MedSpace: Medical Image Analysis with Bayesian Deep Learning

Felix Laumann, *felix@neuralspace.ai*

NeuralSpace

Agenda

1. Why we all are *probably uncertain*
2. Deep Learning & Bayesian Deep Learning
3. Variational inference
    a. Bayes by Backprop
4. Epistemic & Aleatoric uncertainty
5. Medical Image Analysis

PROBABILITIES

Probability

A probability p of an event x is a measure how (un)certain x is.

$$\underbrace{0}_{\text{impossible}} \leq p(\text{event } x) \leq \underbrace{1}_{\text{certain}}$$

Random variable

A named quantity X whose value is uncertain.

⇨ **probability distribution** $p(X)$

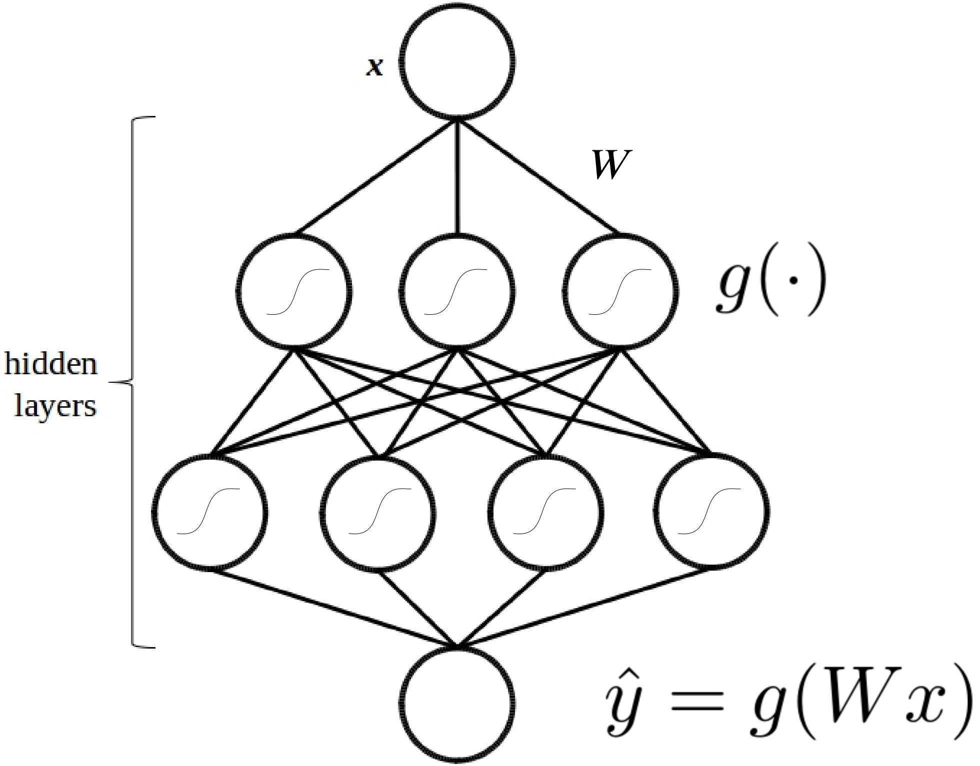
where $X = \underbrace{x_1, x_2, \ldots, x_N}_{\text{events}}$

Predictions, forecasts and decision-making

→ would it have been better, if we knew **uncertainties**?

DEEP LEARNING

Deep neural networks



Backpropagation

1. Measure error

$$C(W) = \frac{1}{2}\sum_{i=1}^{N} \|y_i - \hat{y}_i\|^2$$

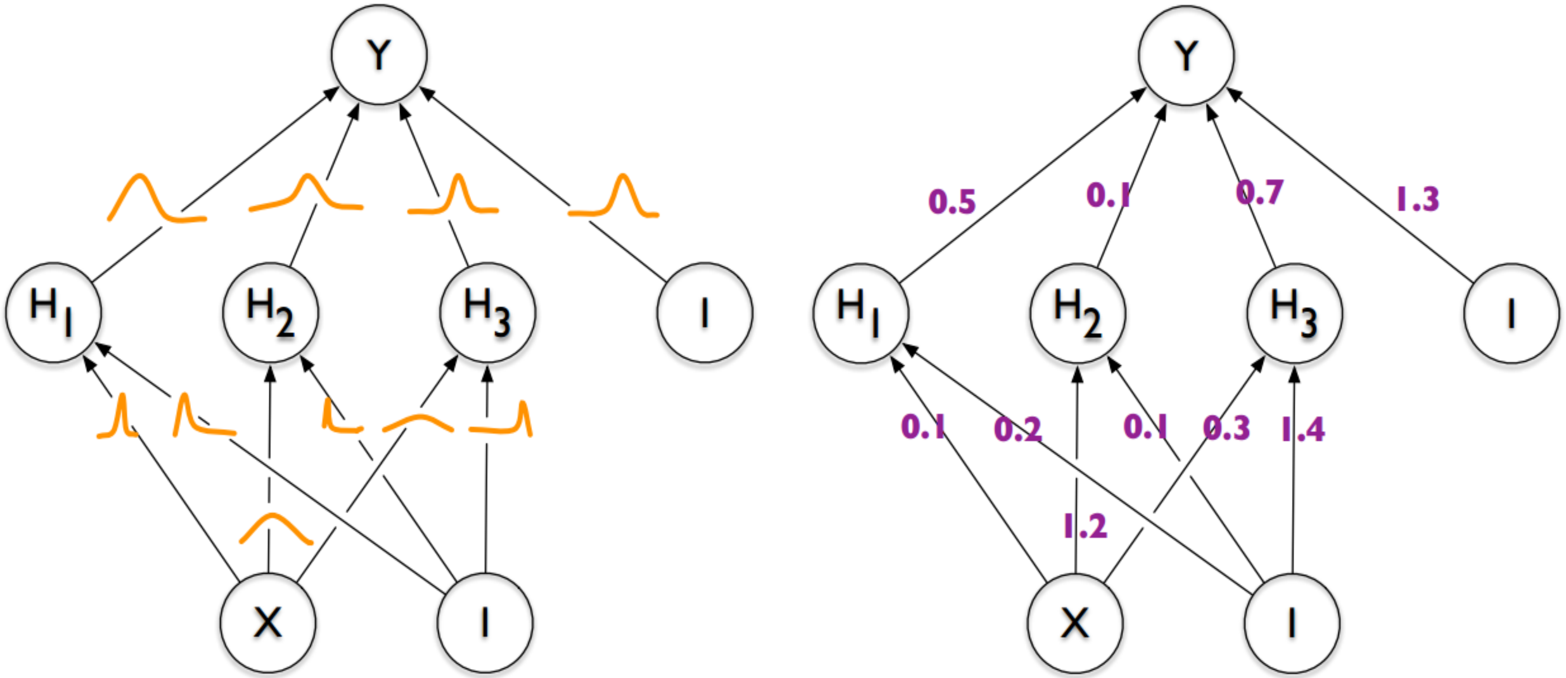
2. Measure error derived by weights and update them

$$w_{11} := w_{11} - \alpha \frac{\partial C}{\partial w_{11}}$$

→ **minimize error**

BAYESIAN DEEP LEARNING

Distributions over weights



(Blundell et al., 2015)

How to learn distributions

Bayes' rule:

$$\underbrace{p(w|\mathcal{D})}_{\text{posterior}} = \frac{\overbrace{p(\mathcal{D}|w)}^{\text{likelihood}}\ \overbrace{p(w)}^{\text{prior}}}{\underbrace{p(\mathcal{D})}_{\text{data}}}$$

INTRACTABLE

INTRACTABLE

What does intractable mean?

Our training set...

... is a part of ***p(D)***

→ will never know the entire population ***p(D)***

Bayes' rule

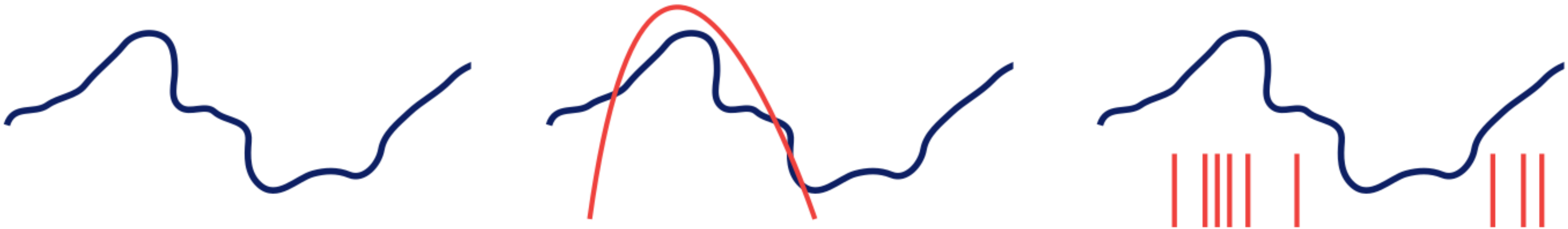
$$\boxed{\underbrace{p(w|\mathcal{D})}_{\text{posterior}}} = \frac{\overbrace{p(\mathcal{D}|w)}^{\text{likelihood}}\ \overbrace{p(w)}^{\text{prior}}}{\underbrace{p(\mathcal{D})}_{\text{data}}}$$

INTRACTABLE

INTRACTABLE

So what to do?

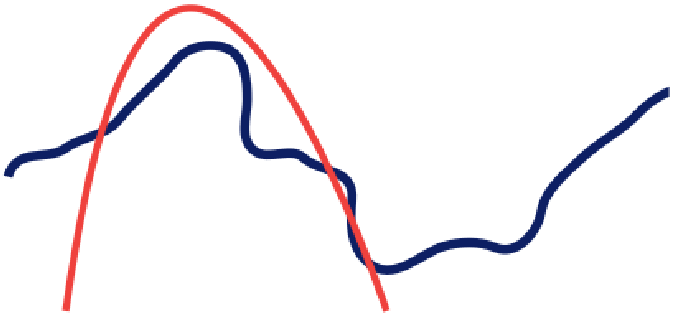
Approximate!



true distribution | variational distribution | Monte Carlo

VARIATIONAL INFERENCE

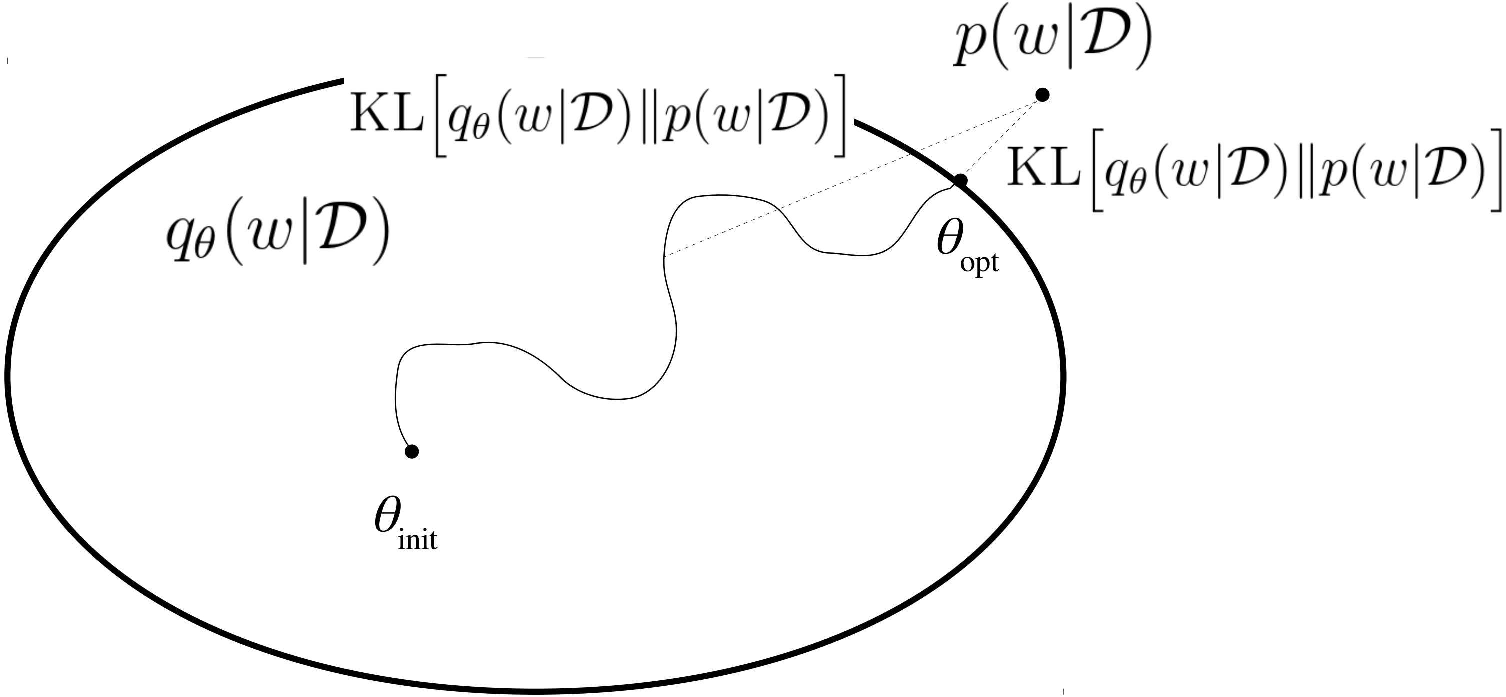
Learning variational distribution



$$q_\theta(w|\mathcal{D}) \approx p(w|\mathcal{D})$$

→ find parameters θ

Graphical intuition



(inspired by Graves, 2011)

KL-divergence

$$\mathrm{KL}\left[q_\theta(w|\mathcal{D})\|p(w|\mathcal{D})\right] = \boxed{\int} q_\theta(w|\mathcal{D}) \log \frac{q_\theta(w|\mathcal{D})}{p(w|\mathcal{D})} dw$$

$$= \mathbb{E}_{q_\theta(w|\mathcal{D})}\left[\log \frac{q_\theta(w|\mathcal{D})}{p(w|\mathcal{D})}\right]$$

INTRACTABLE

Finding θ_{opt}

$$\theta_{opt} = \arg\min_{\theta} \mathbb{E}_{q_\theta(w|\mathcal{D})}\left[\log \frac{q_\theta(w|\mathcal{D})}{p(w|\mathcal{D})}\right]$$

$$= \arg\min_{\theta} \mathbb{E}_{q_\theta(w|\mathcal{D})}\left[\log q_\theta(w|\mathcal{D})\right] - \mathbb{E}_{q_\theta(w|\mathcal{D})}\left[\log p(w|\mathcal{D})\right]$$

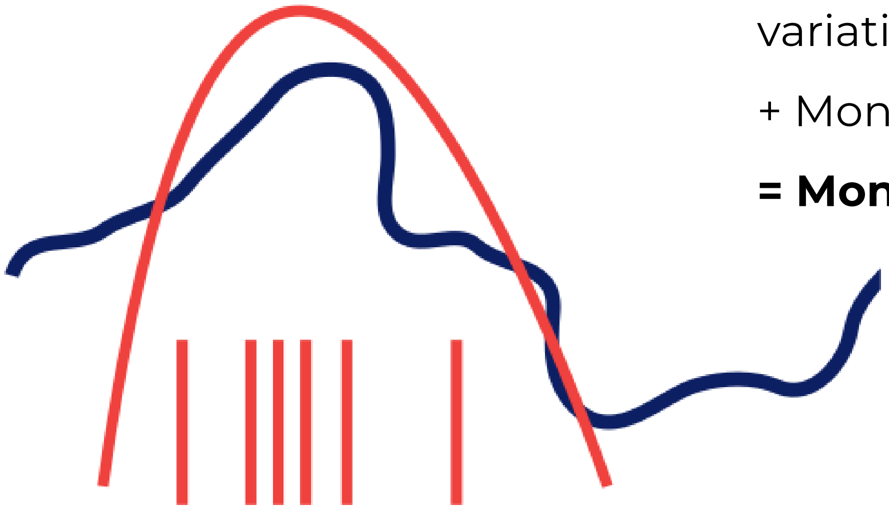
$$= \arg\min_{\theta} \mathbb{E}_{q_\theta(w|\mathcal{D})}\left[\log q_\theta(w|\mathcal{D})\right] - \mathbb{E}_{q_\theta(w|\mathcal{D})}\left[\log \frac{p(w,\mathcal{D})}{p(\mathcal{D})}\right]$$

$$= \arg\min_{\theta} \; -\left(\mathbb{E}_{q_\theta(w|\mathcal{D})}\left[\log p(w,\mathcal{D})\right] - \mathbb{E}_{q_\theta(w|\mathcal{D})}\left[\log q_\theta(w|\mathcal{D})\right]\right) + \underbrace{\log p(\mathcal{D})}_{\text{constant}}$$

REPHRASED, BUT STILL INTRACTABLE

(a) Bayes by Backprop

KL-divergence intractable → Approximate!



variational inference

+ Monte Carlo methods

= Monte Carlo variational inference

Monte Carlo variational inference

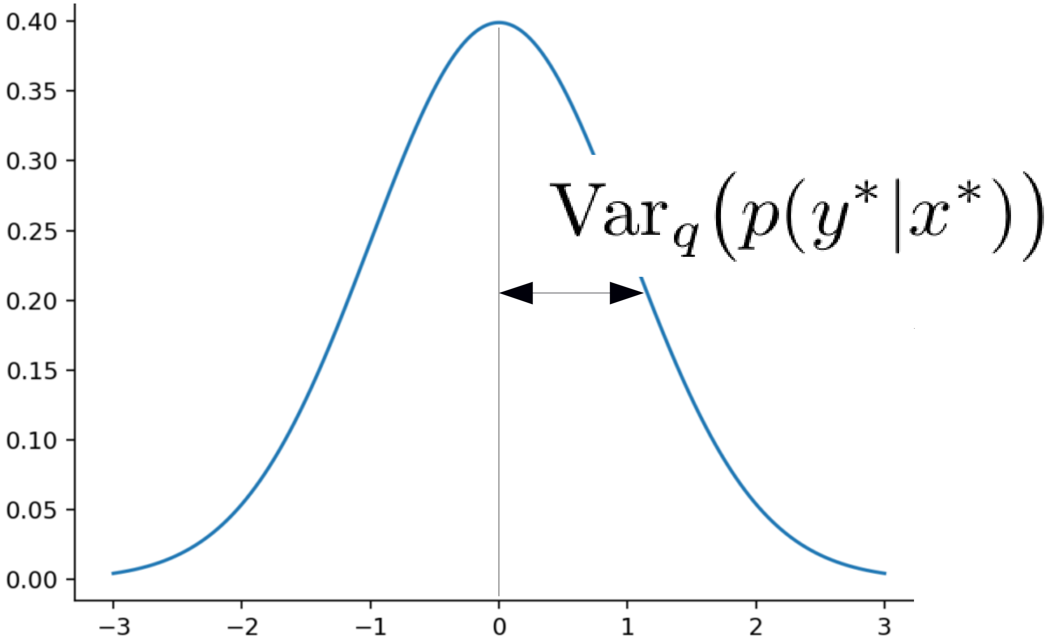
TRACTABLE

$$\theta_{opt} = \arg\min_{\theta} \boxed{\sum_{i=1}^{n}} \log q_\theta(w^{(i)}|\mathcal{D}) - \log p(w^{(i)}) - \log p(\mathcal{D}|w^{(i)})$$

with sample $w^{(i)}$ from $q_\theta(w|\mathcal{D})$

EPISTEMIC & ALEATORIC UNCERTAINTY

When it comes to predicting...



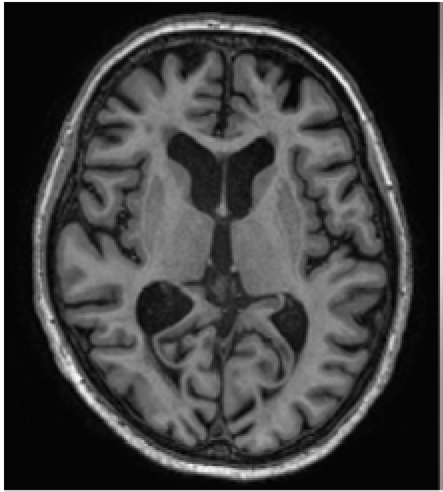
$$\mathrm{Var}_q(p(y^*|x^*)) = \text{aleatoric} + \text{epistemic}$$

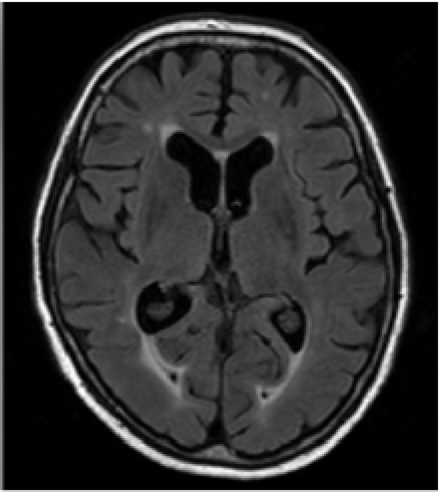
Uncertainties

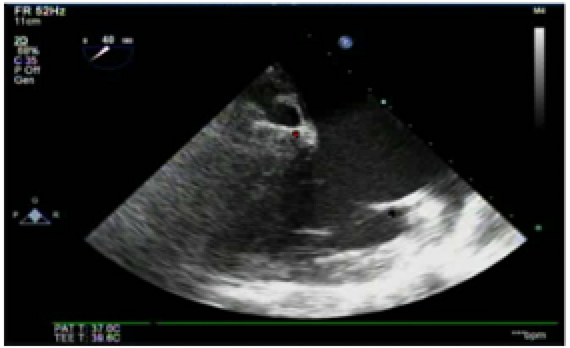
- Aleatoric: **dataset**-caused uncertainty
- Epistemic: **model**-caused uncertainty

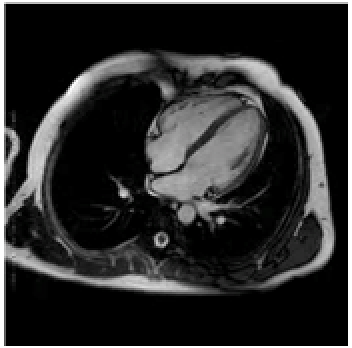
MEDICAL IMAGE ANALYSIS

Examples

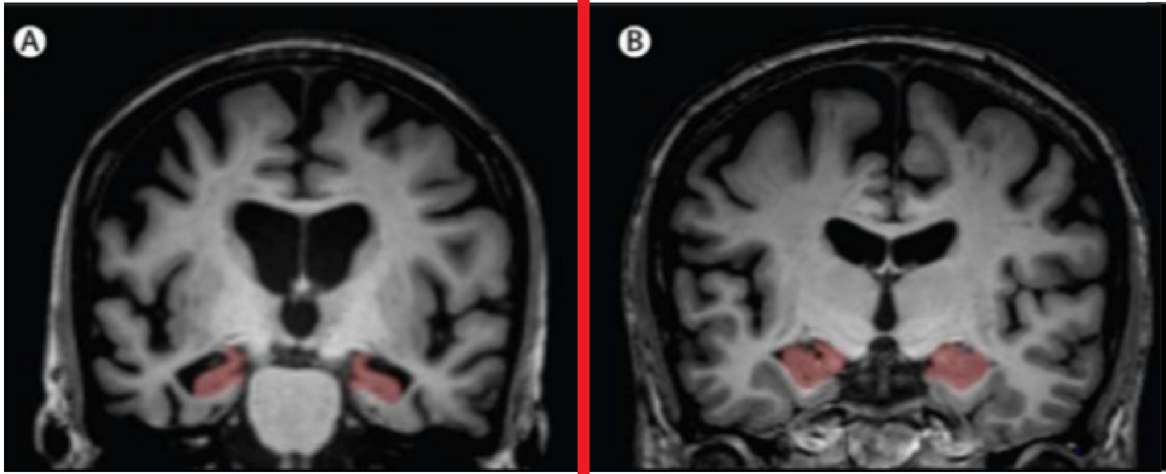








Brain MRI (Alzheimer *detection*)



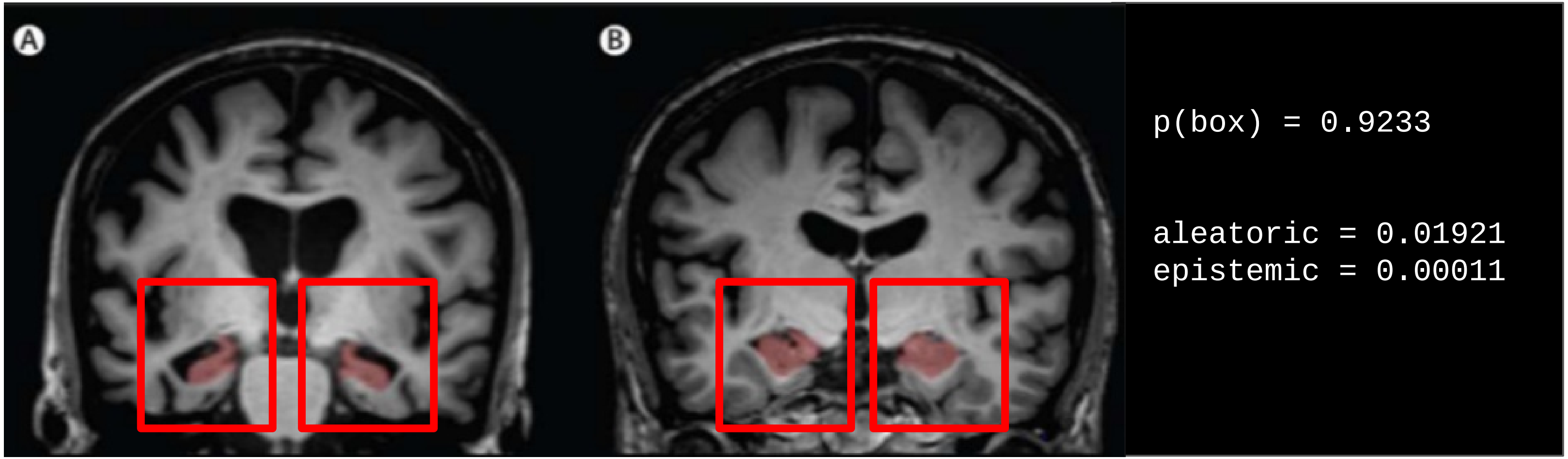
alzheimer

no alzheimer

```
p(alzheimer) = 0.9842

aleatoric = 0.02561
epistemic = 0.00078
```

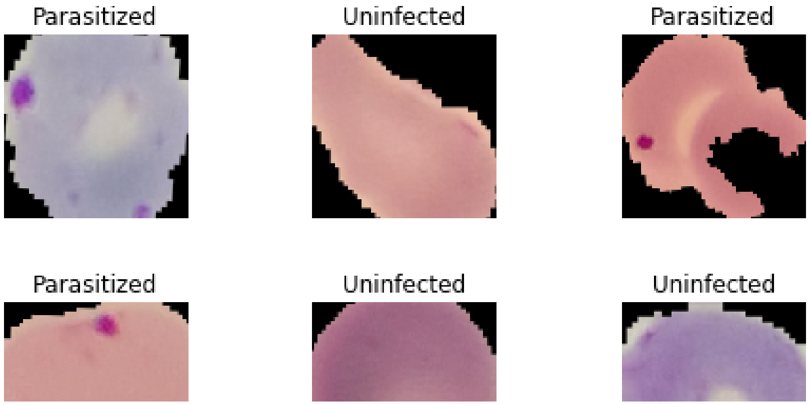
Brain MRI (Alzheimer *localization*)



alzheimer

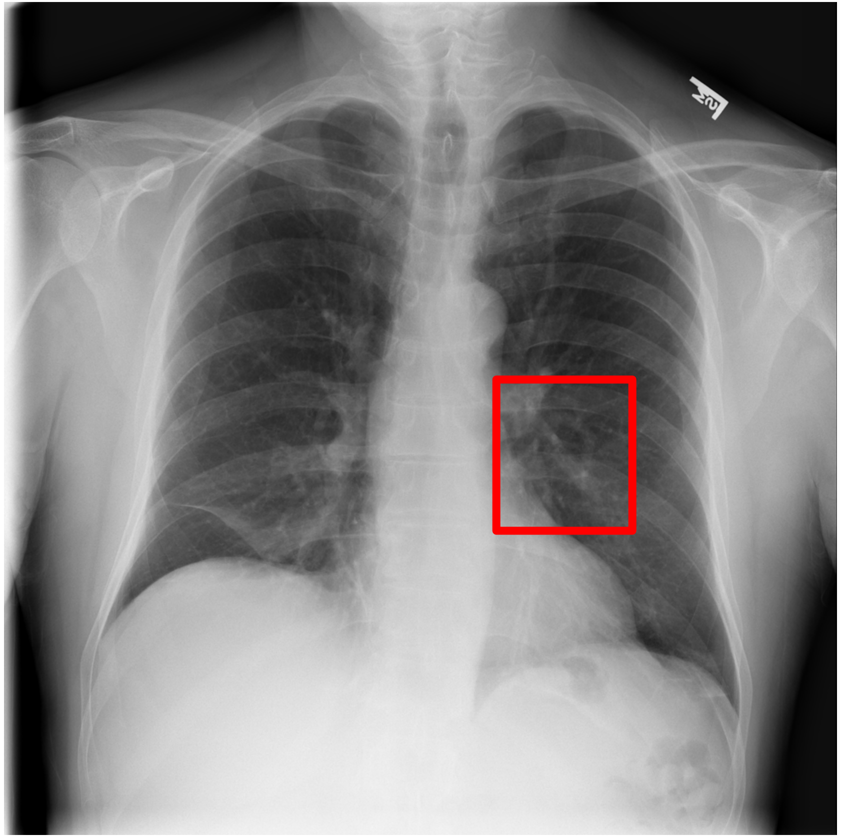
no alzheimer

Human cells (Malaria detection)



Data	100%		50%		25%	
	Parasitized	Uninfected	Parasitized	Uninfected	Parasitized	Uninfected
Bayesian AlexNet (with VI)	93	94	79	82	79	66
Frequentist AlexNet	94	96	89	87	86	85
Epistemic uncertainty ($*10^{-3}$)	0.0005		0.002		0.007	
Aleatoric uncertainty ($*10^{-3}$)	0.8		1.988		3.980	

Chest x-rays (lung diseases)



p(box) = 0.9187

aleatoric = 0.03998
epistemic = 0.00042

References

Kingma, Durk P., Tim Salimans, and Max Welling. "Variational dropout and the local reparameterization trick." Advances in Neural Information Processing Systems. 2015.

Blundell, Charles, et al. "Weight uncertainty in neural networks." arXiv preprint arXiv:1505.05424 (2015).

Graves, Alex. "Practical variational inference for neural networks." Advances in neural information processing systems. 2011.

Laumann, Felix. “Bayesian Convolutional Neural Networks with Bayes by Backprop.” (blog post)

Laumann, Felix. “Probabilistic Deep Learning: Bayes by Backprop.” (blog post)

Laumann, Felix. “What Uncertainties tell you in Bayesian Neural Networks.” (blog post)

THANK YOU

Felix Laumann, *felix@neuralspace.ai*

NeuralSpace

Imperial College London

ETH zürich

NeuralSpace

ELEMENT AI

DFKI

German Research Center for Artificial Intelligence

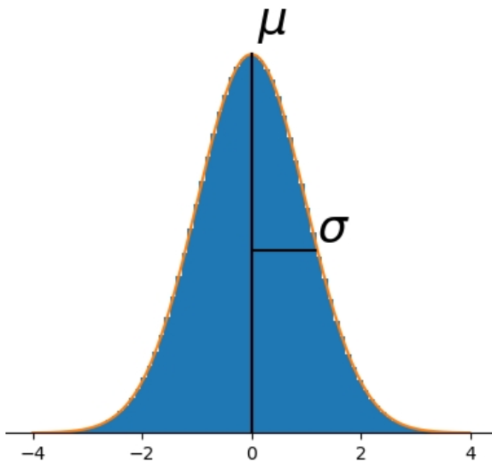
How to learn distributions?

local reparameterization trick

$$\theta = (\mu, \sigma^2)$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

$$f(\epsilon) = w = \mu + \sigma \cdot \epsilon$$



(Kingma et al, 2015)