inovex

Are you sure about that!?

Uncertainty Quantification in AI

Florian Wilhelm

Berlin, October 10th 2019

Dr. Florian Wilhelm

Principal Data Scientist @ inovex

@FlorianWilhelm
FlorianWilhelm
florianwilhelm.info

- Mathematical Modelling
- Data Science to Production
- Recommender Systems
- Uncertainty Quantification & Causality
- Python Data Stack
- Maintainer PyScaffold

Simon Bachstein

Data Scientist @ inovex

@simonbachstein

sbachstein

sbachstein.de

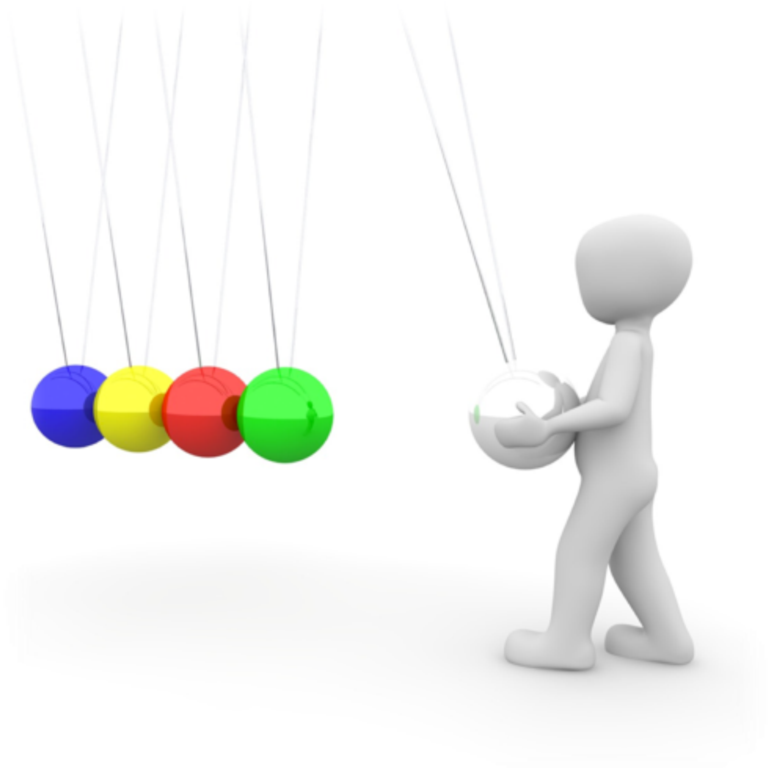
2018/07 – 2019/01 Master Thesis at inovex:

Uncertainty Quantification in Deep Learning

- Blogpost:
  http://inovex.de/blog/uncertainty-quantification-deep-learning
- Master Thesis:
  https://sbachstein.de/master_thesis.pdf

Agenda

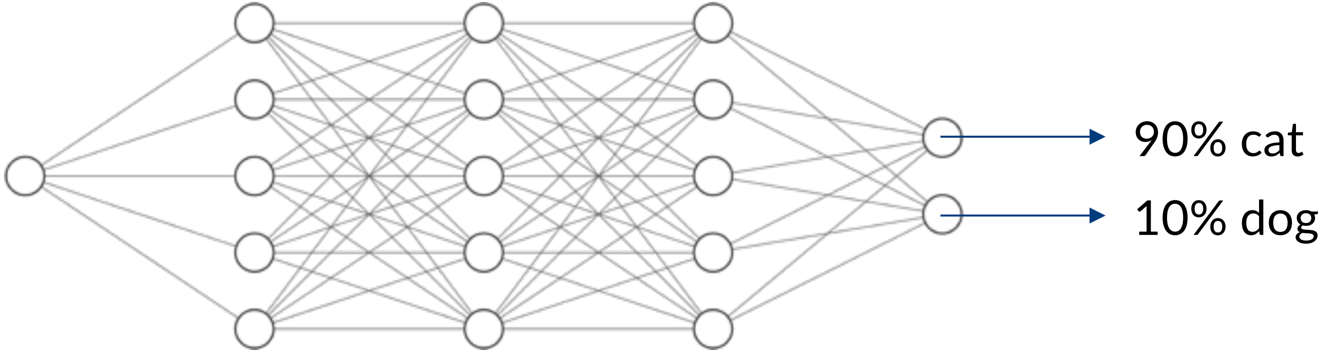
1. Motivation
2. Methods
   a. Gaussian Processes
   b. Monte-Carlo Dropout
   c. Deep Ensembles
   d. Dropout Ensembles
   e. Quantile Regression
3. Experiments
4. Conclusion & Outlook



inovex

Motivation

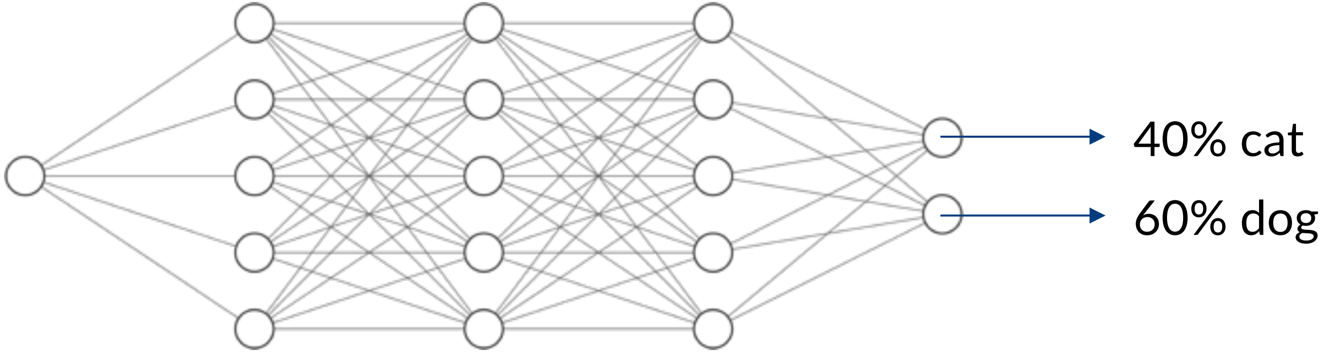
Deep Networks cannot look beyond their horizon



inovex

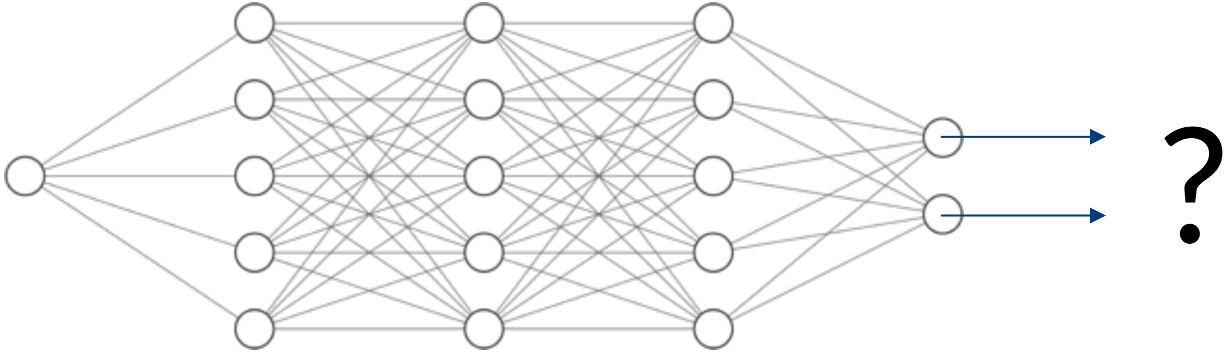
Motivation

Deep Networks cannot look beyond their horizon

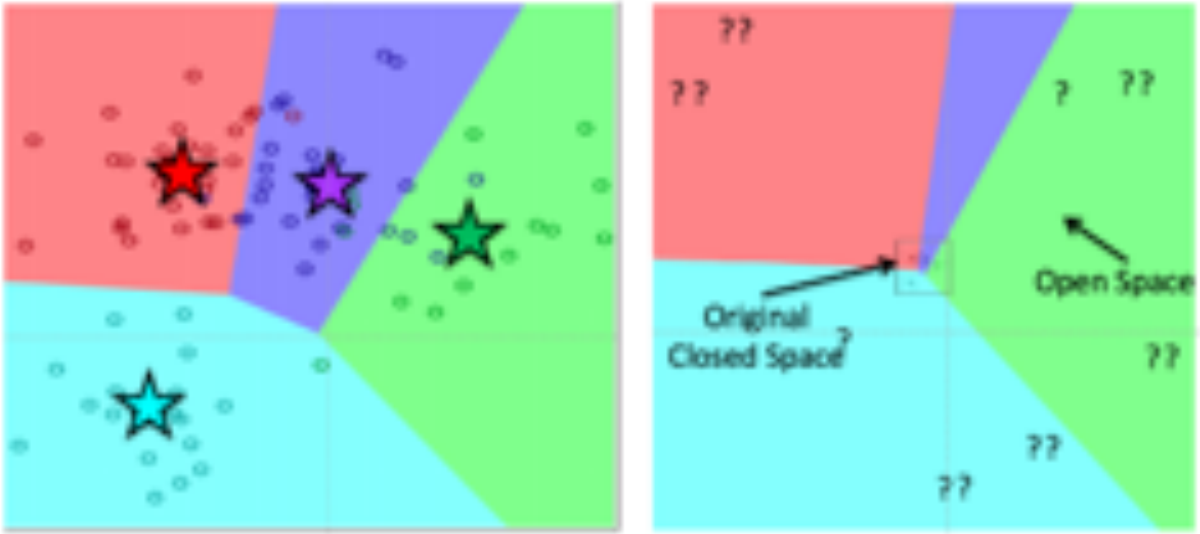


Motivation

Deep Networks cannot look beyond their horizon



Learning and the Unknown

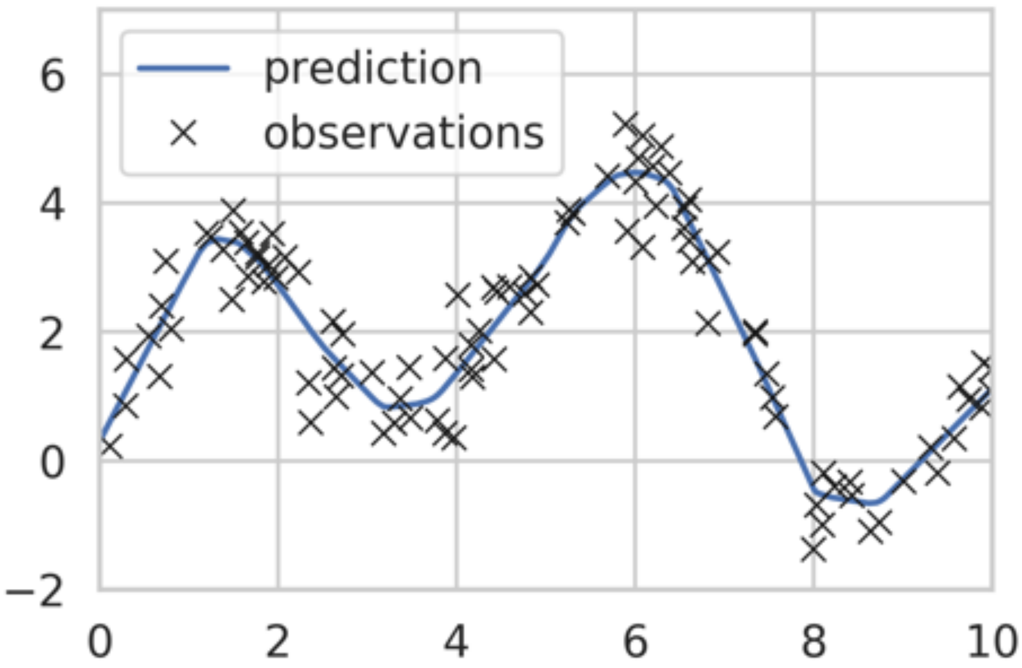


(a) Example four-class model from closed set point of view. (b) Zooming out to show some open space.

Boult, T. E., Cruz, S., Dhamija, A., Gunther, M., Henrydoss, J., & Scheirer, W. (2019). **Learning and the Unknown: Surveying Steps Toward Open World Recognition**. *Aaai*, 1–8. Retrieved from www.aaai.org

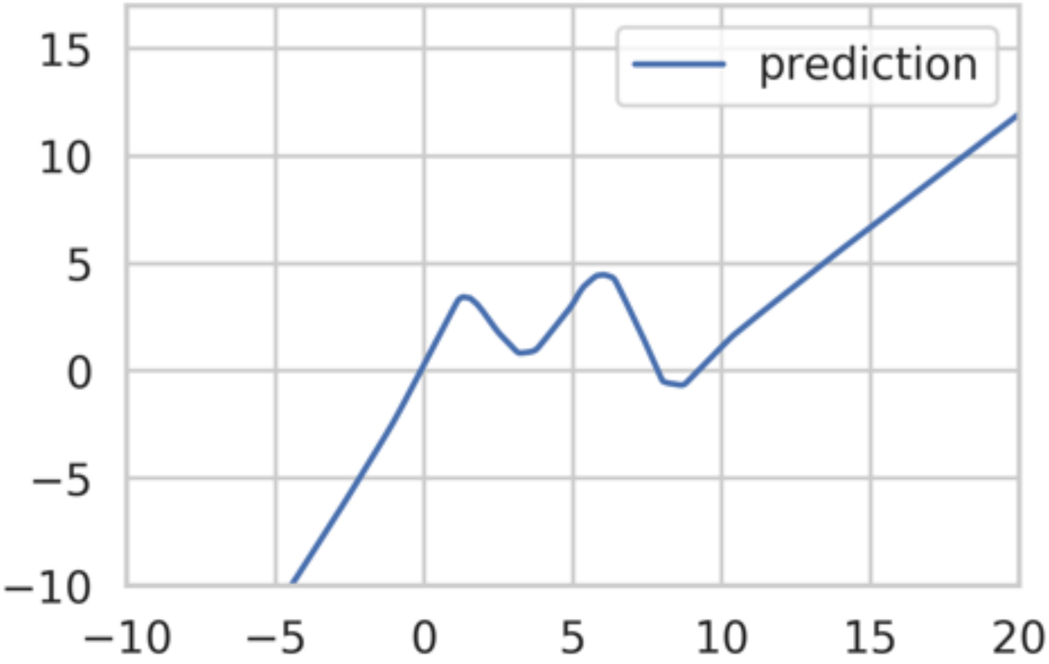
Simple Regression Problem

Interpolation



Simple Regression Problem

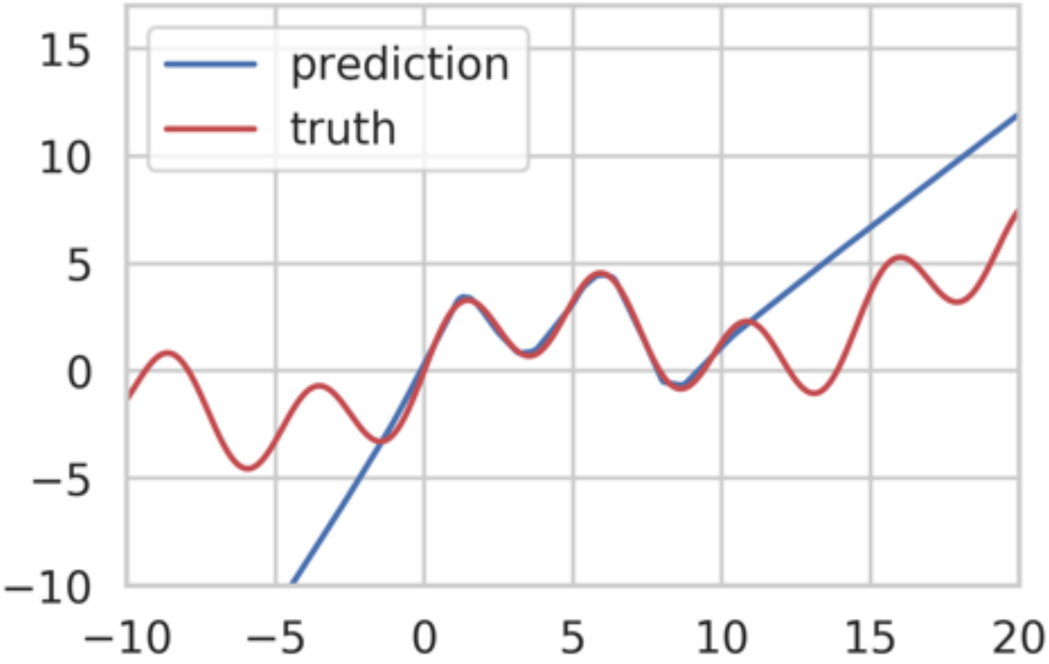
Deep Networks don’t extrapolate



Neural Arithmetic Logic Units, NIPS'18, Andrew Trask et. al.

Simple Regression Problem

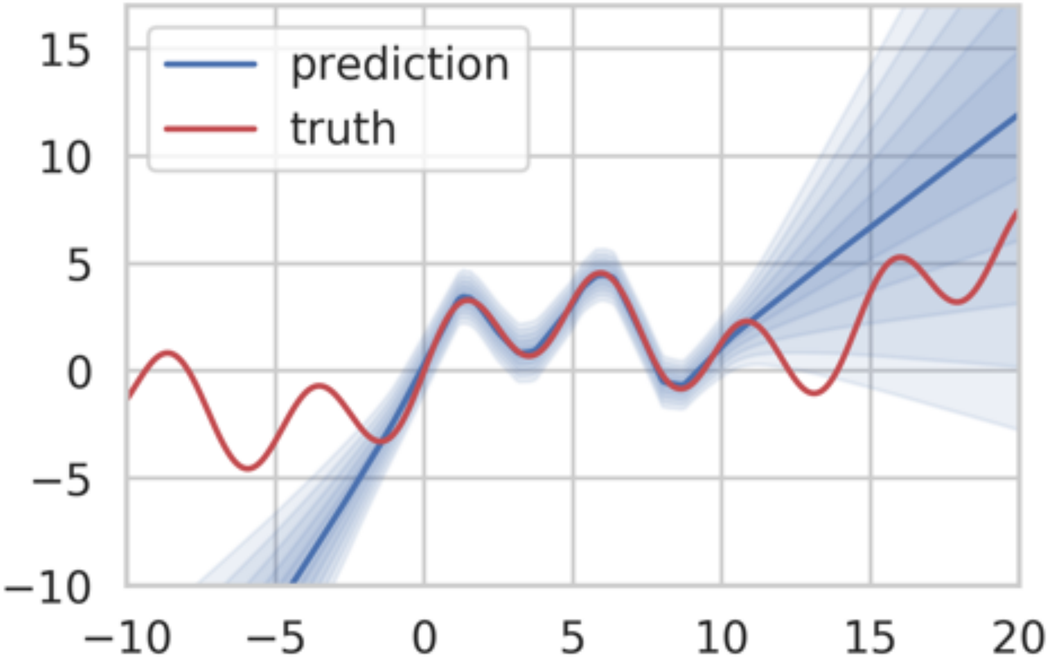
Deep Networks don’t extrapolate



inovex

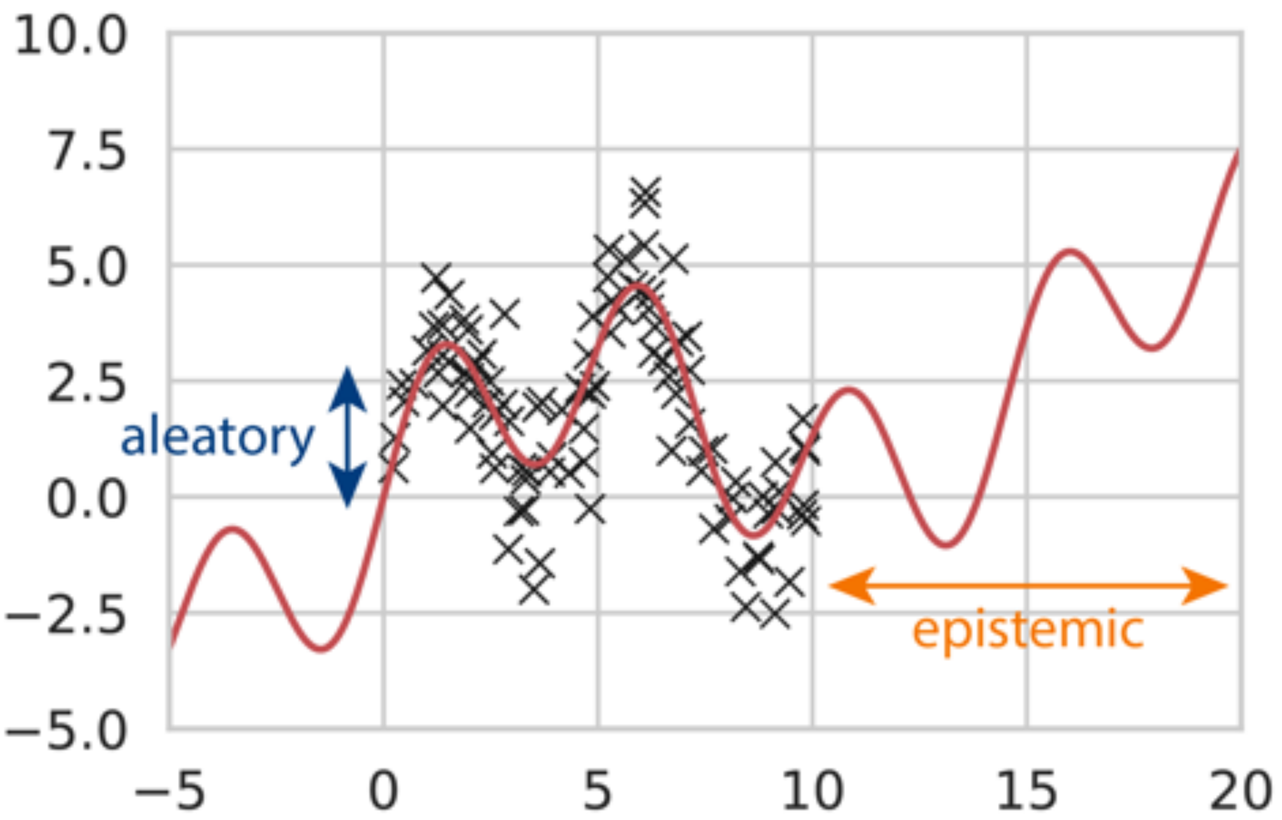
Simple Regression Problem

Uncertainty about interpolation and extrapolation



inovex

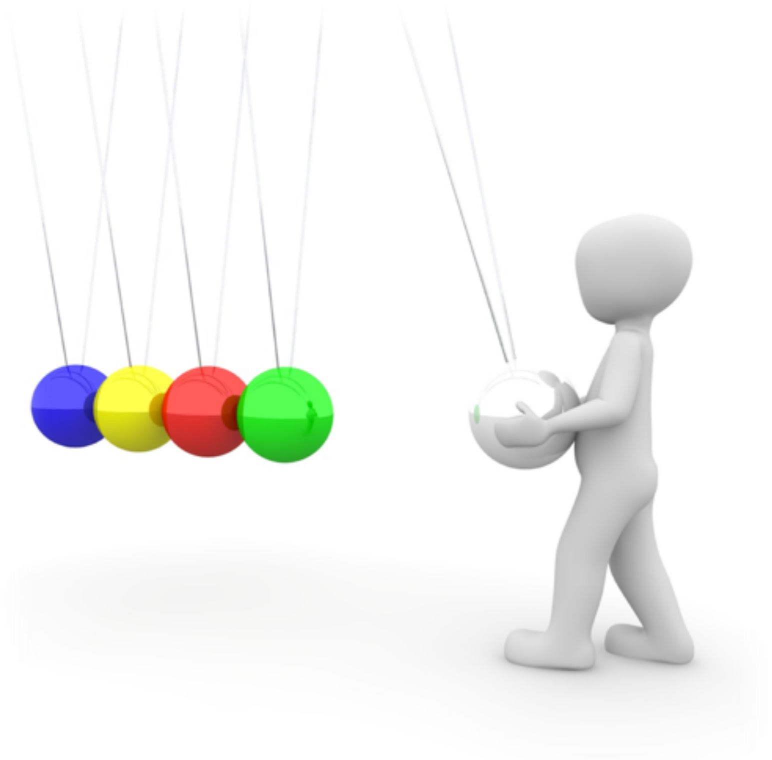
Types of Uncertainty



inovex

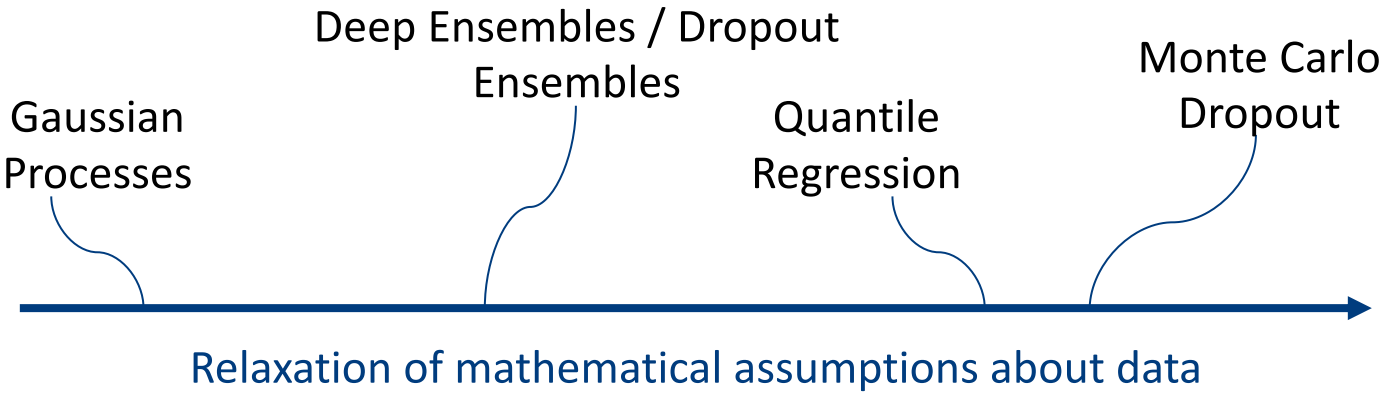
Agenda

1. Motivation
2. Methods
   a. Gaussian Processes
   b. Monte-Carlo Dropout
   c. Deep Ensembles
   d. Dropout Ensembles
   e. Quantile Regression
3. Experiments
4. Conclusion & Outlook



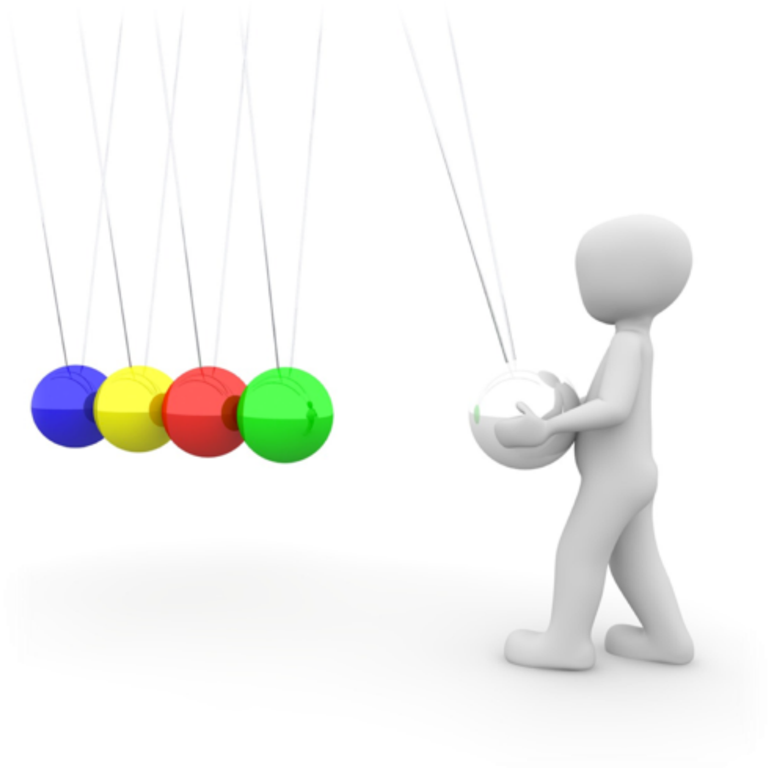
inovex

Methods for Uncertainty Quantification



Agenda

1. Motivation
2. Methods
   a. Gaussian Processes
   b. Monte-Carlo Dropout
   c. Deep Ensembles
   d. Dropout Ensembles
   e. Quantile Regression
3. Experiments
4. Conclusion & Outlook



inovex

Gaussian Processes

Definition

A Gaussian Process can be thought of as a random function

$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$$

which is defined by its mean and covariance functions

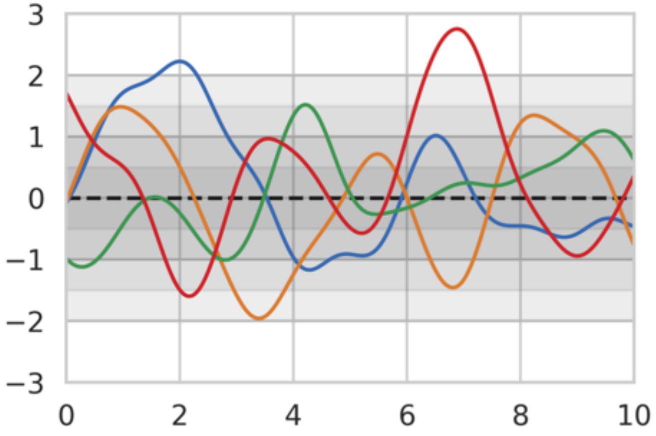
$$m(x) = \mathbb{E}[f(x)], \qquad k(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}(f(\boldsymbol{x}), f(\boldsymbol{x}'))$$
$$= \mathbb{E}[(f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x}') - m(\boldsymbol{x}'))]$$

Gaussian Processes

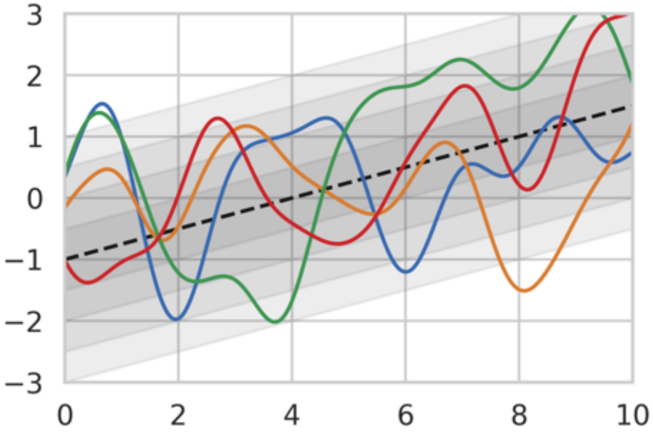
Example

$$k(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\|\boldsymbol{x} - \boldsymbol{x}'\|^2\right)$$

$$m(\boldsymbol{x}) = \boldsymbol{0}$$



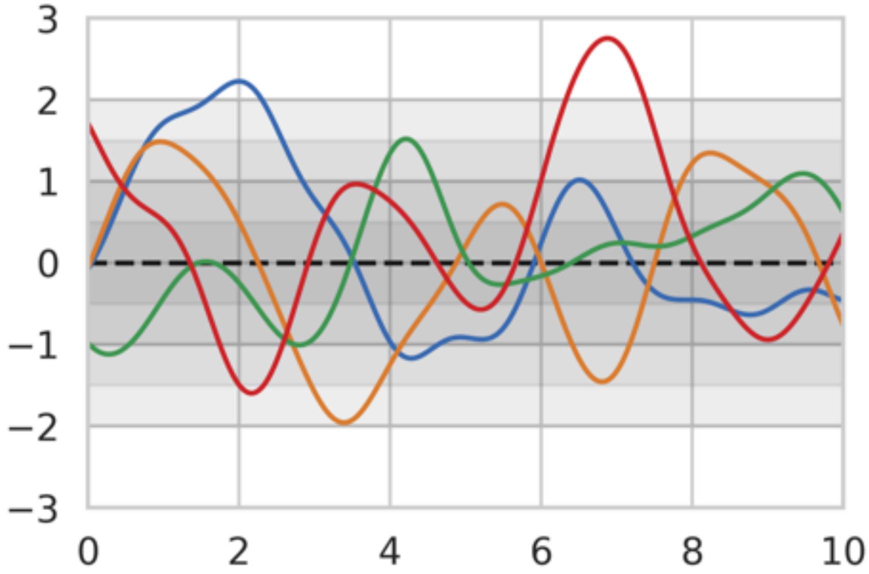
$$m(\boldsymbol{x}) = \frac{\boldsymbol{x}}{4} - 1$$



inovex

Gaussian Processes

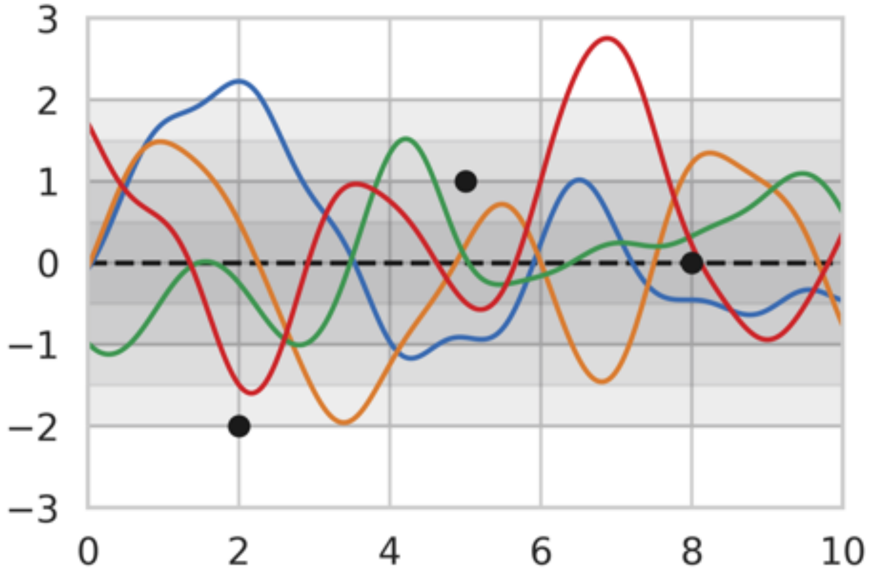
Inference



inovex

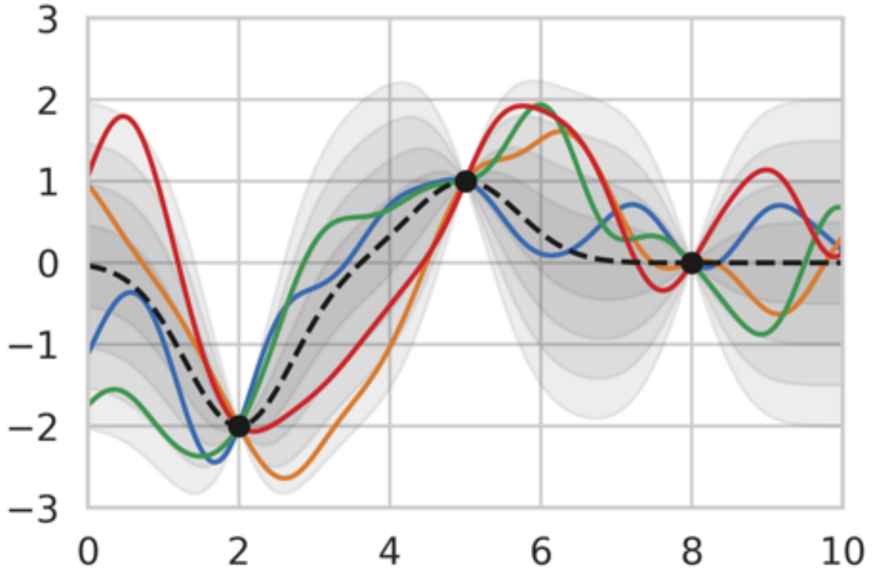
Gaussian Processes

Inference



Gaussian Processes

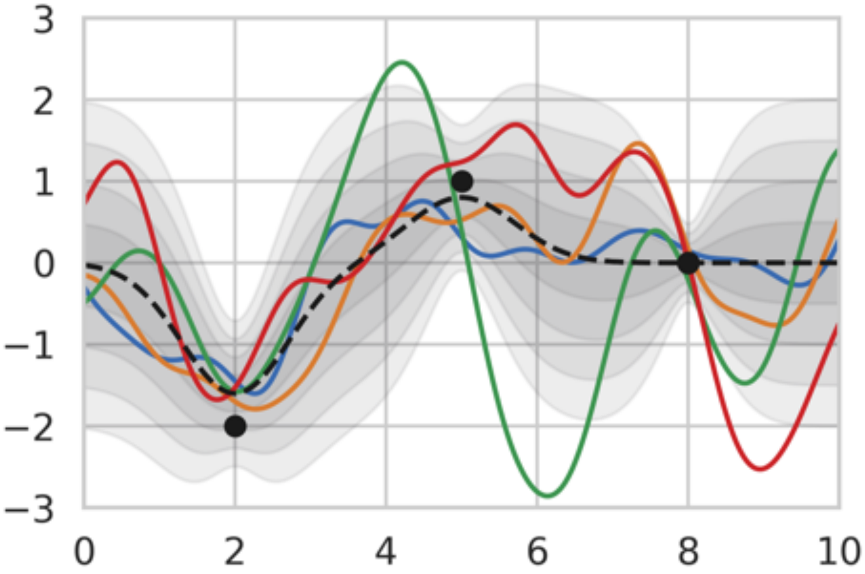
Inference with perfect interpolation



inovex

Gaussian Processes

Inference with noisy observations



inovex

Gaussian Processes

Inference

Inference using given data points can be done analytically. For example, when assuming the (prior) mean function to be zero everywhere, we get:

$$f(X^*)|X, f(X) \sim \mathcal{N}(\mu^*, \Sigma^*)$$

$$\mu^* = K(X^*, X)K(X, X)^{-1} f(X)$$

$$\Sigma^* = K(X^*, X^*) - K(X^*, X)K(X, X)^{-1} K(X, X^*)$$

computationally intense

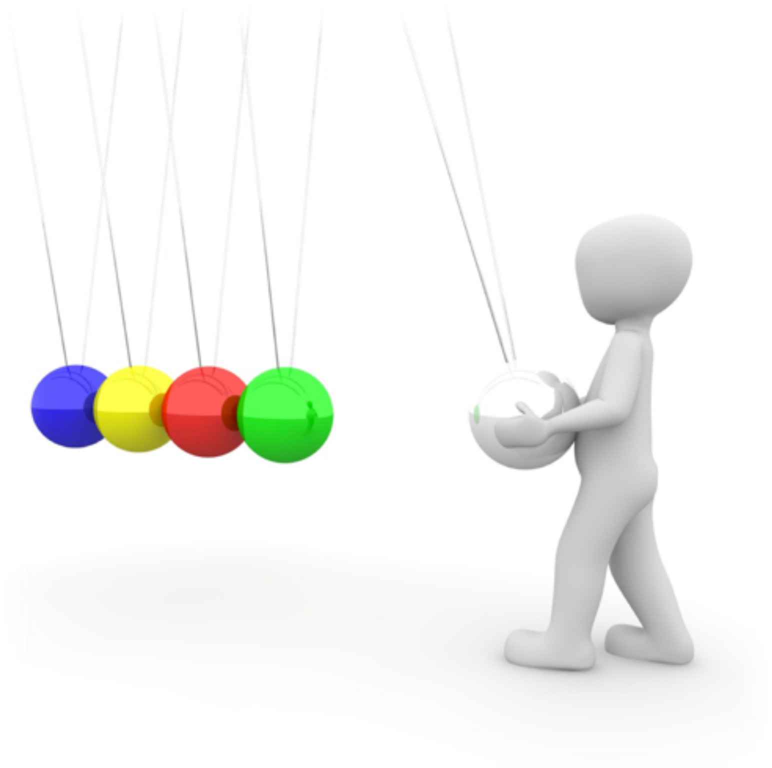
Good introduction:

Bayesian Non-parametric Models for Data Science using PyMC by Christopher Fonnesbeck

- https://www.youtube.com/watch?v=-sIOMs4MSuA
- https://de.slideshare.net/mlreview/bayesian-nonparametric-models-for-data-science-using-pymc

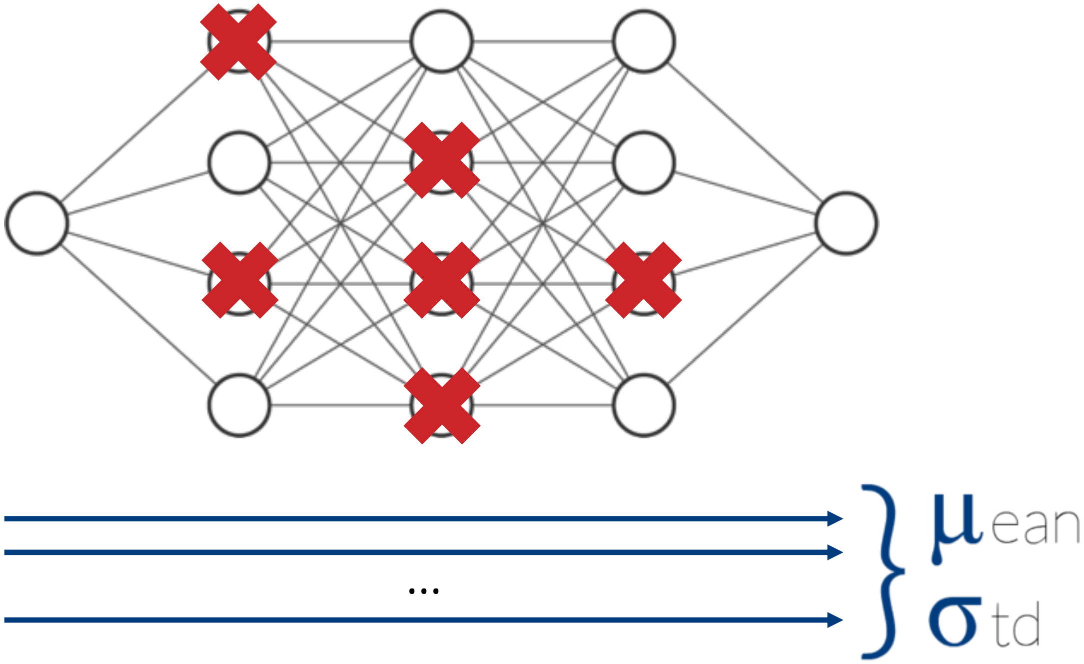
Agenda

1. Motivation
2. Methods
   a. Gaussian Processes
   b. Monte-Carlo Dropout
   c. Deep Ensembles
   d. Dropout Ensembles
   e. Quantile Regression
3. Experiments
4. Conclusion & Outlook



inovex

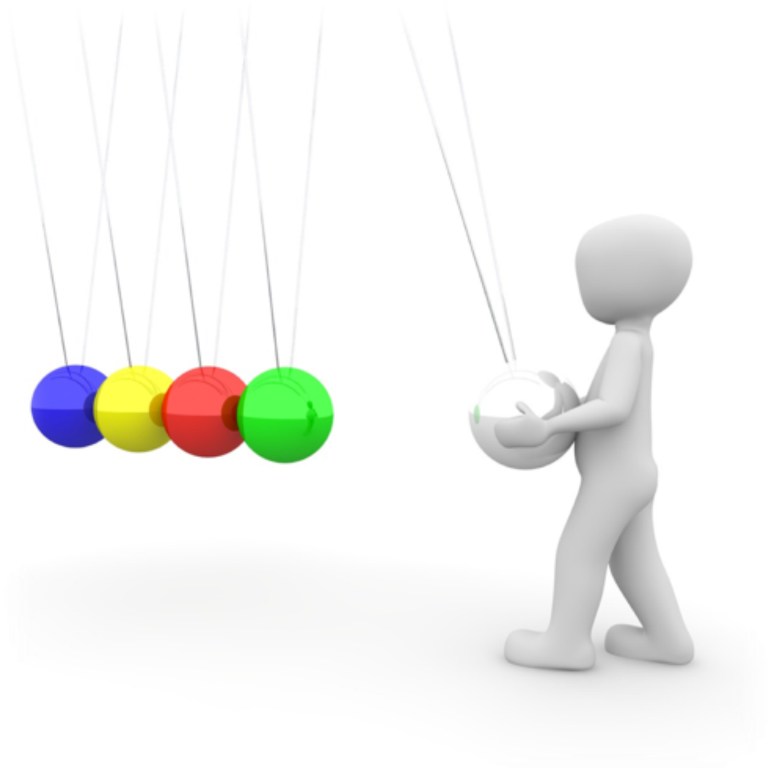
MC Dropout



Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, ICML 2016, Yarin Gal et. al.

Agenda

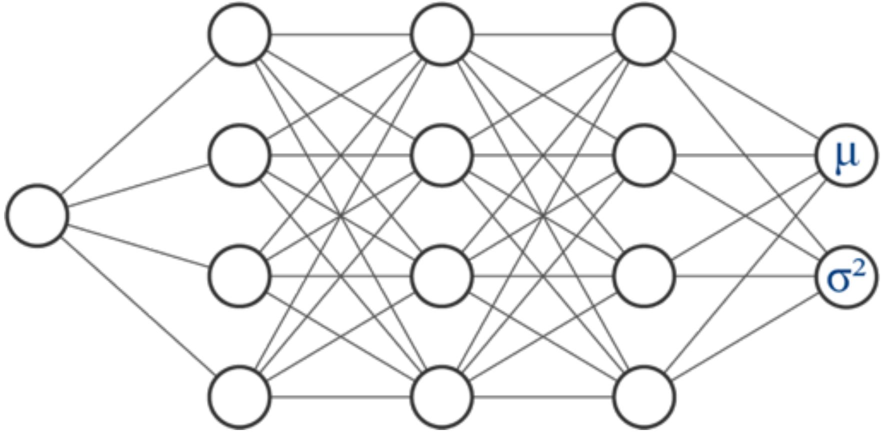
1. Motivation
2. Methods
   a. Gaussian Processes
   b. Monte-Carlo Dropout
   c. Deep Ensembles
   d. Dropout Ensembles
   e. Quantile Regression
3. Experiments
4. Conclusion & Outlook



inovex

Deep Ensembles

Capture uncertainty directly at training time

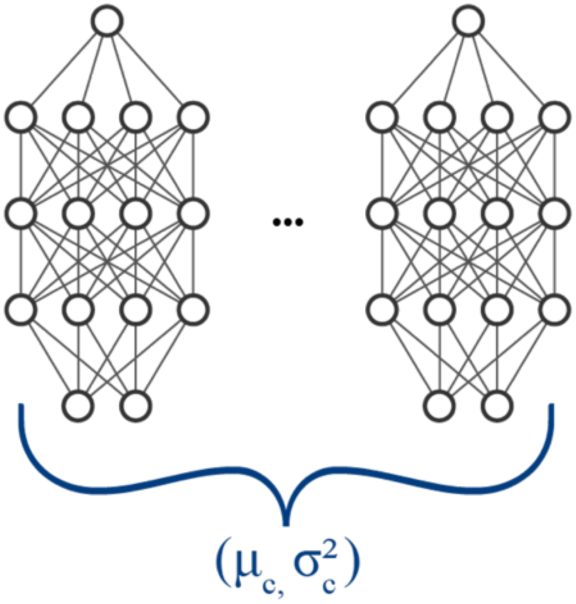


Custom loss function: $$\mathcal{L}(x, y) = \frac{\log(\sigma^2(x))}{2} + \frac{(y - \mu(x))^2}{2\sigma^2(x)}$$

 Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles, NIPS 2017, Balaji Lakshminarayanan et. al

Deep Ensembles

Combine an ensemble of networks



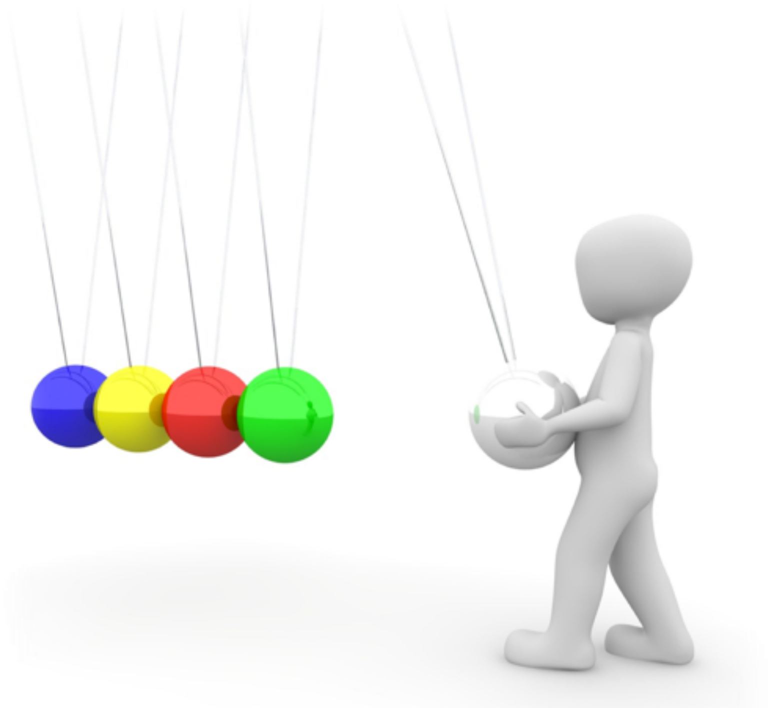
$$\mu_c = \frac{1}{M} \sum_{i=1}^{M} \mu_i$$

$$\sigma_c^2 = \frac{1}{M} \sum_{i=1}^{M} \left(\sigma_i^2 + \mu_i^2\right) - \mu_c^2$$

Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles, NIPS 2017, Balaji Lakshminarayanan et. al

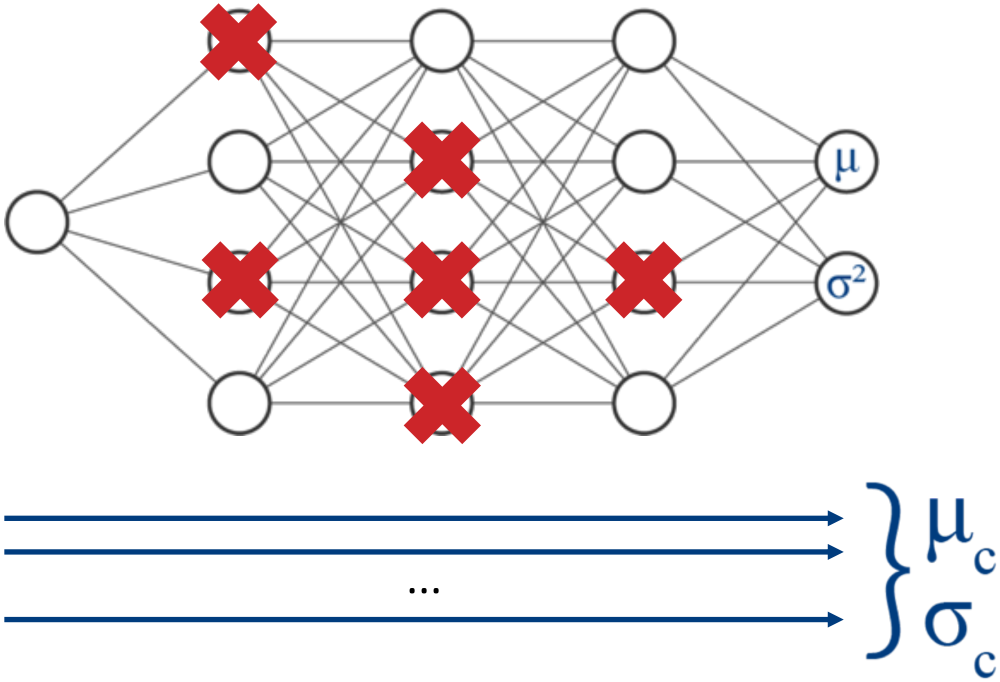
Agenda

1. Motivation
2. Methods
    a. Gaussian Processes
    b. Monte-Carlo Dropout
    c. Deep Ensembles
    d. Dropout Ensembles
    e. Quantile Regression
3. Experiments
4. Conclusion & Outlook



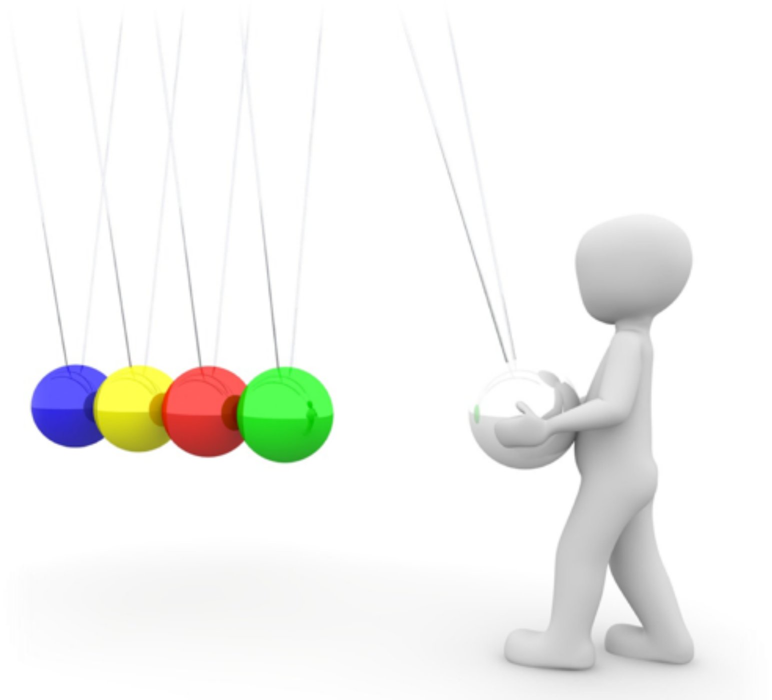
Dropout Ensembles

The best of both worlds?



Agenda

1. Motivation
2. Methods
   a. Gaussian Processes
   b. Monte-Carlo Dropout
   c. Deep Ensembles
   d. Dropout Ensembles
   e. Quantile Regression
3. Experiments
4. Conclusion & Outlook



inovex

Quantile Regression

Using the cumulative distribution function (cdf) of a random variable Y, we define the quantile:

$$F(y|X=x) = P(Y \leq y|X=x) = \tau$$

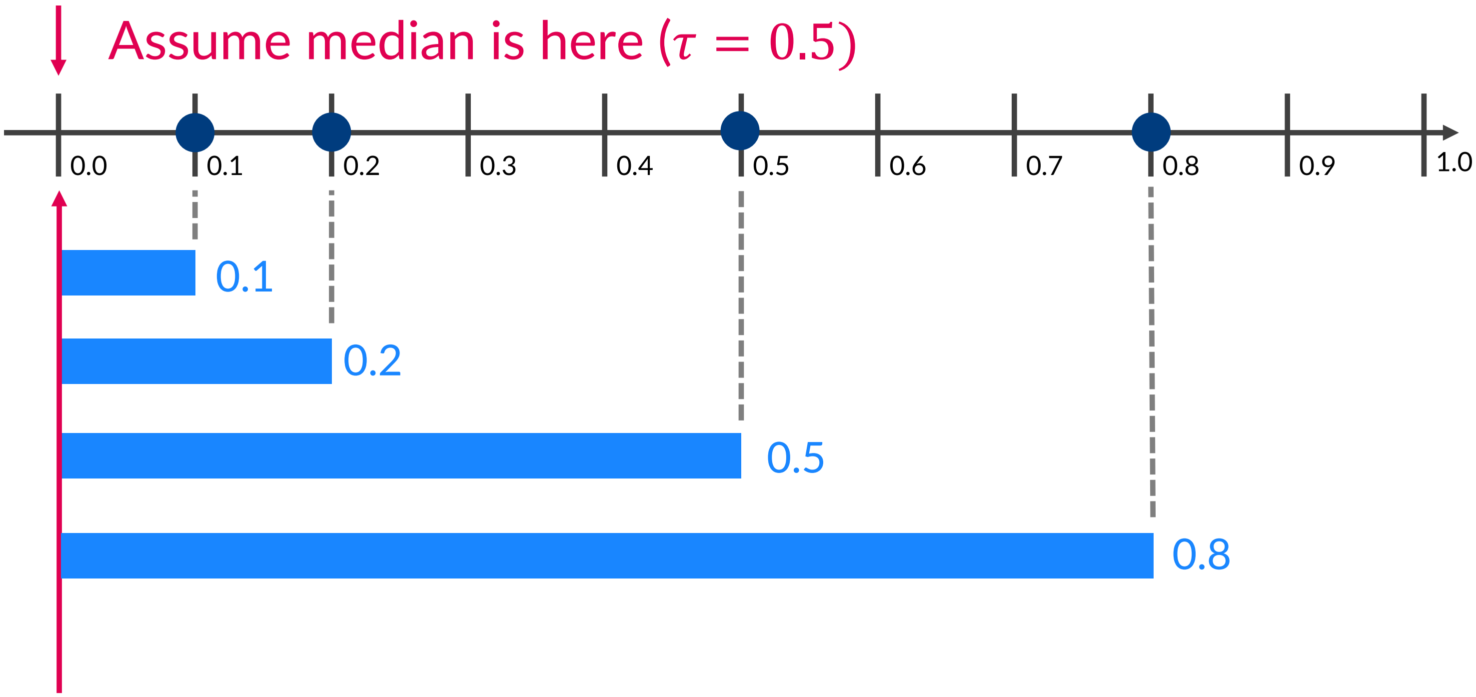
$$\overset{F^{-1}}{\Rightarrow} q_\tau(x) = \inf\{y \in \mathbb{R} | F(y|X=x) \geq \tau\}$$

Loss function to estimate quantile:

$$\mathcal{L}_\tau(x,y) = \begin{cases} \tau|y - \hat{q}_\tau(x)|, & y > \hat{q}_\tau(x) \\ (1-\tau)|y - \hat{q}_\tau(x)|, & y \leq \hat{q}_\tau(x) \end{cases}$$

inovex

Intuition behind Quantile Regression

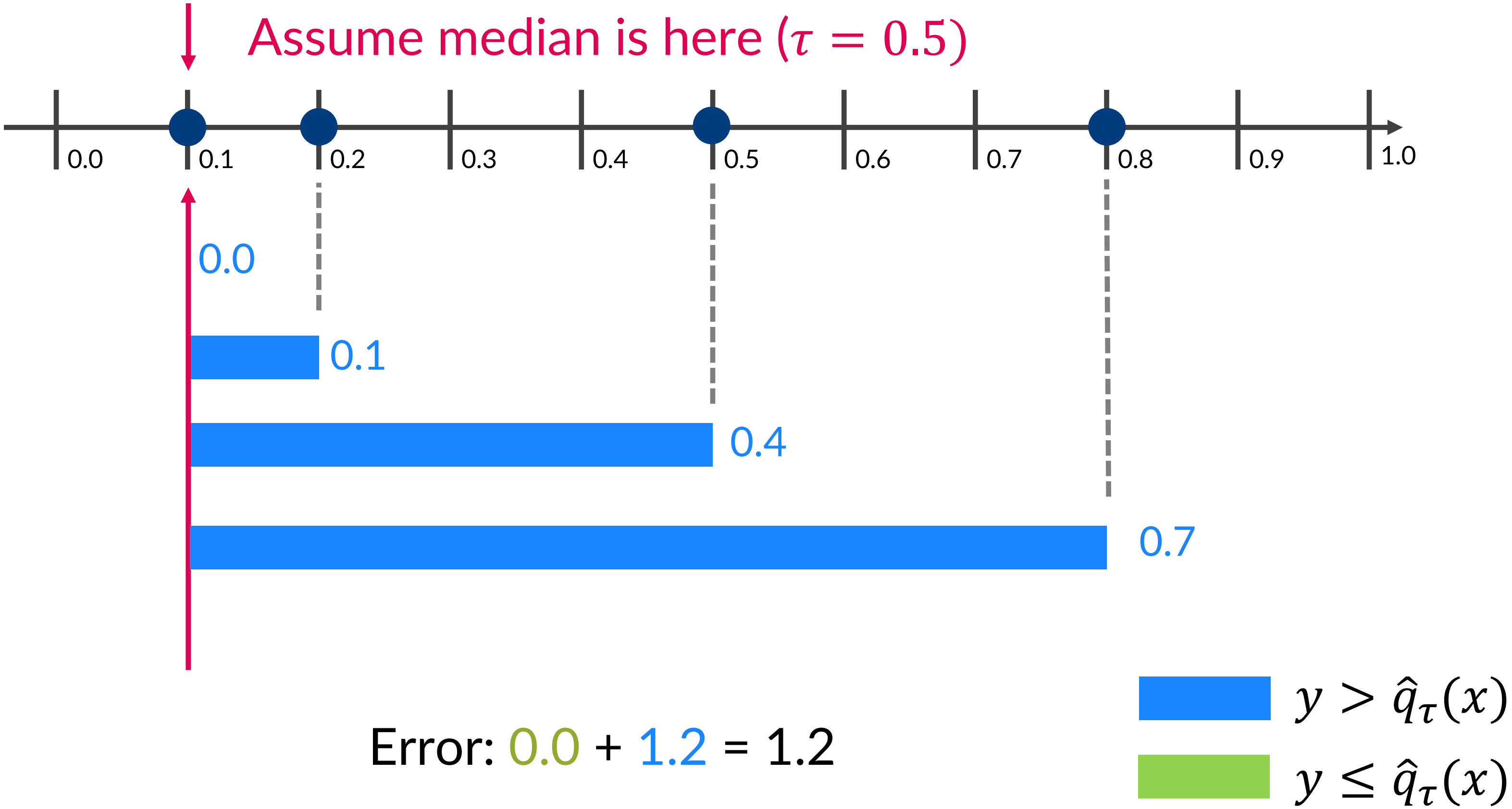


Error: 0.0 + 1.6 = 1.6

$y > \hat{q}_\tau(x)$

$y \leq \hat{q}_\tau(x)$

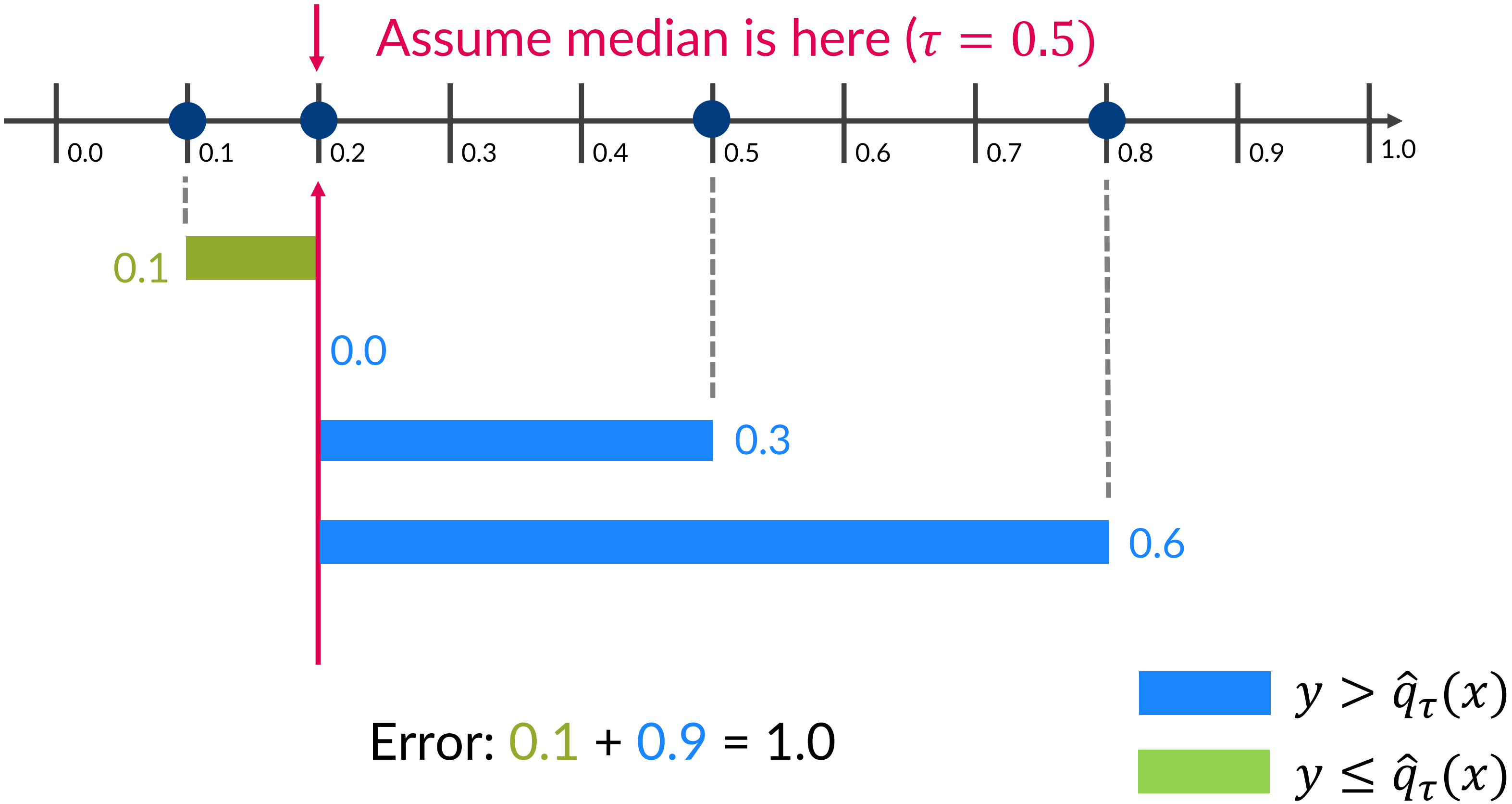
Intuition behind Quantile Regression



Error: 0.0 + 1.2 = 1.2

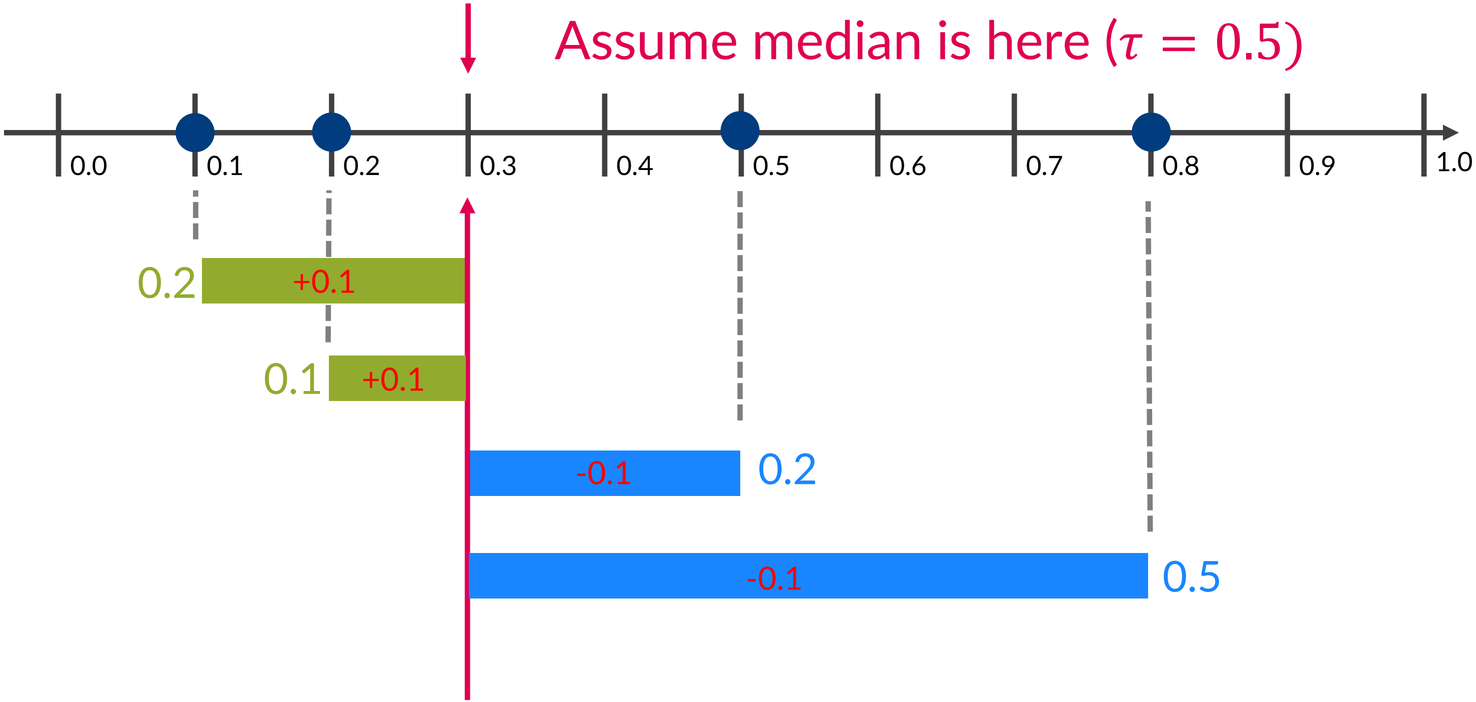
inovex

Intuition behind Quantile Regression



Error: 0.1 + 0.9 = 1.0

Intuition behind Quantile Regression



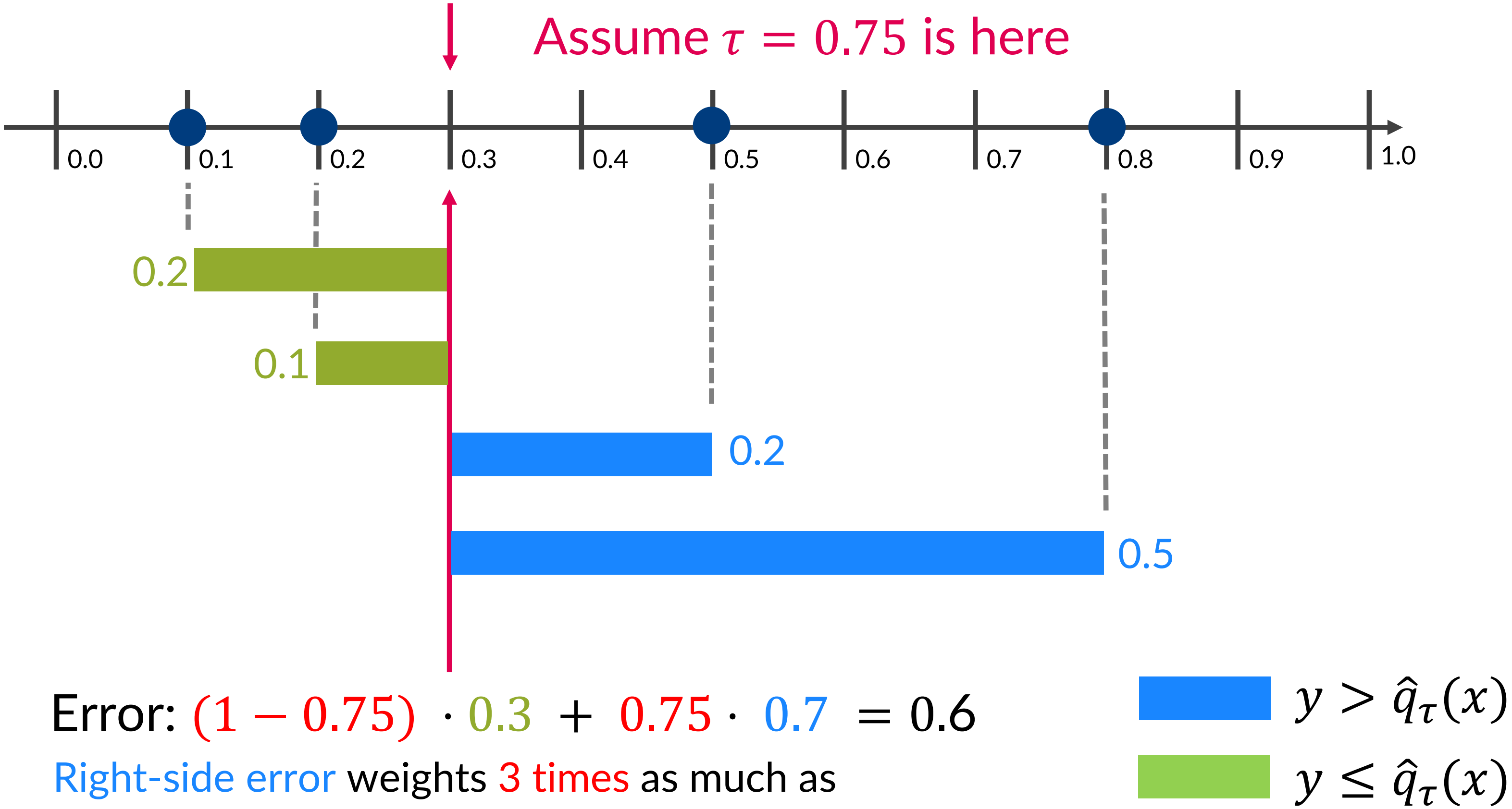
Error: 0.3 + 0.7 = 1.0

No change due to the linearity of the error!

Blue: $y > \hat{q}_\tau(x)$

Green: $y \leq \hat{q}_\tau(x)$

Now the 0.75th Quantile

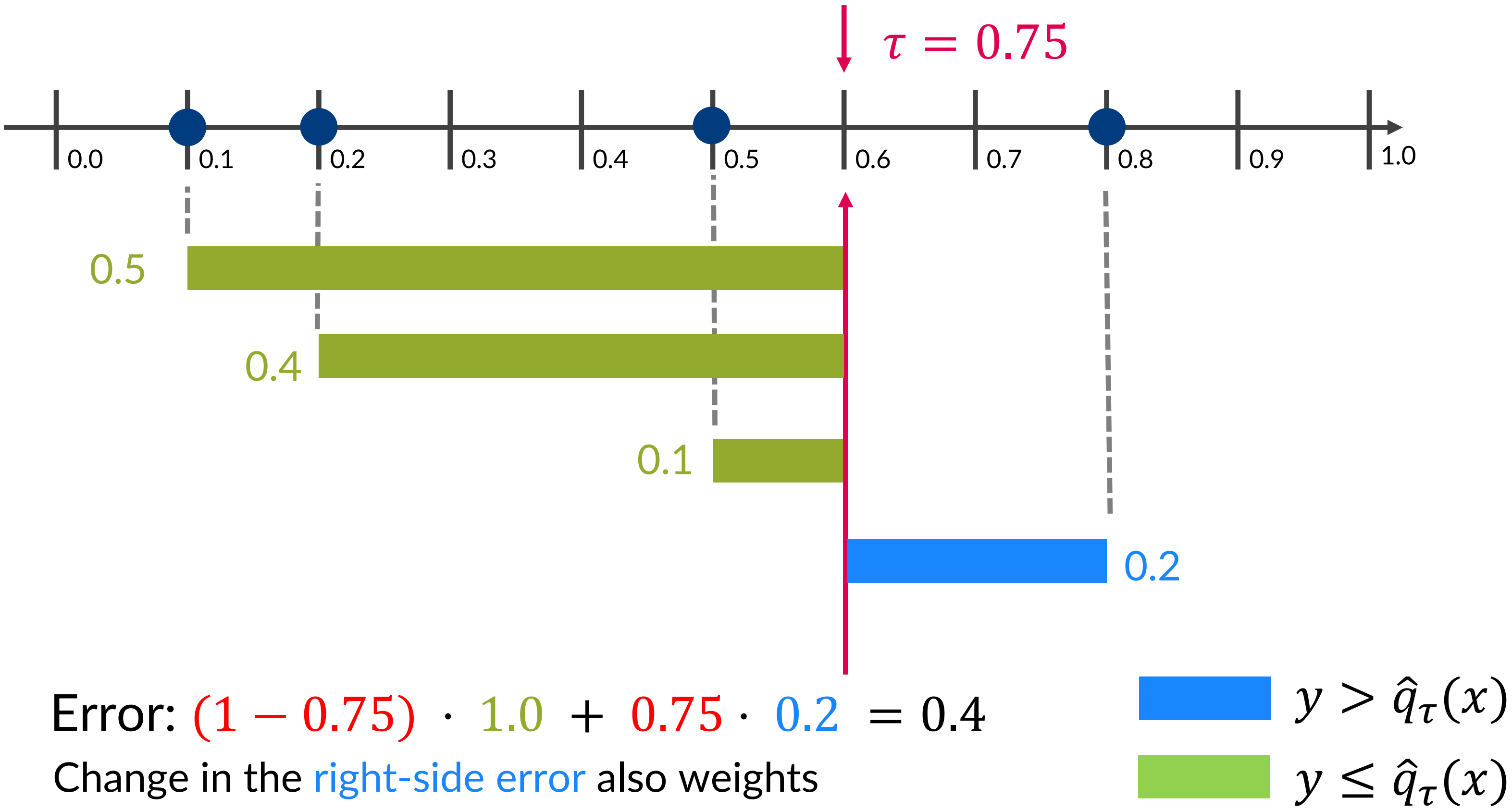


Error: $(1 - 0.75) \cdot 0.3 + 0.75 \cdot 0.7 = 0.6$

Right-side error weights 3 times as much as the left-side error

inovex

Now the 0.75th Quantile

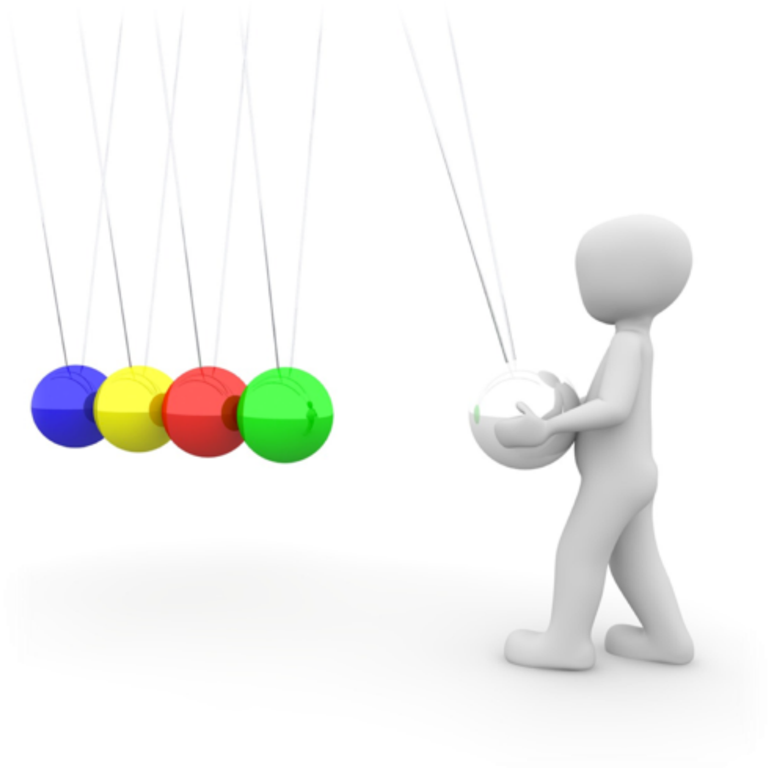


Error: $(1 - 0.75) \cdot 1.0 + 0.75 \cdot 0.2 = 0.4$

Change in the right-side error also weights
3 times as much as the left-side error

Agenda

1. Motivation
2. Methods
   a. Gaussian Processes
   b. Monte-Carlo Dropout
   c. Deep Ensembles
   d. Dropout Ensembles
   e. Quantile Regression
3. Experiments
4. Conclusion & Outlook



Experiments

Dataset

According to the function

$$f(x) = 2 \cdot \left(\frac{x}{10} + \sin\left(\frac{4x}{10}\right) + \sin\left(\frac{13x}{10}\right) \right)$$

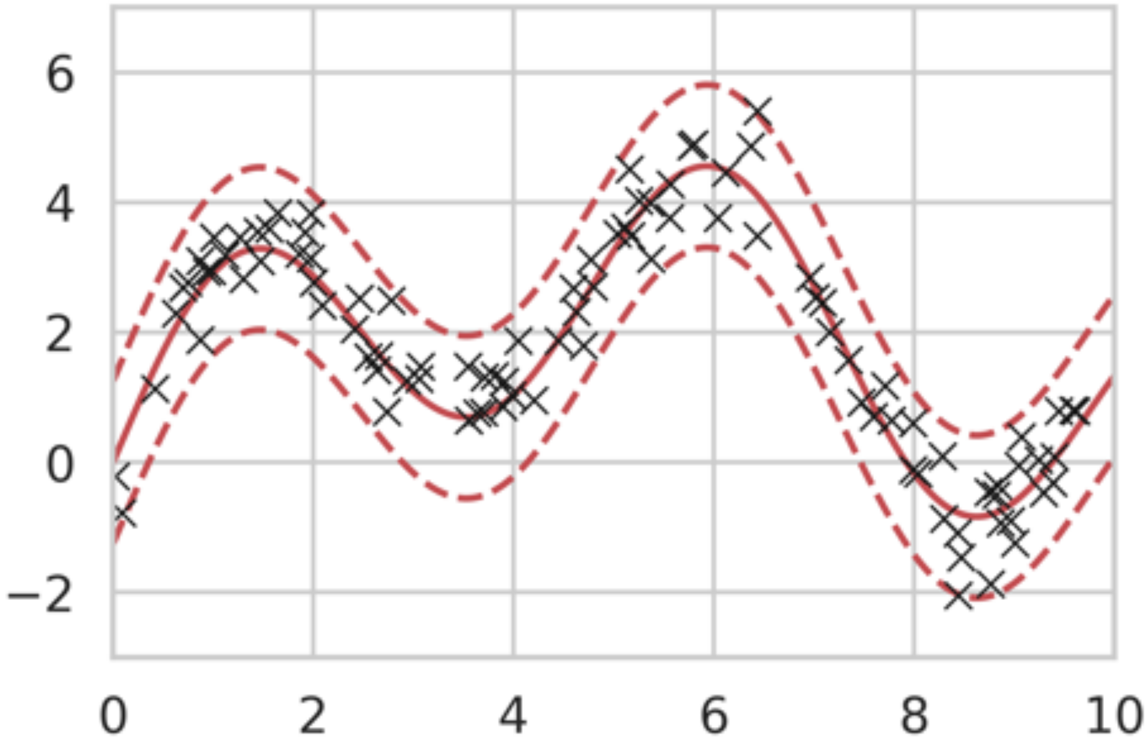
Samples are generated as follows:

$$x^{(i)} \sim \mathcal{U}([0, 10]), \qquad y^{(i)} = f(x^{(i)}) + \epsilon^{(i)}, \quad \epsilon^{(i)} \sim \mathcal{N}(0, 0.5^2)$$

Uncertainty in Deep Learning (Phd thesis), Yarin Gal, http://mlg.eng.cam.ac.uk/yarin/blog_2248.html

Experiments

Dataset



Experiments

Network setup and hyperparameters

Neural networks

- 2 hidden layers with 20 ReLU neurons each
- 5 networks for Deep Ensembles
- 100 iterations for Dropout predictions
- Adam optimizer with batch size of 128
- LR, weight decay, dropout probability are optimized

Gaussian Processes

- squared exponential covariance and zero mean function prior
- covariance function parameters and aleatory noise are optimized

Experiments

Measures for generalization quality

Mean squared error (MSE)

$$\mathrm{MSE}_\mu = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\mu}(x^{(i)}) - f(x^{(i)})\right)^2, \quad \mathrm{MSE}_\sigma = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\sigma}(x^{(i)}) - 0.5\right)^2$$

Mean negative log likelihood (MNLL)

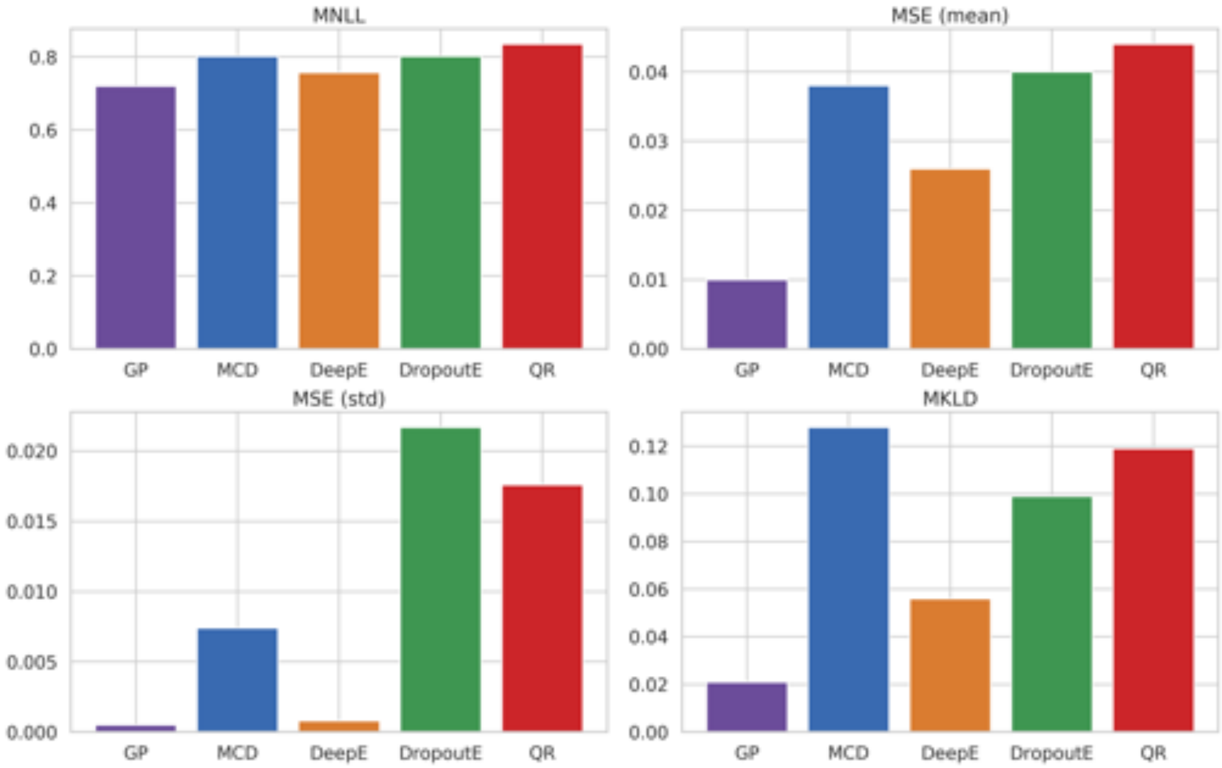
$$\mathrm{MNLL} = -\frac{1}{N}\sum_{i=1}^{N}\phi(y^{(i)}; \hat{\mu}(x^{(i)}), \hat{\sigma}^2(x^{(i)}))$$

Mean Kullback-Leibler (KL) divergence

$$MD_{KL} = \frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{\hat{\sigma}(x^{(i)})}{0.5}\right) + \frac{0.5^2 + (f(x^{(i)}) - \hat{\mu}(x^{(i)}))^2}{2\hat{\sigma}^2(x^{(i)})} - \frac{1}{2}$$

Experiments

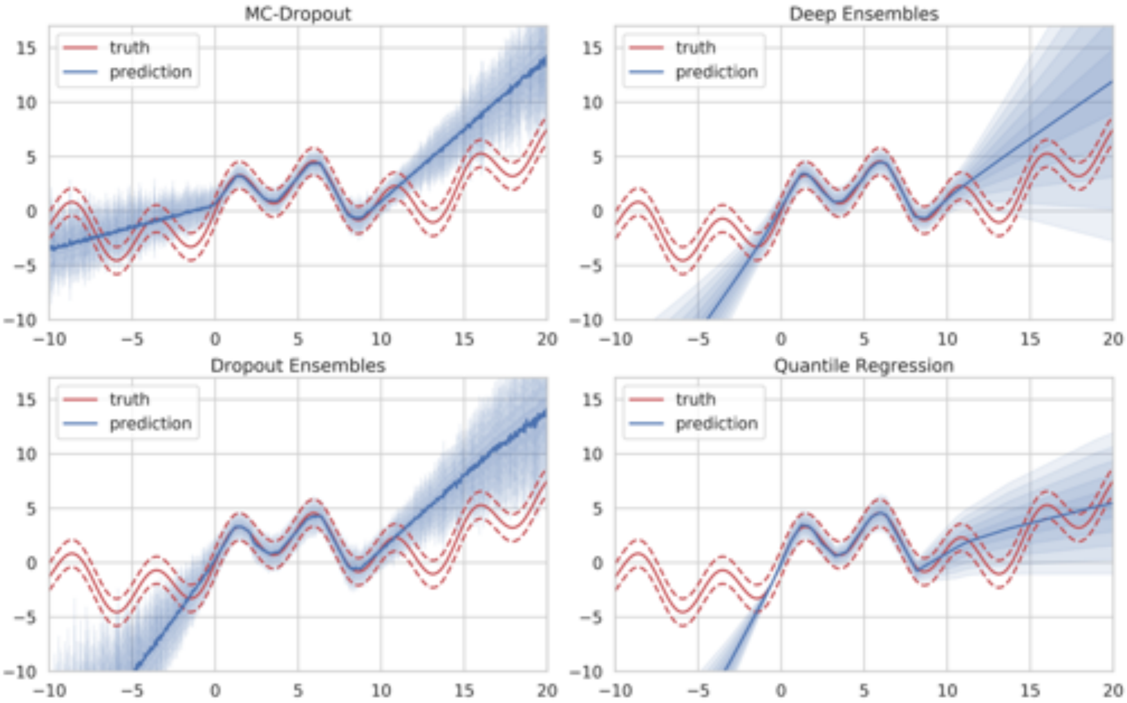
Interpolation



inovex

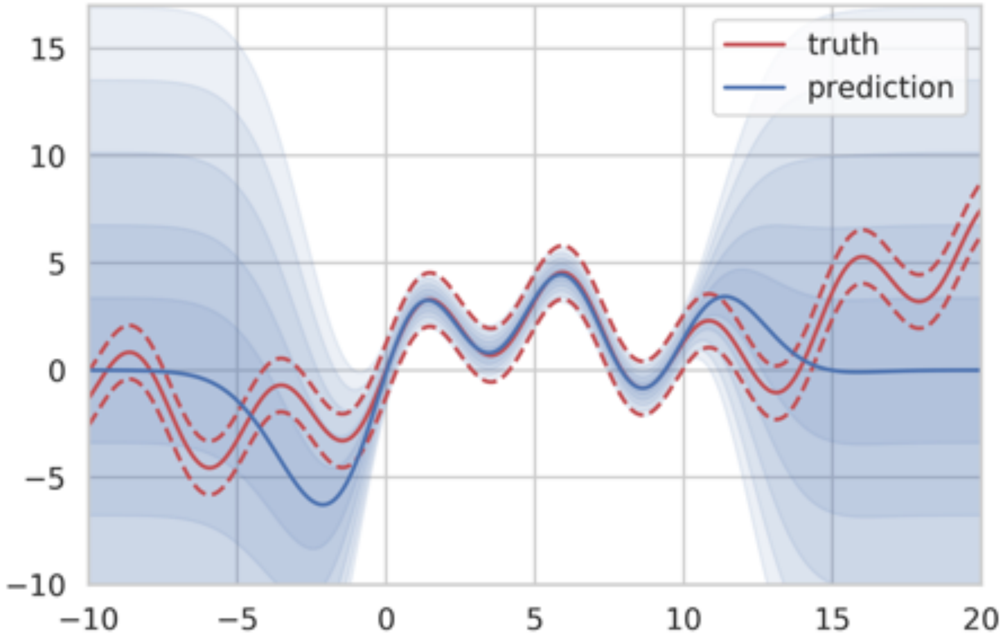
Experiments

They still don’t extrapolate and they don’t quite realize



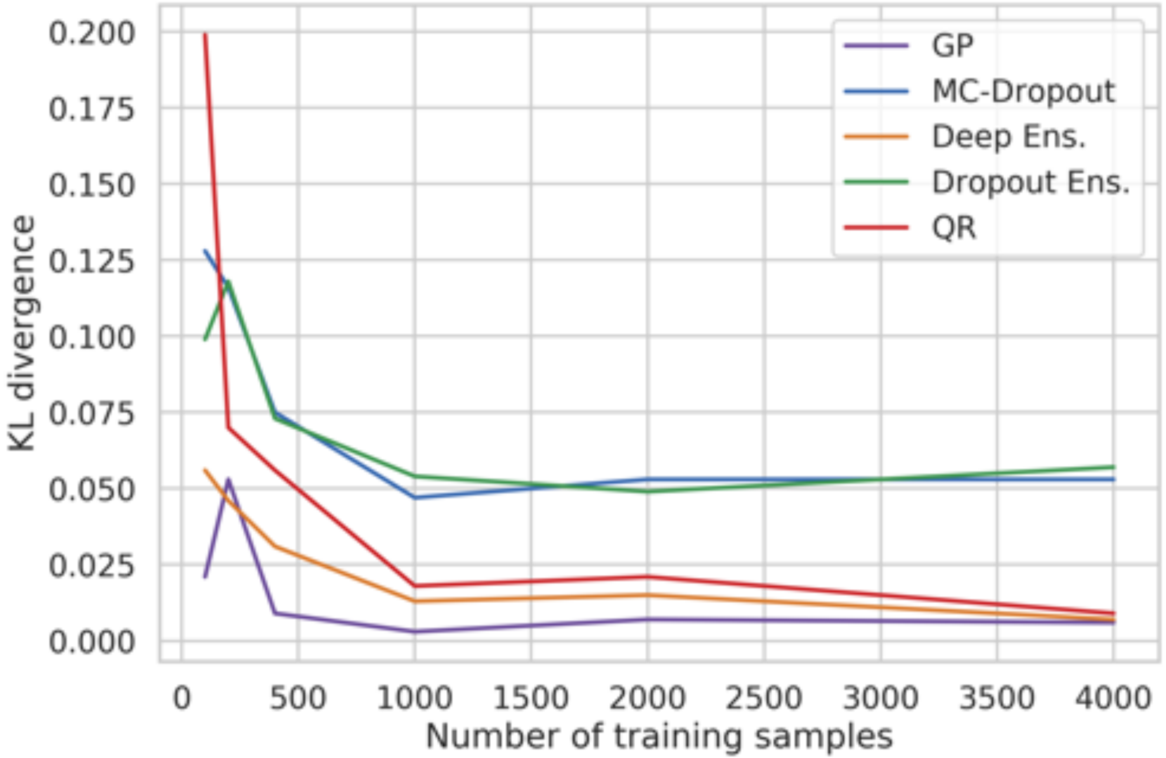
Experiments

Gaussian Process



Experiments

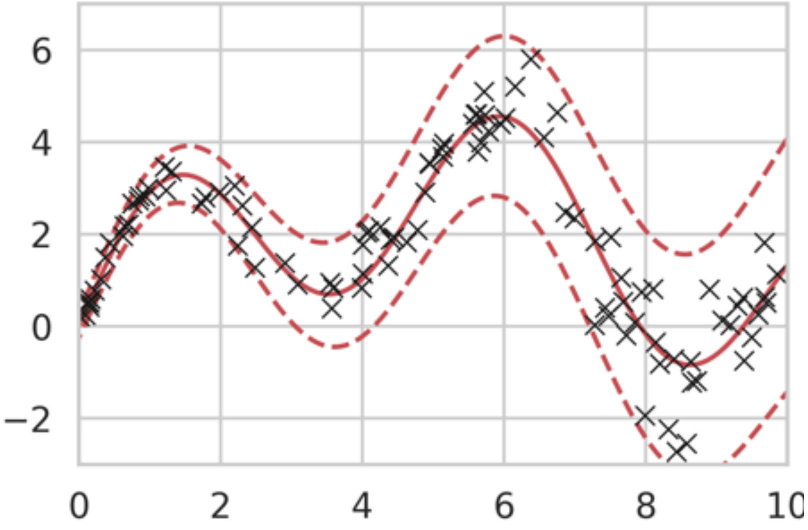
Convergence

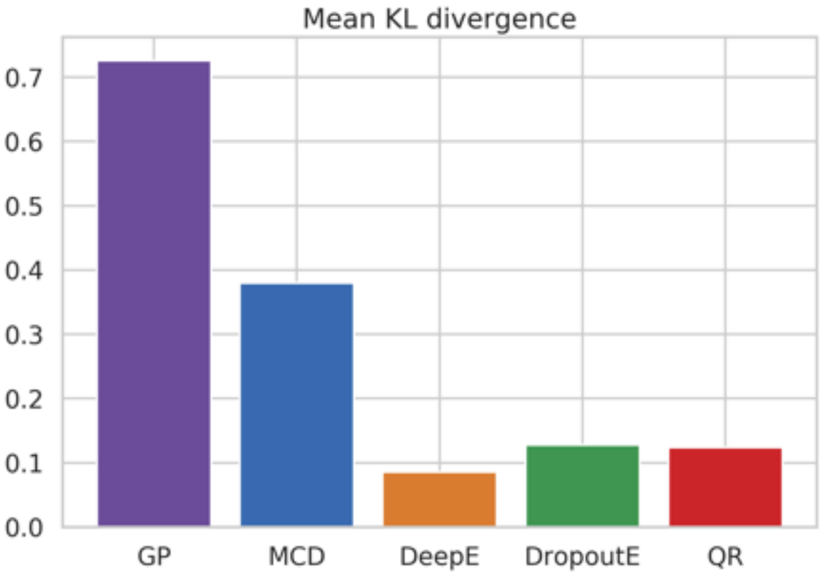


inovex

Experiments

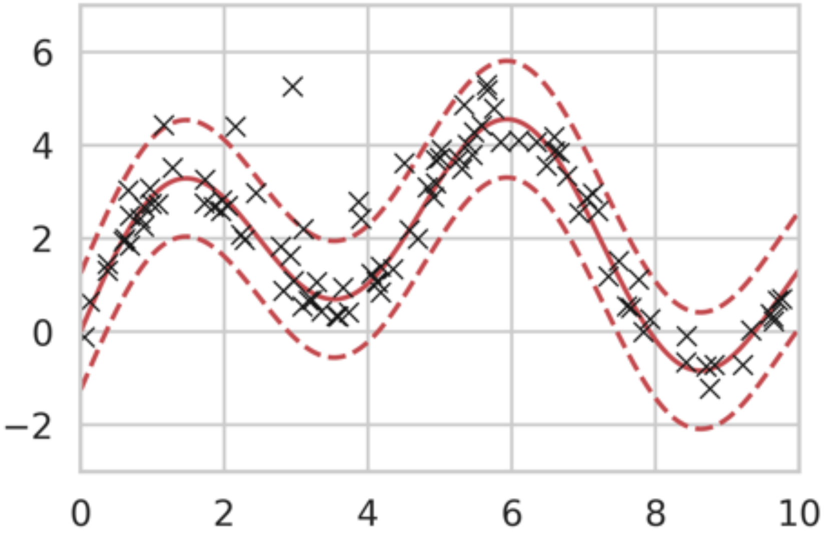
Heteroscedastic noise

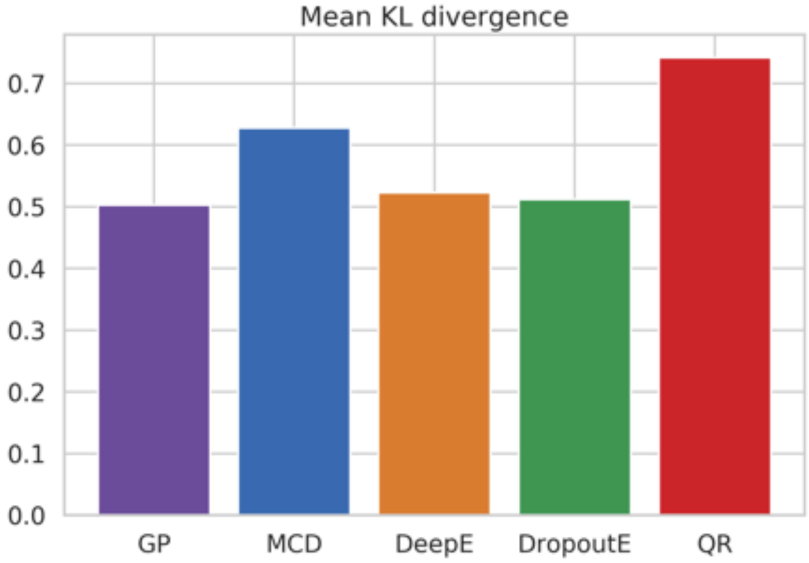




Experiments

Non-Gaussian noise





inovex

Experiments

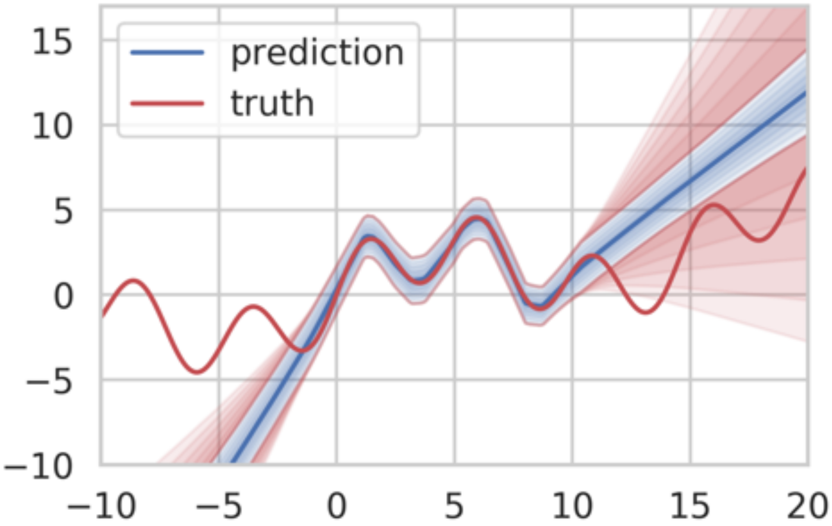
Uncertainty split

$$\sigma_c^2 = \frac{1}{M}\sum_{i=1}^{M}(\sigma_i^2 + \mu_i^2) - \mu_c^2$$

$$= \frac{1}{M}\sum_{i=1}^{M}\sigma_i^2 + \left[\frac{1}{M}\sum_{i=1}^{M}\mu_i^2 - \mu_c^2\right]$$

aleatoric

epistemic



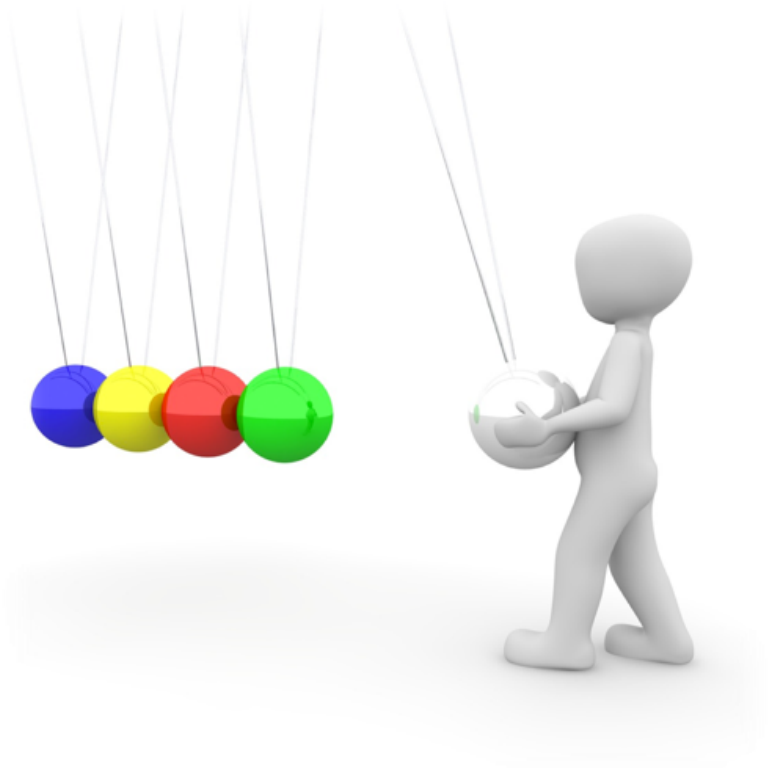
inovex

Summary

	GP	MCD	DeepE	DropoutE	QR
Homoscedastic noise	++	o	+	o	o
Heteroscedastic noise	--	-	++	+	+
Non-Gaussian noise	+	o	+	+	-
Convergence	++	-	+	-	+
Speed	(--)	+	- / (+)	+	++
Uncertainty split	yes	no	yes	yes	no

Agenda

1. Motivation
2. Methods
   a. Gaussian Processes
   b. Monte-Carlo Dropout
   c. Deep Ensembles
   d. Dropout Ensembles
   e. Quantile Regression
3. Experiments
4. Conclusion & Outlook



inovex

Conclusion

There is work to be done

- Neural network approaches discussed here are very aware of aleatory uncertainty, however, not capable of correctly estimating epistemic uncertainty
- Gaussian Processes give clear signals about ignorance but do not scale

A combined solution needs to be developed because uncertainty estimation is needed in critical applications

Outlook

Other approaches

- Bayesian Neural Networks (e.g. with PyMC)
- Sparse Gaussian Process approximations
- Gaussian Processes on top of neural networks

Thank You!

Florian Wilhelm
Principal Data Scientist

inovex GmbH
Schanzenstraße 6-20
Kupferhütte 1.13
51063 Köln

florian.wilhelm@inovex.de

inovex