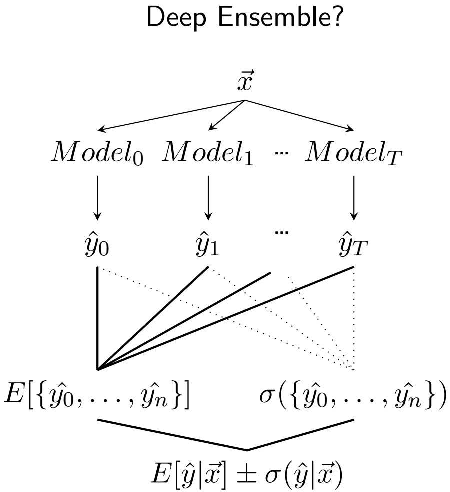
Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)

Deep Ensemble?



- use existing model and training loop
- train T models (different seeds)
- **trick:** for each $\vec{x}$ get one prediction from each model
- obtain mean prediction $E[\hat{y}]$ and prediction deviation $\sigma(\hat{y})$

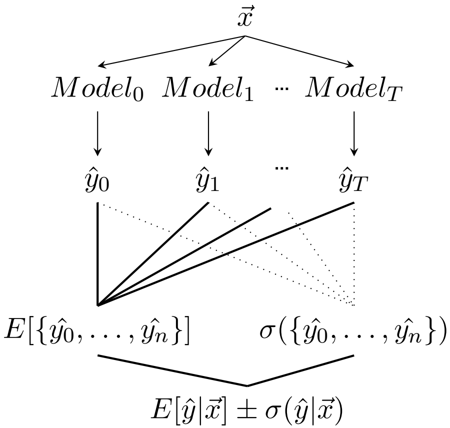
Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)



- use existing model and training loop
- train T models (different seeds)
- **trick:** for each $\vec{x}$ get one prediction from each model
- obtain mean prediction $E[\hat{y}]$ and prediction deviation $\sigma(\hat{y})$
- **epistemic uncertainty**

Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)

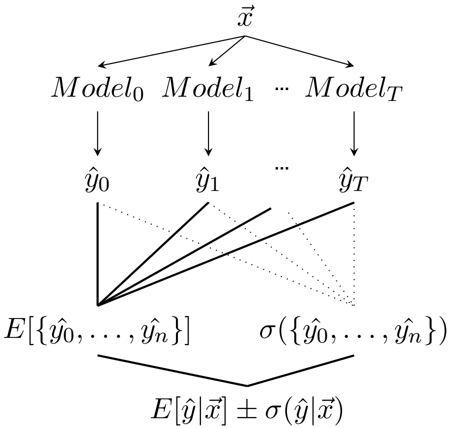
Deep Ensemble?



- use existing model and training loop
- train T models (different seeds)
- **trick:** for each $\vec{x}$ get one prediction from each model
- obtain mean prediction $E[\hat{y}]$ and prediction deviation $\sigma(\hat{y})$
- **epistemic uncertainty**

Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)

Deep Ensemble?



- use existing model and training loop
- train T models (different seeds)
- **trick:** for each $\vec{x}$ get one prediction from each model
- obtain mean prediction $E[\hat{y}]$ and prediction deviation $\sigma(\hat{y})$
- **epistemic uncertainty**

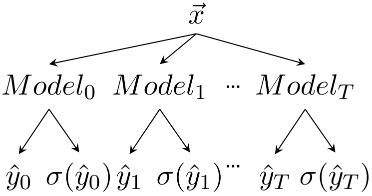
Our nickname: "Simple Ensembles"!

Try Simple Ensembles for yourself

Please open `simpleensembles_1D_regression_vanilla.ipynb`!
Go through the notebook.

Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)

Deep Ensemble!



Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)

Deep Ensemble!

$$
\begin{tikzcd}
 & & & \vec{x} \arrow[dlll] \arrow[dl] \arrow[drr] & & & \\
Model_0 \arrow[d] \arrow[dr] & & Model_1 \arrow[d] \arrow[dr] & \cdots & & Model_T \arrow[d] \arrow[dr] & \\
\hat{y}_0 & \sigma(\hat{y}_0) & \hat{y}_1 & \sigma(\hat{y}_1) & \cdots & \hat{y}_T & \sigma(\hat{y}_T)
\end{tikzcd}
$$

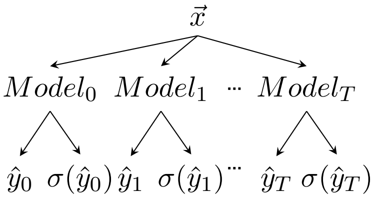
$$l_{nll} = -\log p = \frac{\log \sigma(\hat{y})^2}{2} + \frac{1}{2}\frac{(y-\hat{y})^2}{\sigma(\hat{y})^2}$$

$$E_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T} \hat{y}_t$$

$$\sigma_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T} (\sigma(\hat{y}_t)^2 + \hat{y}_t^2) - E_{ens}[y_n]^2$$

Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)

Deep Ensemble!



$$l_{nll} = -\log p = \frac{\log \sigma(\hat{y})^2}{2} + \frac{1}{2}\frac{(y-\hat{y})^2}{\sigma(\hat{y})^2}$$

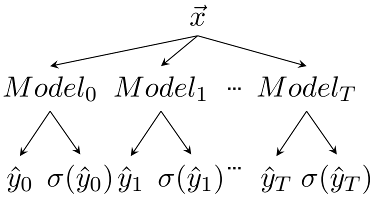
$$E_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T}\hat{y}_t$$

$$\sigma_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T}(\sigma(\hat{y}_t)^2 + \hat{y}_t^2) - E_{ens}[y_n]^2$$

- use expanded model to predict expectation and std deviation (mean variance estimation, MVE)
- train T models (different seeds)

Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)

Deep Ensemble!



$$l_{nll} = -\log p = \frac{\log \sigma(\hat{y})^2}{2} + \frac{1}{2}\frac{(y-\hat{y})^2}{\sigma(\hat{y})^2}$$

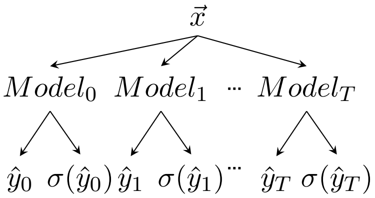
$$E_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T}\hat{y}_t$$

$$\sigma_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T}(\sigma(\hat{y}_t)^2 + \hat{y}_t^2) - E_{ens}[y_n]^2$$

- use expanded model to predict expectation and std deviation (mean variance estimation, MVE)
- train T models (different seeds)
- **trick:** for each $\vec{x}$ get one prediction and one sigma from each model

Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)

Deep Ensemble!



$$l_{nll} = -\log p = \frac{\log \sigma(\hat{y})^2}{2} + \frac{1}{2}\frac{(y-\hat{y})^2}{\sigma(\hat{y})^2}$$

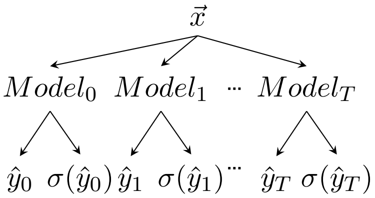
$$E_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T} \hat{y}_t$$

$$\sigma_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T}(\sigma(\hat{y}_t)^2 + \hat{y}_t^2) - E_{ens}[y_n]^2$$

- use expanded model to predict expectation and std deviation (mean variance estimation, MVE)
- train T models (different seeds)
- **trick:** for each $\vec{x}$ get one prediction and one sigma from each model
- allows to learn model of **aleatoric uncertainty** (MVE)

Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)

Deep Ensemble!



$$l_{nll} = -\log p = \frac{\log \sigma(\hat{y})^2}{2} + \frac{1}{2}\frac{(y-\hat{y})^2}{\sigma(\hat{y})^2}$$

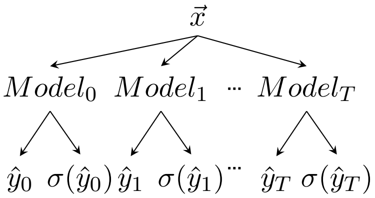
$$E_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T}\hat{y}_t$$

$$\sigma_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T}(\sigma(\hat{y}_t)^2 + \hat{y}_t^2) - E_{ens}[y_n]^2$$

- use expanded model to predict expectation and std deviation (mean variance estimation, MVE)
- train T models (different seeds)
- **trick:** for each $\vec{x}$ get one prediction and one sigma from each model
- allows to learn model of **aleatoric uncertainty** (MVE)
- ensemble allows to assess **epistemic uncertainty**

Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)

Deep Ensemble!



$$l_{nll} = -\log p = \frac{\log \sigma(\hat{y})^2}{2} + \frac{1}{2}\frac{(y-\hat{y})^2}{\sigma(\hat{y})^2}$$

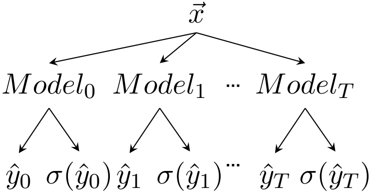
$$E_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T} \hat{y}_t$$

$$\sigma_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T}(\sigma(\hat{y}_t)^2 + \hat{y}_t^2) - E_{ens}[y_n]^2$$

- use expanded model to predict expectation and std deviation (mean variance estimation, MVE)
- train T models (different seeds)
- **trick:** for each $\vec{x}$ get one prediction and one sigma from each model
- allows to learn model of **aleatoric uncertainty** (MVE)
- ensemble allows to assess **epistemic uncertainty**
- but: different loss, different model

Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)

Deep Ensemble!



$$l_{nll} = -\log p = \frac{\log \sigma(\hat{y})^2}{2} + \frac{1}{2}\frac{(y-\hat{y})^2}{\sigma(\hat{y})^2}$$

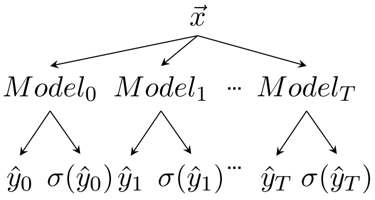
$$E_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T} \hat{y}_t$$

$$\sigma_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T}(\sigma(\hat{y}_t)^2 + \hat{y}_t^2) - E_{ens}[y_n]^2$$

- use expanded model to predict expectation and std deviation (mean variance estimation, MVE)
- train T models (different seeds)
- **trick:** for each $\vec{x}$ get one prediction and one sigma from each model
- allows to learn model of **aleatoric uncertainty** (MVE)
- ensemble allows to assess **epistemic uncertainty**
- but: different loss, different model

Obtaining Uncertainties in Ensembles (Lakshminarayanan *et al.*, 2017)

Deep Ensemble!



$$l_{nll} = -\log p = \frac{\log \sigma(\hat{y})^2}{2} + \frac{1}{2}\frac{(y-\hat{y})^2}{\sigma(\hat{y})^2}$$

$$E_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T} \hat{y}_t$$

$$\sigma_{ens}[y_n] = \frac{1}{T}\sum_{t=0}^{T} (\sigma(\hat{y}_t)^2 + \hat{y}_t^2) - E_{ens}[y_n]^2$$

- use expanded model to predict expectation and std deviation (mean variance estimation, MVE)
- train T models (different seeds)
- **trick:** for each $\vec{x}$ get one prediction and one sigma from each model
- allows to learn model of **aleatoric uncertainty** (MVE)
- ensemble allows to assess **epistemic uncertainty**
- but: different loss, different model

These are "Deep Ensembles"!

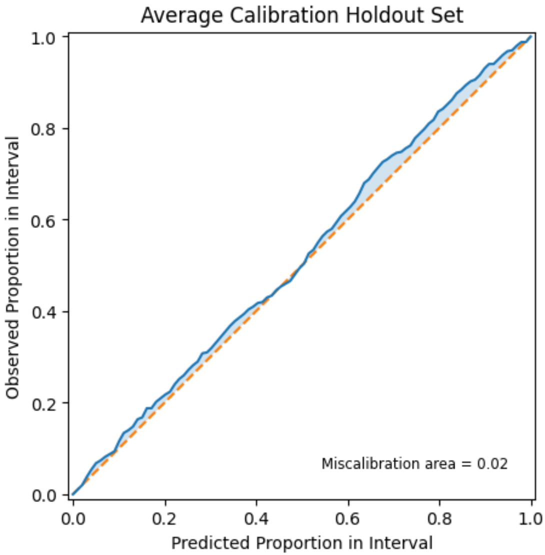
Try Deep Ensembles for yourself

Please open

`deepensembles_1D_regression_vanilla.ipynb`

Go through the notebook.

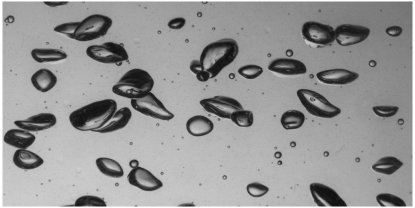
Well Calibrated Uncertainties with DeepEnsembles



Use Case: Predictive Uncertainties for Instance Segmentations

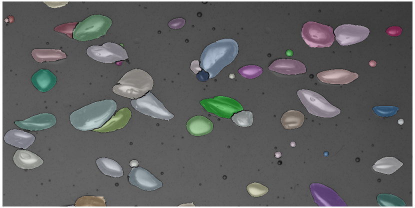
Instance Segmentation Task

Input



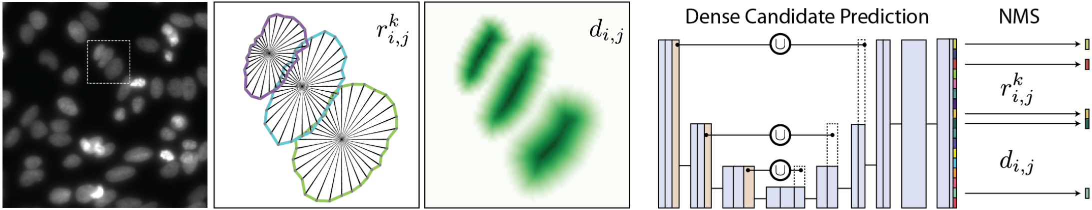
(Hessenkemper *et al.*, 2022)

Labels



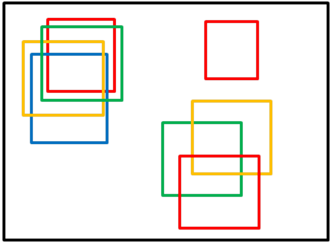
- goal: accurate spatial prediction
- adding uncertainty = reliable and robust prediction

Instance Segmentation Tooling: StarDist



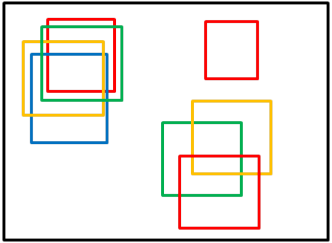
Obtaining Uncertainties for StarDist (Siddiqui *et al.*, 2023)

ensemble predictions
provide multitude of labels

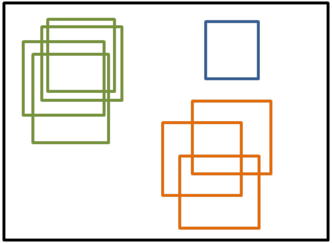


Obtaining Uncertainties for StarDist (Siddiqui *et al.*, 2023)

ensemble predictions
provide multitude of labels

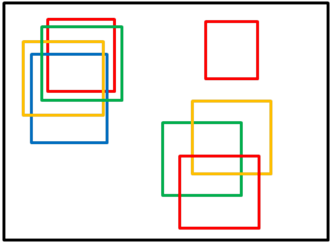


clustering for
homogenous instance
labels

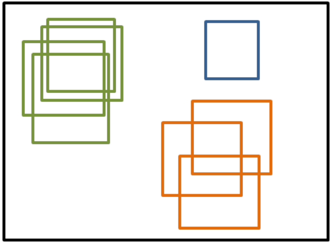


Obtaining Uncertainties for StarDist (Siddiqui *et al.*, 2023)

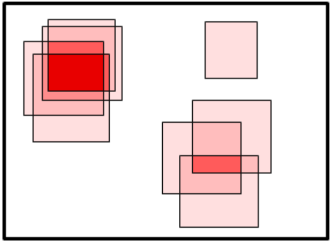
ensemble predictions provide multitude of labels



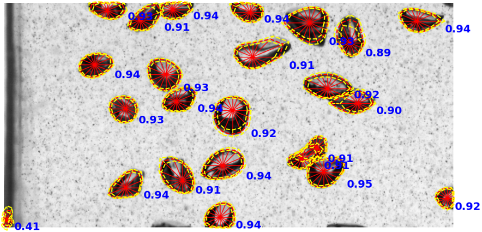
clustering for homogenous instance labels



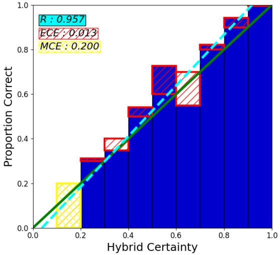
calibrated **certainty scores** by region of most overlap



Informed Predictions with Uncertainties and Calibration Plots

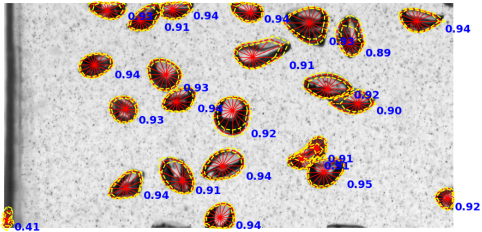


benign segmentation
(bubble segmentation)

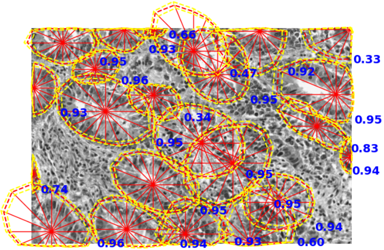


(Siddiqui *et al.*, 2023)

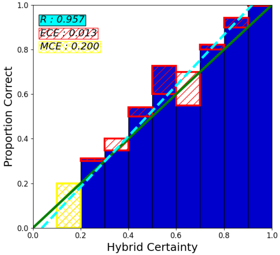
Informed Predictions with Uncertainties and Calibration Plots

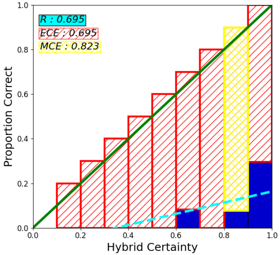


benign segmentation
(bubble segmentation)



malignant segmentation
(gland tumor cell segmentation)





(Siddiqui *et al.*, 2023)

Summary

- **uncertainties express variability** in Machine Learning predictions (induction, limits of dataset)

Summary

- **uncertainties express variability** in Machine Learning predictions (induction, limits of dataset)
- uncertainties should become first class citizens in Machine Learning

Summary

- **uncertainties express variability** in Machine Learning predictions (induction, limits of dataset)
- uncertainties should become first class citizens in Machine Learning
- established methods for obtaining **predictive uncertainties** (MCDropout, SimpleEnsembles, DeepEnsembles)

Summary

- **uncertainties express variability** in Machine Learning predictions (induction, limits of dataset)
- uncertainties should become first class citizens in Machine Learning
- established methods for obtaining **predictive uncertainties** (MCDropout, SimpleEnsembles, DeepEnsembles)
- depending on UQ method chosen, uncertainties refer to different things

  (epistemic vs aleatoric uncertainties for example)

Summary

- **uncertainties express variability** in Machine Learning predictions (induction, limits of dataset)
- uncertainties should become first class citizens in Machine Learning
- established methods for obtaining **predictive uncertainties** (MCDropout, SimpleEnsembles, DeepEnsembles)
- depending on UQ method chosen, uncertainties refer to different things

  (epistemic vs aleatoric uncertainties for example)

Summary

- **uncertainties express variability** in Machine Learning predictions (induction, limits of dataset)
- uncertainties should become first class citizens in Machine Learning
- established methods for obtaining **predictive uncertainties** (MCDropout, SimpleEnsembles, DeepEnsembles)
- depending on UQ method chosen, uncertainties refer to different things

  (epistemic vs aleatoric uncertainties for example)

Thank you for your attention!

Looking forward to questions, feedback and comments.

Bibliography I

Bouthillier, Xavier, Delaunay, Pierre, Bronzi, Mirko, Trofimov, Assya, Nichyporuk, Brennan, Szeto, Justin, Sepah, Naz, Raff, Edward, Madan, Kanika, Voleti, Vikram, Kahou, Samira Ebrahimi, Michalski, Vincent, Serdyuk, Dmitriy, Arbel, Tal, Pal, Chris, Varoquaux, Gaël, & Vincent, Pascal. 2021. Accounting for Variance in Machine Learning Benchmarks. *CoRR*, **abs/2103.03098**.

Gal, Yarin, & Ghahramani, Zoubin. 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. *Pages 1050–1059 of: international conference on machine learning*. PMLR.

Gawlikowski, Jakob, Tassi, Cedrique Rovile Njieutcheu, Ali, Mohsin, Lee, Jongseok, Humt, Matthias, Feng, Jianxiang, Kruspe, Anna, Triebel, Rudolph, Jung, Peter, Roscher, Ribana, Shahzad, Muhammad, Yang, Wen, Bamler, Richard, & Zhu, Xiao Xiang. *A Survey of Uncertainty in Deep Neural Networks*.

Bibliography II

Hessenkemper, Hendrik, Starke, Sebastian, Atassi, Yazan, Ziegenhein, Thomas, & Lucas, Dirk. 2022. Bubble identification from images with machine learning methods. *International Journal of Multiphase Flow*, **155**, 104169.

Howard, Jeremy, *et al.* 2022. *Imagenette*. Commit at time of writing.

Hüllermeier, Eyke, & Waegeman, Willem. 2021. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. **110**(3), 457–506.

Kuleshov, Volodymyr, Fenner, Nathan, & Ermon, Stefano. 2018. Accurate Uncertainties for Deep Learning Using Calibrated Regression. *CoRR*, **abs/1807.00263**.

Lakshminarayanan, Balaji, Pritzel, Alexander, & Blundell, Charles. 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, **30**.

Levi, Dan, Gispan, Liran, Giladi, Niv, & Fetaya, Ethan. 2020. *Evaluating and Calibrating Uncertainty Prediction in Regression Tasks*.

Bibliography III

Liu, Zhuang, Mao, Hanzi, Wu, Chao-Yuan, Feichtenhofer, Christoph, Darrell, Trevor, & Xie, Saining. 2022. A ConvNet for the 2020s. *CoRR*, **abs/2201.03545**.

Park, Namuk, & Kim, Songkuk. 2022. *How Do Vision Transformers Work?*

Raschka, Sebastian. 2018. Model evaluation, model selection, and algorithm selection in machine learning.

Rohatgi, Ankit. 2021. *Webplotdigitizer: Version 4.5*.

Russell, Stuart J, & Norvig, Peter. 2010. *Artificial intelligence a modern approach*. London.

Siddiqui, Qasim M. K., Starke, Sebastian, & Steinbach, Peter. 2023. *Uncertainty Estimation in Instance Segmentation with Star-convex Shapes*. in review at WACV'24.

Steinbach, Peter, Gernhardt, Felicita, Tanveer, Mahnoor, Schmerler, Steve, & Starke, Sebastian. 2022. Machine learning state-of-the-art with uncertainties. *ICML*, **ML Evaluation Standards workshop**.

Bibliography IV

Tan, Aik Rui, Urata, Shingo, Goldman, Samuel, Dietschreit, Johannes C. B., & Gómez-Bombarelli, Rafael. 2023. Single-model uncertainty quantification in neural network potentials does not consistently outperform model ensembles. **9**(1), 1–11. Number: 1 Publisher: Nature Publishing Group.

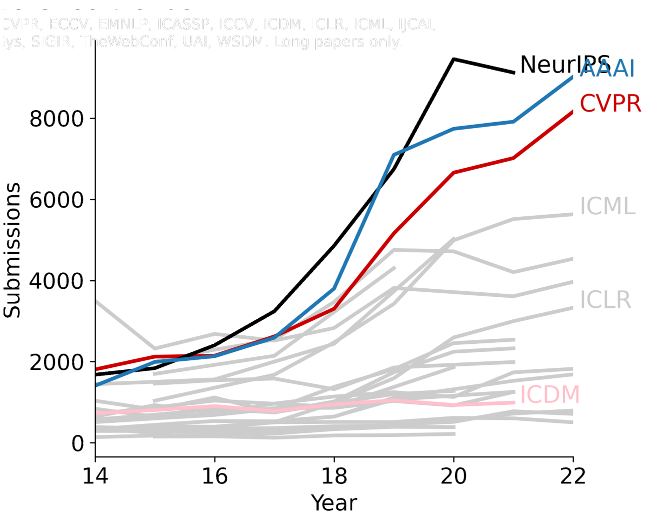
Tomasev, Nenad, Bica, Ioana, McWilliams, Brian, Buesing, Lars, Pascanu, Razvan, Blundell, Charles, & Mitrovic, Jovana. 2022. Pushing the limits of self-supervised ResNets: Can we outperform supervised learning without labels on ImageNet? *CoRR*, **abs/2201.05119**.

Xin, Li. 2022. *Statistics of acceptance rate for the main AI conferences*. Commit at time of writing.

Appendix

Derived Uncertainties

High Demand for Reviewing

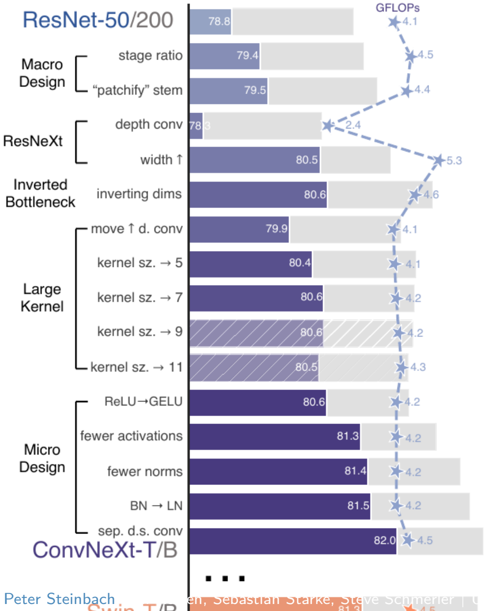


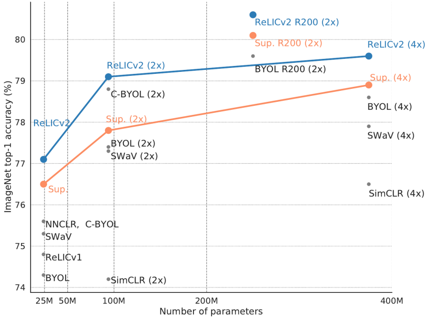
adapted from Xin (2022)

International Conference on Learning Representations 2022 (2023)

- accepted/submitted 1095/3328 (1574/4956)
- acceptance rate 32.9% (32.0%)
- 54 (91) orals
- 176 (280) spotlights
- 865 (1203) posters

State-of-the-Art, SOTA





from Tomasev *et al.* (2022)

- SOTA = (uncertified) reference to check for progress
- accuracy often a central figure of merit
- differences in accuracy small $\mathcal{O}(1\%)$

A classification SOTA for demonstration

image classification on imagenette (Howard *et al.*, 2022)

...

A classification SOTA for demonstration

image classification on imagenette (Howard *et al.*, 2022)




...

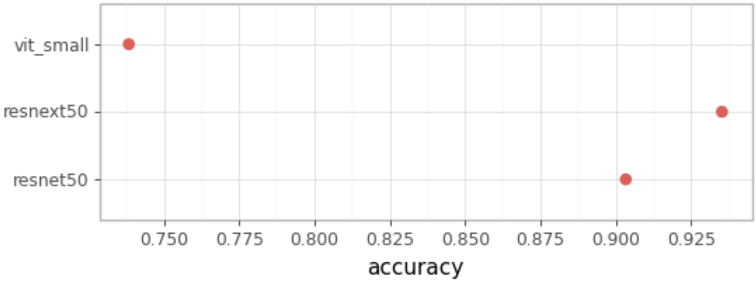


Figure 1 (a): Accuracy estimates on 10-class image classification for three different ML architectures. Taken from (Steinbach *et al.*, 2022)

Accuracies with Uncertainties from Cross-Validation

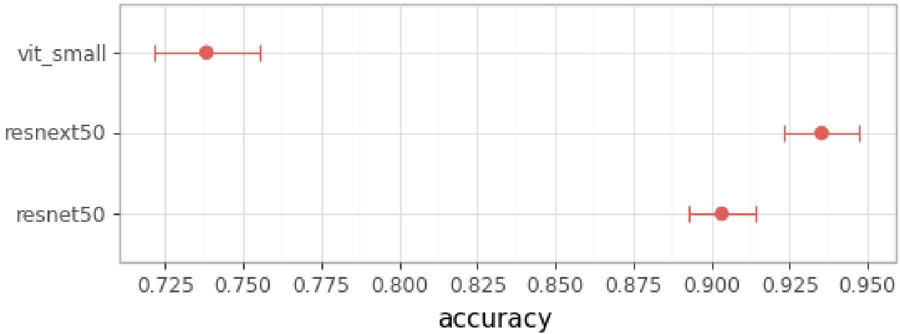


Figure 1 (b): Accuracy estimates on 10-class image classification for three different ML architectures. Point estimates and confidence intervals obtained from 20-fold cross validation is shown. Taken from (Steinbach *et al.*, 2022)

Approximated Uncertainties $\hat{\sigma}$

Approximate Accuracy as a Bernoulli probability

$$\mu_{\mathrm{ACC}} \pm \hat{\sigma}_{\mathrm{ACC}} = \mu_{\mathrm{ACC}} \pm z \sqrt{\frac{1}{n_{holdout}} \mathrm{ACC}_{holdout} \left(1 - \mathrm{ACC}_{holdout}\right)}$$

In the limit of large numbers, this converges to a normal distribution. Use z to construct confidence interval assuming normality.

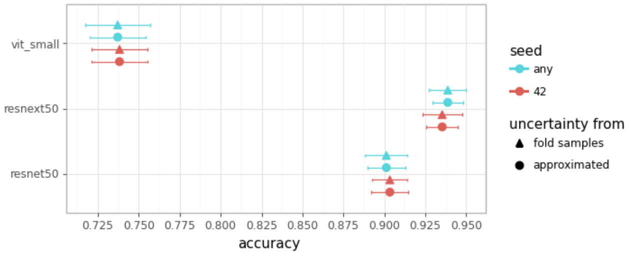
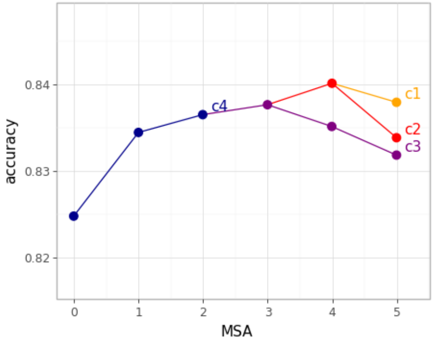


Figure 5: Comparison of fold sample based uncertainty with approximated uncertainty using eq. (1) (Raschka, 2018). Each estimate was obtained for one seed (42) or any seed available (total 6 seeds). The uncertainty plotted for seed 42 was obtained using the approximation in eq. (1). The uncertainty plotted for all seeds was obtained using the sample standard deviation. Taken from (Steinbach *et al.*, 2022)

How Do Vision Transformers Work? (Park & Kim, 2022)



How Do Vision Transformers Work? (Park & Kim, 2022)

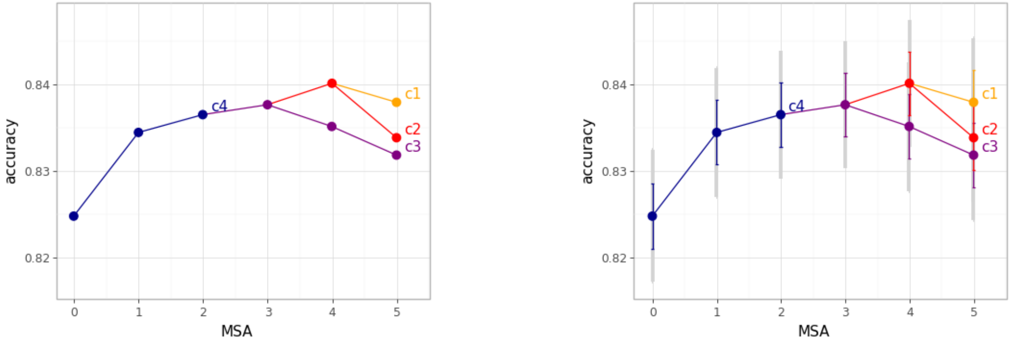


Figure 2: Reproduction of figure 12a from Park & Kim (2022) (left). Augmentation of the same figure with estimated accuracy calculated using eq. (1) using a one-sigma 68.2% (colored) and two-sigma 95% (grey) confidence interval (right). Data to reproduce these figures was obtained by using Rohatgi (2021) on the figures from the preprint PDF. Taken from

(Steinbach *et al.*, 2022)

More Sources of Variance

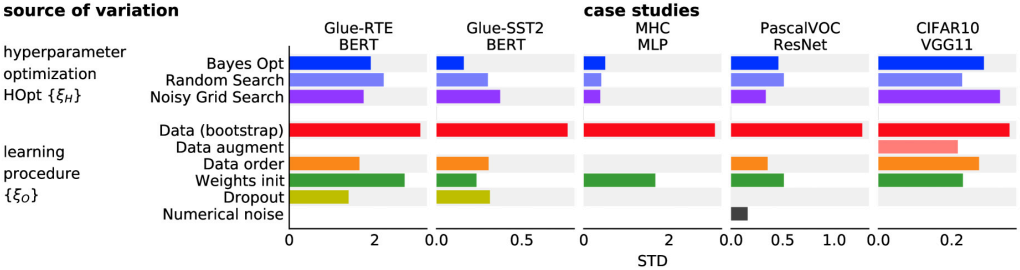


Figure 1 of (Bouthillier *et al.*, 2021): Different sources of variation of the measured performance: across our different case studies, as a fraction of the variance induced by bootstrapping the data. For hyperparameter optimization, we studied several algorithms.

Takeaways: Let's "increase the quality of evidence"[1]

- **uncertainties are essential**

  (strong hint for communicating and reviewing academic results)

[1] G. Varoquaux at ICLR's ML Eval workshop 2022

Takeaways: Let's "increase the quality of evidence"[1]

- **uncertainties are essential**

  (strong hint for communicating and reviewing academic results)

- **uncertainties can be laborious**

  (cross-validation, running training multipe times)

[1] G. Varoquaux at ICLR's ML Eval workshop 2022

Takeaways: Let's "increase the quality of evidence"[1]

- **uncertainties are essential**

  (strong hint for communicating and reviewing academic results)
- **uncertainties can be laborious**

  (cross-validation, running training multipe times)
- **approximations for uncertainties** provide a solution **with minimal runtime cost**

[1] G. Varoquaux at ICLR's ML Eval workshop 2022

Takeaways: Let's "increase the quality of evidence"[1]

- **uncertainties are essential**

  (strong hint for communicating and reviewing academic results)

- **uncertainties can be laborious**

  (cross-validation, running training multipe times)

- **approximations for uncertainties** provide a solution **with minimal runtime cost**

[1] G. Varoquaux at ICLR's ML Eval workshop 2022

Takeaways: Let's "increase the quality of evidence"[1]

- **uncertainties are essential**

  (strong hint for communicating and reviewing academic results)

- **uncertainties can be laborious**

  (cross-validation, running training multipe times)

- **approximations for uncertainties** provide a solution **with minimal runtime cost**

See (Steinbach *et al.*, 2022) for more details!

[1] G. Varoquaux at ICLR's ML Eval workshop 2022