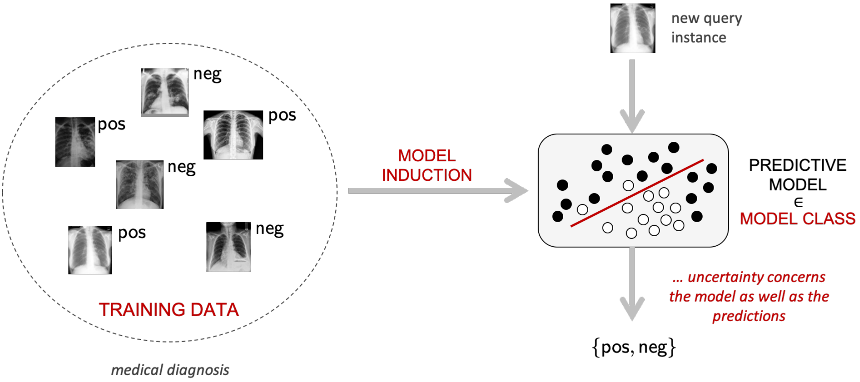
Aleatoric and Epistemic Uncertainty in Machine Learning

Eyke Hüllermeier

Artificial Intelligence and Machine Learning
Institute of Informatics
LMU Munich

German Data Science Days, Munich, March 17, 2021

Machine learning is inseparably connected with uncertainty



Trustworthy machine learning

- Many applications require safe and reliable predictions, and hence a certain level of **self-awareness** of ML systems:
  - equip predictions with an appropriate **quantification of uncertainty**,
  - **reject** a decision in cases of high uncertainty (abstention) ,
  - deliver a credible **set-valued prediction** (partial abstention),
  - ...

Driver assistance systems: a safety-critical application

Lack of uncertainty-awareness

Example of a lack of “uncertainty-awareness”:

Predictions by EfficientNet (Tan and Le, 2019) on test images from ImageNet: For the left image, the neural network predicts “typewriter keyboard” with certainty 83.14 %, for the right image “stone wall” with certainty 87.63 %.

Aleatoric versus epistemic uncertainty

- Traditional approaches in ML fail to distinguish inherently different sources of uncertainty, often referred to as **aleatoric** and **epistemic** uncertainty (Hora, 1996; Der Kiureghian and Ditlevsen, 2009).
- Motivated in the context of ML for medical diagnosis by Senge et al. (2014), increasing attention more recently due to interest by the deep learning community (Kendall and Gal, 2017).

ELSEVIER

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Reliable classification: Learning classifiers that distinguish aleatoric and epistemic uncertainty

CrossMark

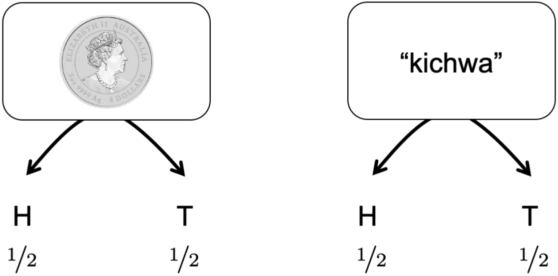
Robin Senge [a], Stefan Bösner [b], Krzysztof Dembczyński [c], Jörg Haasenritter [b], Oliver Hirsch [b], Norbert Donner-Banzhoff [b], Eyke Hüllermeier [a,*]

[a] *Department of Mathematics and Computer Science, University of Marburg, Hans-Meerwein-Str., 35032 Marburg, Germany*
[b] *Department of Family Medicine, University of Marburg, Karl-von-Frisch-Str. 4, 35043 Marburg, Germany*
[c] *Institute of Computing Science, Poznań University of Technology, Piotrowo 2, 60-965 Poznan, Poland*

Aleatoric versus epistemic uncertainty

- **Aleatoric** (*aka* statistical) uncertainty refers to the notion of **randomness**, that is, the variability in the outcome of an experiment which is due to inherently random effects.
- **Epistemic** (*aka* systematic) uncertainty refers to uncertainty caused by a **lack of knowledge**, i.e., to the epistemic state of the agent.
- As opposed to aleatoric uncertainty, epistemic uncertainty can in principle be reduced on the basis of additional information.

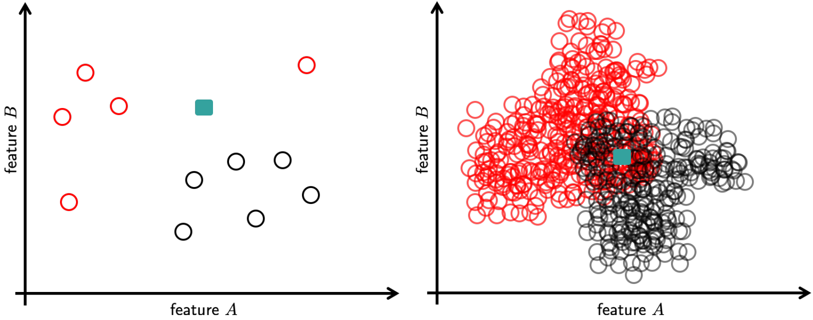
Aleatoric versus epistemic uncertainty



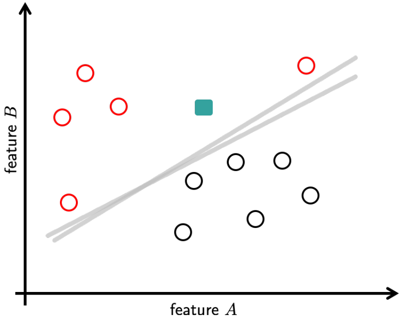
“Not knowing the chance of mutually exclusive events and knowing the chance to be equal are two quite different states of knowledge"

Ronald Fisher (1890-1962)

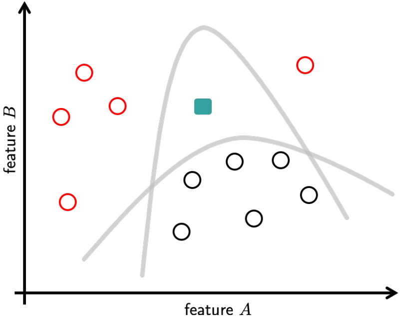
Aleatoric versus epistemic uncertainty in ML



Aleatoric versus epistemic uncertainty in ML

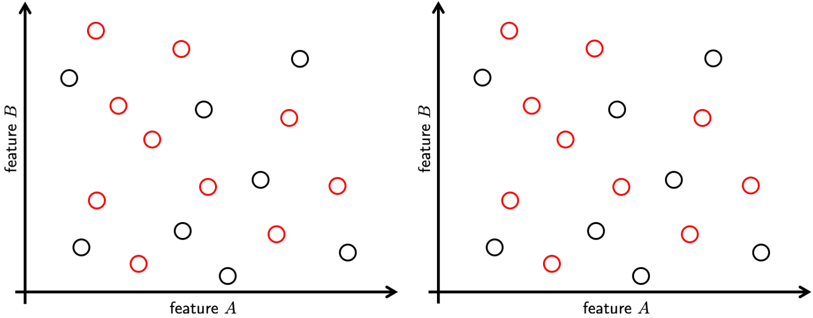


strong prior



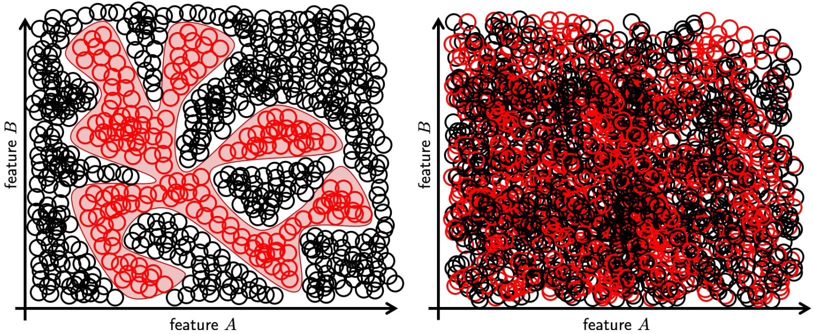
weaker prior

Aleatoric versus epistemic uncertainty in ML



Is the uncertainty aleatoric or epistemic?

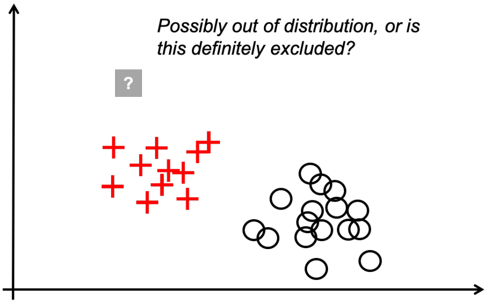
Aleatoric versus epistemic uncertainty in ML



Like random versus pseudo-random numbers ...

Problem setting and assumptions

- A precise specification of the **problem setting** and **underlying assumptions** is an important prerequisite, not only for providing learning guarantees, but also for **uncertainty quantification**.



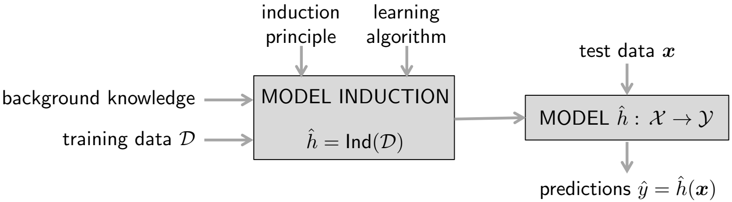
- Here, one might be quite sure about the class of the query under standard assumptions of binary classification, but much less so in a setting of **novelty detection**, where new classes may emerge.

Supervised learning and predictive uncertainty

- Uncertainty occurs in various facets in machine learning, and different **settings and learning problems** will usually require a different handling from an uncertainty modeling point of view.

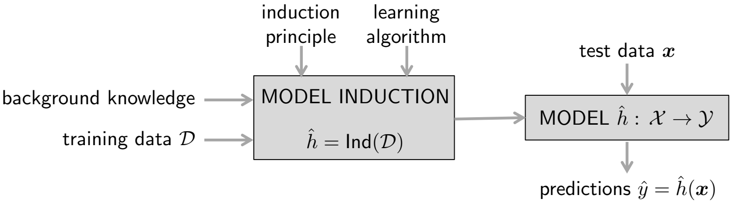
Supervised learning and predictive uncertainty

- Uncertainty occurs in various facets in machine learning, and different **settings and learning problems** will usually require a different handling from an uncertainty modeling point of view.
- Here, we focus on the standard setting of **supervised learning** and **predictive uncertainty**:



Supervised learning and predictive uncertainty

- Uncertainty occurs in various facets in machine learning, and different **settings and learning problems** will usually require a different handling from an uncertainty modeling point of view.
- Here, we focus on the standard setting of **supervised learning** and **predictive uncertainty**:



- Assuming probabilistic data generation $\mathbf{P}(\boldsymbol{x}, y) = \mathbf{P}(\boldsymbol{x})\mathbf{P}(y \mid \boldsymbol{x})$, **probabilistic predictors** (estimating $\mathbf{P}(y \mid \boldsymbol{x})$) are natural primitives.

Supervised learning and predictive uncertainty

- A learner is given access to a set of (i.i.d.) **training data**

$$\mathcal{D} := \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\} \subset \mathcal{X} \times \mathcal{Y} \ ,$$

where $\mathcal{X}$ is an instance space and $\mathcal{Y}$ the set of outcomes.

Supervised learning and predictive uncertainty

- A learner is given access to a set of (i.i.d.) **training data**

$$\mathcal{D} := \left\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\right\} \subset \mathcal{X} \times \mathcal{Y} \ ,$$

  where $\mathcal{X}$ is an instance space and $\mathcal{Y}$ the set of outcomes.
- Given a **hypothesis space** $\mathcal{H} \subset \mathcal{Y}^{\mathcal{X}}$ and a loss function

$$\ell : \mathcal{Y} \times \mathcal{Y} \longrightarrow \mathbb{R} \,,$$

  the goal of the learner is to induce a hypothesis $h^* \in \mathcal{H}$ with low risk

$$R(h) := \int_{\mathcal{X} \times \mathcal{Y}} \ell(h(\boldsymbol{x}), y) \, d\, \mathbf{P}(\boldsymbol{x}, y) \ .$$

Supervised learning and predictive uncertainty

- A learner is given access to a set of (i.i.d.) **training data**

$$\mathcal{D} := \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\} \subset \mathcal{X} \times \mathcal{Y} \ ,$$

  where $\mathcal{X}$ is an instance space and $\mathcal{Y}$ the set of outcomes.

- Given a **hypothesis space** $\mathcal{H} \subset \mathcal{Y}^{\mathcal{X}}$ and a loss function

$$\ell : \mathcal{Y} \times \mathcal{Y} \longrightarrow \mathbb{R} \, ,$$

  the goal of the learner is to induce a hypothesis $h^* \in \mathcal{H}$ with low risk

$$R(h) := \int_{\mathcal{X} \times \mathcal{Y}} \ell(h(\boldsymbol{x}), y) \, d\, \mathbf{P}(\boldsymbol{x}, y) \ .$$

- The learner's choice is commonly guided by the **empirical risk**

$$R_{emp}(h) := \frac{1}{N} \sum_{i=1}^{N} \ell(h(\boldsymbol{x}_i), y_i) \ .$$

Supervised learning and predictive uncertainty

- Yet, since $R_{emp}(h)$, or any variant $\hat{R}_{emp}$, is only an estimation of the true risk $R(h)$, the hypothesis (e.g., the **ERM**)

$$\hat{h} := \arg\min_{h \in \mathcal{H}} \hat{R}_{emp}(h)$$

will normally not coincide with the true **risk minimizer**

$$h^* := \arg\min_{h \in \mathcal{H}} R(h) \, .$$

Supervised learning and predictive uncertainty

- Yet, since $R_{emp}(h)$, or any variant $\hat{R}_{emp}$, is only an estimation of the true risk $R(h)$, the hypothesis (e.g., the **ERM**)

$$\hat{h} := \arg\min_{h \in \mathcal{H}} \hat{R}_{emp}(h)$$

will normally not coincide with the true **risk minimizer**

$$h^* := \arg\min_{h \in \mathcal{H}} R(h)\,.$$

- Correspondingly, there remains **uncertainty** regarding h^* as well as the approximation quality of $\hat{h}$ (in the sense of its proximity to h^*) and its true risk $R(\hat{h})$.

Supervised learning and predictive uncertainty

- Yet, since $R_{emp}(h)$, or any variant $\hat{R}_{emp}$, is only an estimation of the true risk $R(h)$, the hypothesis (e.g., the **ERM**)

$$\hat{h} := \arg\min_{h \in \mathcal{H}} \hat{R}_{emp}(h)$$

will normally not coincide with the true **risk minimizer**

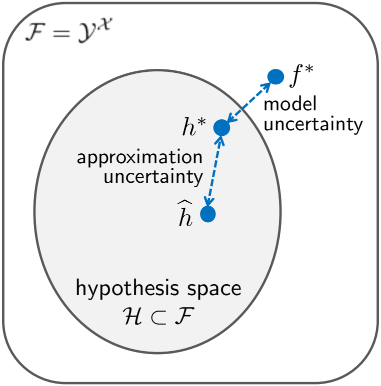
$$h^* := \arg\min_{h \in \mathcal{H}} R(h) \, .$$

- Correspondingly, there remains **uncertainty** regarding h^* as well as the approximation quality of $\hat{h}$ (in the sense of its proximity to h^*) and its true risk $R(\hat{h})$.
- Eventually, one is often interested in the **predictive uncertainty**, i.e., the uncertainty related to the prediction $\hat{y}_q$ for an **individual** (query) instance $\boldsymbol{x}_q \in \mathcal{X}$.

Agenda

1. Introduction
2. **Sources of uncertainty in supervised learning**
3. Modeling approximation uncertainty
4. Ensemble methods for uncertainty quantification
5. Conclusion and outlook

Sources of uncertainty



	point prediction	probability
ground truth	$f^*(\boldsymbol{x})$	$\mathbf{p}(\cdot \mid \boldsymbol{x})$
best possible	$h^*(\boldsymbol{x})$	$\mathbf{p}(\cdot \mid \boldsymbol{x}, h^*)$
induced predictor	$\hat{h}(\boldsymbol{x})$	$\mathbf{p}(\cdot \mid \boldsymbol{x}, \hat{h})$

Sources of uncertainty

- A query instance $\boldsymbol{x}_q$ gives rise to a conditional probability on $\mathcal{Y}$:

$$\mathbf{p}(y \mid \boldsymbol{x}_q) = \frac{\mathbf{p}(\boldsymbol{x}_q, y)}{\mathbf{p}(\boldsymbol{x}_q)}$$

Sources of uncertainty

- A query instance $\boldsymbol{x}_q$ gives rise to a conditional probability on $\mathcal{Y}$:

$$\mathbf{p}(y \,|\, \boldsymbol{x}_q) = \frac{\mathbf{p}(\boldsymbol{x}_q, y)}{\mathbf{p}(\boldsymbol{x}_q)}$$

- Thus, even given full information in the form of the measure $\mathbf{P}$ (and its density $\mathbf{p}$), uncertainty about the actual outcome y remains.
- This uncertainty is of an **aleatoric** nature.

Sources of uncertainty

- A query instance $\boldsymbol{x}_q$ gives rise to a conditional probability on $\mathcal{Y}$:

$$\mathbf{p}(y \,|\, \boldsymbol{x}_q) = \frac{\mathbf{p}(\boldsymbol{x}_q, y)}{\mathbf{p}(\boldsymbol{x}_q)}$$

- Thus, even given full information in the form of the measure $\mathbf{P}$ (and its density $\mathbf{p}$), uncertainty about the actual outcome y remains.
- This uncertainty is of an **aleatoric** nature.
- The best point predictions (minimizing expected loss) are prescribed by the **pointwise Bayes predictor** f^*:

$$f^*(\boldsymbol{x}) := \arg\min_{\hat{y} \in \mathcal{Y}} \int_{\mathcal{Y}} \ell(y, \hat{y}) \, d\mathbf{P}(y \,|\, \boldsymbol{x}) \, .$$

Sources of uncertainty

- The **Bayes predictor** does not necessarily coincide with the pointwise Bayes predictor.
- This discrepancy between h^* and f^* is connected to the uncertainty regarding the **right type of model** to be fit, and hence the choice of the hypothesis space $\mathcal{H}$.
- We shall refer to this uncertainty as **model uncertainty**.

Sources of uncertainty

- The **Bayes predictor** does not necessarily coincide with the pointwise Bayes predictor.
- This discrepancy between h^* and f^* is connected to the uncertainty regarding the **right type of model** to be fit, and hence the choice of the hypothesis space $\mathcal{H}$.
- We shall refer to this uncertainty as **model uncertainty**.
- Due to model uncertainty, one cannot guarantee

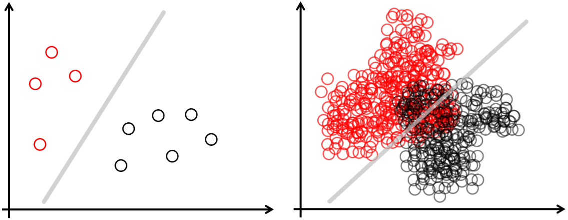
$$h^*(\boldsymbol{x}) = f^*(\boldsymbol{x}) \,,$$

or, in the case of probabilistic predictions $\mathbf{p}(y \,|\, \boldsymbol{x}, h^*)$, that

$$\mathbf{p}(\cdot \,|\, \boldsymbol{x}, h^*) = \mathbf{p}(\cdot \,|\, \boldsymbol{x}) \,.$$

Sources of uncertainty

- Hypothesis $\hat{h}$ produced by the learner is an estimate of h^*.
- The quality of this estimate strongly depends on the quality and the amount of training data.



- We refer to the uncertainty about the discrepancy between $\hat{h}$ and h^* as **approximation uncertainty**.
- Both model and approximation uncertainty are of **epistemic** nature.

Reducible versus irreducible uncertainty

- One way to characterize uncertainty as **aleatoric** or **epistemic** is to ask whether or not it can be reduced through additional information.
- Aleatoric uncertainty refers to the **irreducible** part of the uncertainty, which is due to the stochastic dependency between instances $\boldsymbol{x}$ and outcomes y.

flipping a biased coin

- Model uncertainty and approximation uncertainty are subsumed under the notion of epistemic uncertainty, that is, uncertainty due to a **lack of knowledge** about the perfect predictor.
- In principle, these uncertainties are **reducible**.

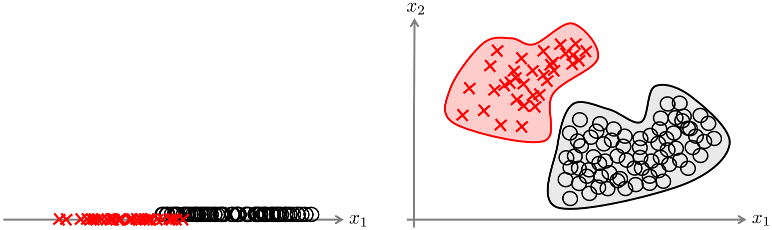
Reducible versus irreducible uncertainty

- But what does “reducible” actually mean?
- An obvious source of additional information is the **training data** $\mathcal{D}$: Uncertainty can be reduced by observing more data, ...
- ... while the problem setting $(\mathcal{X}, \mathcal{Y}, \mathcal{H}, \mathbf{P})$ remains fixed.

Reducible versus irreducible uncertainty

- But what does "reducible" actually mean?
- An obvious source of additional information is the **training data** $\mathcal{D}$: Uncertainty can be reduced by observing more data, ...
- ... while the problem setting $(\mathcal{X}, \mathcal{Y}, \mathcal{H}, \mathbf{P})$ remains fixed.
- In practice, this is of course not always the case.
- For example, a learner may decide to extend the description of instances by **additional features**, thereby replacing the current instance space $\mathcal{X}$ by another space $\mathcal{X}'$.
- Thus, aleatoric and epistemic uncertainty should not be seen as absolute notions. Instead, they are **context-dependent** in the sense of depending on the setting $(\mathcal{X}, \mathcal{Y}, \mathcal{H}, \mathbf{P})$.

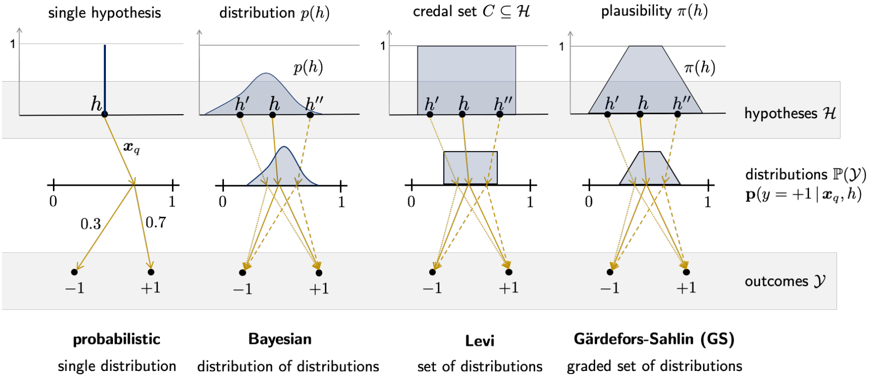
Reducible versus irreducible uncertainty



Left: The two classes are overlapping, which causes (aleatoric) uncertainty in a certain region of the instance space. Right: By adding a second feature, and hence embedding the data in a higher-dimensional space, the two classes become separable, and the uncertainty can be resolved.

Agenda

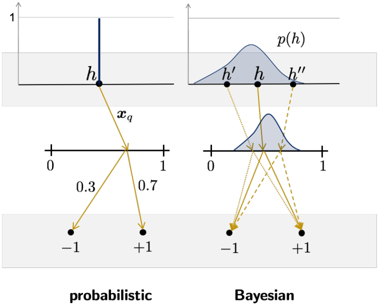
Approaches for representing uncertainty in ML



Approaches for representing uncertainty in ML

single hypothesis

single distribution
on $\mathcal{Y} = \{0, 1\}$



distribution $p(h)$

distribution of
distributions

single distribution
through BME

Bayesian agents

- Explicit attempts at **uncertainty quantification** separating between aleatoric and epistemic uncertainty were made by Mobiny et al. (2017) and Depeweg et al. (2018).
- Here, in the context of regression with DNNs, epistemic uncertainty corresponds to uncertainty about network weights, but the idea can be generalized toward other models.

Bayesian agents

- Explicit attempts at **uncertainty quantification** separating between aleatoric and epistemic uncertainty were made by Mobiny et al. (2017) and Depeweg et al. (2018).

- Here, in the context of regression with DNNs, epistemic uncertainty corresponds to uncertainty about network weights, but the idea can be generalized toward other models.

- Measuring total uncertainty in a prediction $Y = Y|X$ in terms of Shannon entropy of $\mathbf{p} = h(\boldsymbol{x})$,

$$S(Y) = S(\,\mathbf{p}\,) = -\sum_{y \in \mathcal{Y}} \mathbf{p}(y) \log_2 \mathbf{p}(y)\,,$$

  the idea is to exploit the following information-theoretic decomposition:

$$\underbrace{S(Y)}_{total\ \ uncertainty} \quad = \quad \underbrace{I(Y,H)}_{epistemic} + \underbrace{S(Y \mid H)}_{aleatoric}\,.$$

Bayesian agents

- $I(Y, H)$ is the **mutual information** between hypotheses and outcomes (i.e., Kullback-Leibler divergence between joint distribution of outcomes and hypotheses and product of marginals):

$$I(Y, H) = \mathbf{E}_{p(y,h)} \left\{ \log_2 \left(\frac{p(y,h)}{p(y)p(h)} \right) \right\} .$$

- Intuitively, epistemic uncertainty thus captures the amount of information about the hypothesis that would be gained through knowledge of the true outcome y.

- The conditional entropy is given by

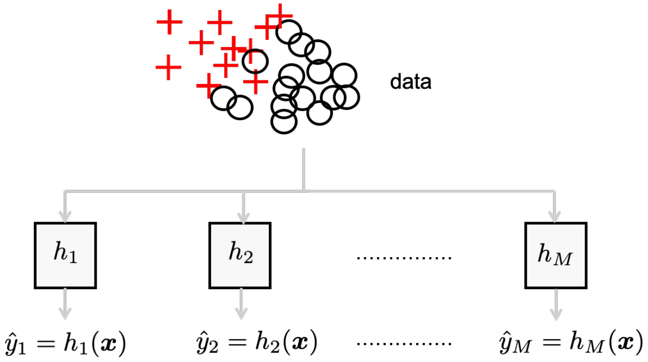
$$S(Y \mid H) = \mathbf{E}_{p(h \mid \mathcal{D})} \left\{ S(\mathbf{p}(y \mid h)) \right\} =$$

$$= - \int_{\mathcal{H}} p(h \mid \mathcal{D}) \left(\sum_{y \in \mathcal{Y}} \mathbf{p}(y \mid h) \log_2 \mathbf{p}(y \mid h) \right) \, d\, h \ .$$

Agenda

1. Introduction
2. Sources of uncertainty in supervised learning
3. Modeling approximation uncertainty
4. **Ensemble methods for uncertainty quantification**
5. Conclusion and outlook

Ensemble methods for uncertainty quantification



- Ensemble can be seen as an approximation of a distribution.
- Intuitively, diversity is an indicator for epistemic uncertainty.

Bayesian agents: Ensemble-based approximation

- Recall what is needed for the proposed uncertainty quantification:
  - Probabilities $\mathbf{p}(y) = \mathbf{p}(y \,|\, \boldsymbol{x})$ to compute entropy $S(Y)$:

$$\mathbf{p}(y) = \int_{\mathcal{H}} \mathbf{p}(y \,|\, h) \, dP(h \,|\, \mathcal{D})$$

  - Expectation for the conditional entropy:

$$S\big(Y \,|\, H\big) = \int_{\mathcal{H}} S(Y \,|\, h) \, dP(h \,|\, \mathcal{D})$$

- The idea is to approximate the **integrals** by (weighted) **averages** over the ensemble members.

Bayesian agents: Ensemble-based approximation

- Based on an ensemble $H = \{h_1, \dots, h_M\}$ of hypotheses, an approximation of **conditional entropy** can be obtained by

$$\mathrm{AU}(\boldsymbol{x}) := -\frac{1}{M} \sum_{i=1}^{M} \left(\sum_{y \in \mathcal{Y}} \mathbf{p}(y \mid h_i) \log_2 \mathbf{p}(y \mid h_i) \right) ,$$

an approximation of **total uncertainty** (Shannon entropy) by

$$\mathrm{U}(\boldsymbol{x}) := -\sum_{y \in \mathcal{Y}} \underbrace{\left(\frac{1}{M} \sum_{i=1}^{M} \mathbf{p}(y \mid h_i) \right)}_{\mathbf{p}(y)} \log_2 \underbrace{\left(\frac{1}{M} \sum_{i=1}^{M} \mathbf{p}(y \mid h_i) \right)}_{\mathbf{p}(y)} ,$$

and an approximation of epistemic uncertainty (mutual information) by the difference, which is equivalent to **Jensen-Shannon divergence** of the distributions $\mathbf{p}(y \mid h_i, \boldsymbol{x})$, $i = 1, \dots, M$.

Bayesian agents: Ensemble-based approximation

	y_1	y_2	$\dots$	y_K	entropy
$h_1(\boldsymbol{x})$	$p_{1,1}$	$p_{1,2}$	$\dots$	$p_{1,K}$	s_1
$h_2(\boldsymbol{x})$	$p_{2,1}$	$p_{2,2}$	$\dots$	$p_{2,K}$	s_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$h_M(\boldsymbol{x})$	$p_{M,1}$	$p_{M,2}$	$\dots$	$p_{M,K}$	s_M
h	p_1	p_2	$\dots$	p_K	$s \mid \bar{s}$

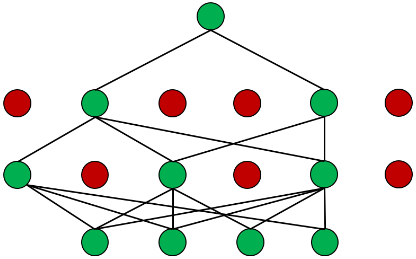
$$\mathrm{U}(\boldsymbol{x}) = s = \text{entropy of average probabilities}$$

$$\mathrm{AU}(\boldsymbol{x}) = \bar{s} = \text{average of entropies}$$

$$\mathrm{EU}(\boldsymbol{x}) = \mathrm{U}(\boldsymbol{x}) - \mathrm{AU}(\boldsymbol{x})$$

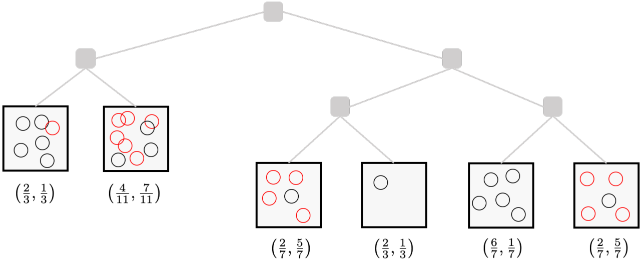
Bayesian agents: Ensemble-based approximation

- For neural networks, it has been shown that techniques such as **Dropout** (Gal and Ghahramani, 2016) and **DropConnect** (Mobiny et al., 2017) can be interpreted as (implicit) ensemble methods, and can hence be used to implement this approach.



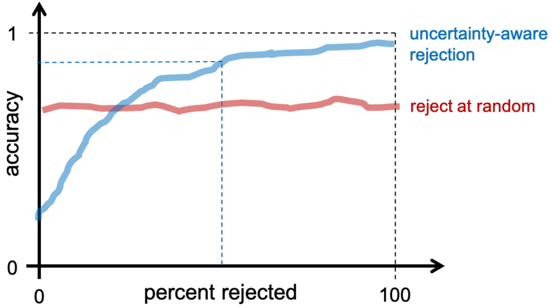
Bayesian agents: Ensemble-based approximation

- Of course, any other ensemble technique could be used as well.
- We proposed an implementation based on **Random Forests**, using decision trees that predict probabilities in terms of (Laplace-corrected) relative frequencies (Shaker and Hüllermeier, 2020).



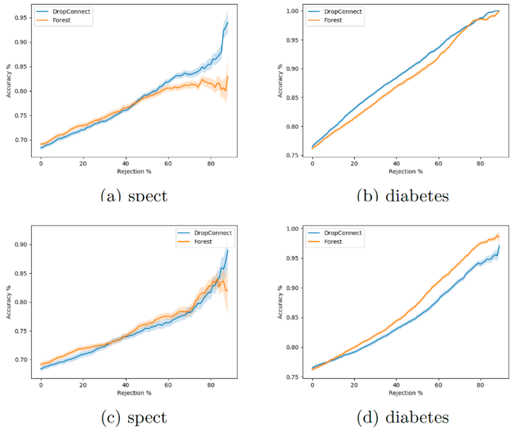
Evaluation

- Quality of uncertainty quantification was evaluated (indirectly) in terms of **accuracy-rejection curves**.



Evaluation

- Quality of uncertainty quantification was evaluated (indirectly) in terms of **accuracy-rejection curves**.
- Results for two approaches, DNN with DropConnect and Random Forests, both for aleatoric (above) and epistemic (below) uncertainty:

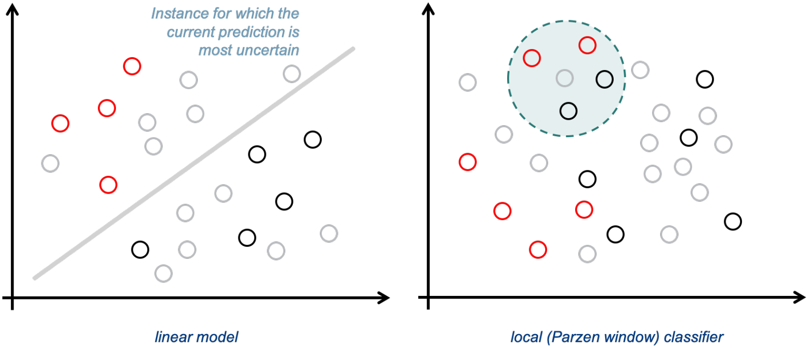


(a) spect (b) diabetes

(c) spect (d) diabetes

Epistemic uncertainty sampling

- The idea of **epistemic uncertainty sampling** is to use a measure of epistemic (instead of total) uncertainty in uncertainty sampling for active learning (Nguyen et al., 2019).



Conclusion and outlook

- We highlighted the importance of uncertainty in ML and the benefits of distinguishing between different types of uncertainty, notably **aleatoric and epistemic**.
- In a Bayesian setting, epistemic uncertainty is reflected by the "peakedness" of the posterior $\mathbf{p}(h \,|\, \mathcal{D})$ on $\mathcal{H}$ resp. $\mathbf{p}(y \,|\, \boldsymbol{x})$ on $\mathcal{Y}$.
- We considered an information-theoretic approach to uncertainty quantification and its realization by means of **ensemble learning**.
- Ongoing work on **generalizations** (Levi and GS agents), building on generalized uncertainty calculi.
- **Model uncertainty** is also important, but difficult to capture.
- Many **applications** can benefit from "uncertainty-informed" decisions.

References

Springer Link

Search Log in

Open Access | Published: 08 March 2021

Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods

Eyke Hüllermeier & Willem Waegeman

Machine Learning (2021) | Cite this article

Download PDF

Sections | Figures | References

Abstract

The notion of uncertainty is of major importance in machine learning and constitutes a key element of machine learning methodology. In line with the statistical tradition, uncertainty has long been perceived as almost synonymous with standard probability and probabilistic predictions. Yet, due to the steadily increasing relevance of machine learning for practical applications and related issues such as safety requirements, new problems and challenges have recently been identified by machine learning scholars, and these problems may call for new methodological developments. In particular, this includes the importance of distinguishing between (at least) two different types of uncertainty, often referred to as *aleatoric* and *epistemic*. In this paper, we provide an introduction to the topic of uncertainty

References

Abellan J, Klir J, Moral S (2006) Disaggregated total uncertainty measure for credal sets. International Journal of General Systems 35(1)

Depeweg S, Hernandez-Lobato J, Doshi-Velez F, Udluft S (2018) Decomposition of uncertainty in Bayesian deep learning for efficient and risk-sensitive learning. In: Proc. ICML, Stockholm, Sweden

Der Kiureghian A, Ditlevsen O (2009) Aleatory or epistemic? does it matter? Structural Safety 31:105–112

Gal Y, Ghahramani Z (2016) Bayesian convolutional neural networks with Bernoulli approximate variational inference. In: Proc. of the ICLR Workshop Track

Hora S (1996) Aleatory and epistemic uncertainty in probability elicitation with an example from hazardous waste management. Reliability Engineering and System Safety 54(2–3):217–223

Kendall A, Gal Y (2017) What uncertainties do we need in Bayesian deep learning for computer vision? In: Proc. NIPS, pp. 5574–5584

Klir G, Mariano M (1987) On the uniqueness of possibilistic measure of uncertainty and information. Fuzzy Sets and Systems 24(2):197–219

Mobiny A, Nguyen H, Moulik S, Garg N, Wu C (2017) DropConnect is effective in modeling uncertainty of Bayesian networks. CoRR abs/1906.04569

Nguyen V, Destercke S, Hüllermeier E (2019) Epistemic uncertainty sampling. In: Proc. DS 2019, 22nd International Conference on Discovery Science, Split, Croatia

Senge R, Bösner S, Dembczynski K, Haasenritter J, Hirsch O, Donner-Banzhoff N, Hüllermeier E (2014) Reliable classification: Learning classifiers that distinguish aleatoric and epistemic uncertainty. Information Sciences 255:16–29

Shaker M, Hüllermeier E (2020) Aleatoric and epistemic uncertainty with random forests. In: Proc. IDA 2020, 18th Int. Symposium on Intelligent Data Analysis, Springer, Konstanz, Germany, pp. 444–456

Tan M, Le Q (2019) EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proc. ICML, Long Beach, California

Zaffalon M (2002) The naive credal classifier. Journal of Statistical Planning and Inference 105(1):5–21