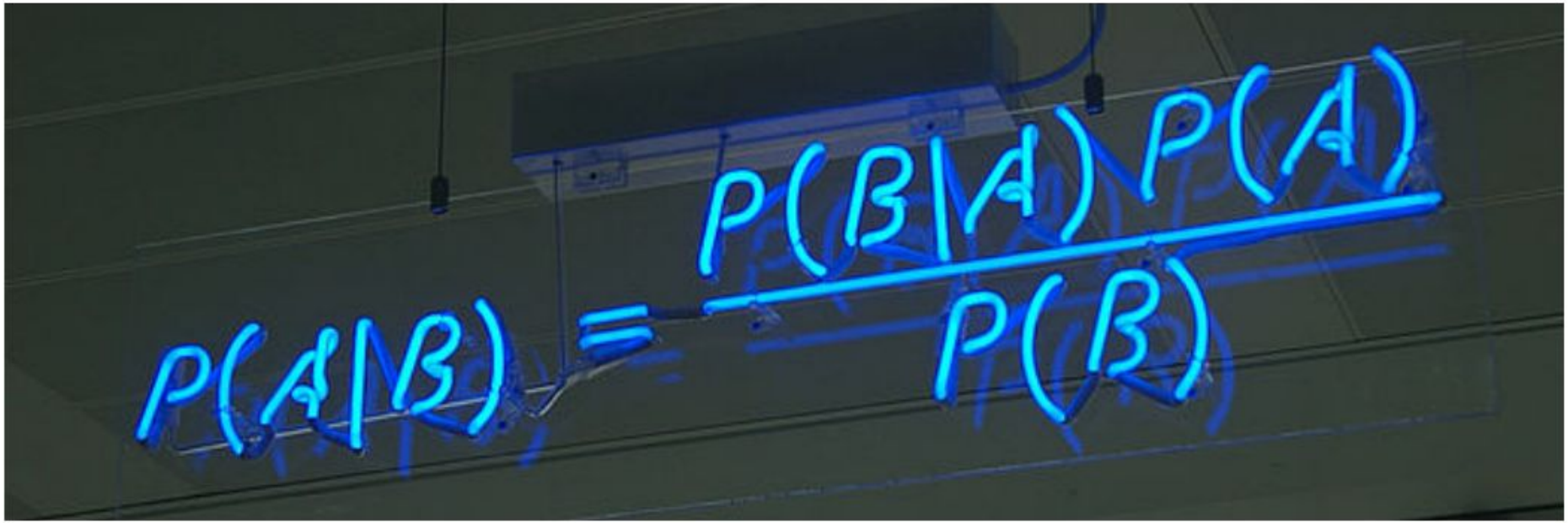
Bayesian Deep Learning

Mohd Adnan

Problems With Deep Learning

- What does a model not know?
- Uninterpretable black-boxes
- Easily fooled (AI safety)
- Lacks solid mathematical foundation
- Crucially relies on big dat
- Why does my model work
- What does my model know?

Bayesian Deep Learning



Bayesian Deep Learning

- Observed inputs X = {xi} and outputs Y = {yi}
- Capture stochastic process believed to have generated outputs
- Def. ω model parameters as random variable
- Prior dist. over ω: p(ω)
- Likelihood: p(Y|ω, X)
- Posterior: p(ω|X, Y) = p(Y|ω,X)p(ω) p(Y|X) (Bayes' theorem)
- Predictive distribution given new input x ∗ p(y ∗ |x ∗ , X, Y) = Z p(y ∗ |x ∗ , ω) p(ω|X, Y) | {z } posterior dω

Bayesian Deep Learning

Predictive distribution given new input $\mathbf{x}^*$

$$p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{X}, \mathbf{Y}) = \int p(\mathbf{y}^*|\mathbf{x}^*, \omega) \underbrace{\boxed{p(\omega|\mathbf{X}, \mathbf{Y})}}_{\text{posterior}} \mathrm{d}\omega$$

Why use Deep Network for Bayesian Learning?

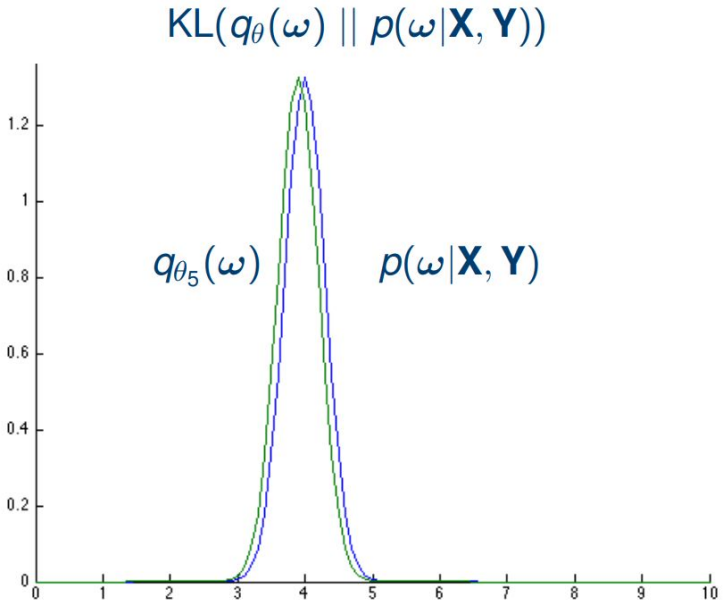
Predictive distribution given new input $\mathbf{x}^*$

$$p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{X}, \mathbf{Y}) = \int p(\mathbf{y}^*|\mathbf{x}^*, \omega) \underbrace{\boxed{p(\omega|\mathbf{X}, \mathbf{Y})}}_{\text{posterior}} \mathrm{d}\omega$$

Posterior is Intractable

Approximating Posterior with Deep Neural Networks

- Approximate p(ω|X, Y) with simple dist. q(ω)
- Minimise divergence from posterior



Advantages of Bayesian Deep Learning

- Can model uncertainty (Adversarial Attacks)
- Less prone to over-fitting due to prior distribution P(w)
- **With Bayesian modelling we can explain why**

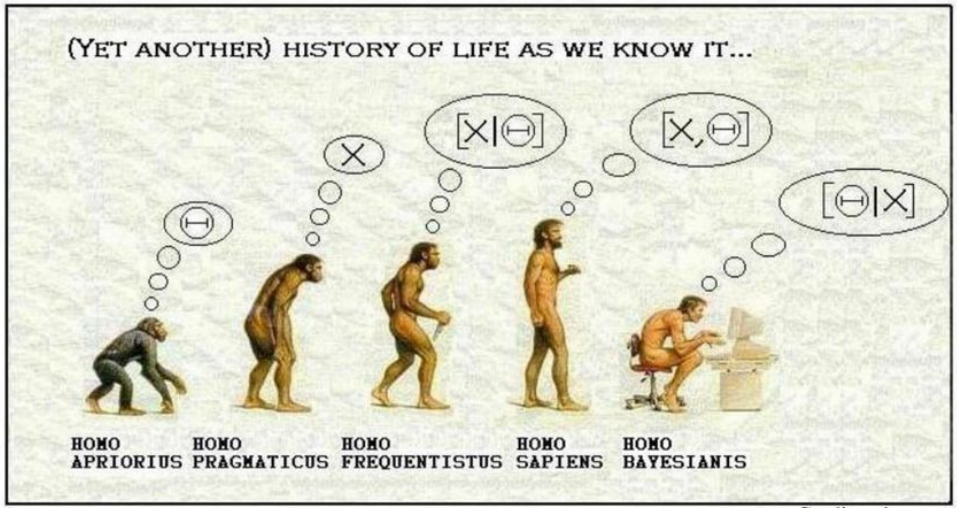
Fun Fact

Dropout is Bayesian Approximation

Deep Learning (Frequentist) vs Bayesian

Bayesian Methods

This is probably a more apt meme for us:



Credit:unknown

Bayesian Deep Learning: Two Schools of Thought

1. Bayesian Deep Learning is not useful unless you have a well defined prior.

2. Bayesian Deep Learning is useful as it act as ensemble of models

Carles Gelada
@carlesgelada

Good uncertainties are profoundly connected to generalization. If the prior used in BNNs isn't, the uncertainties will be useless. @jacobmbuckman and I provide a mathematical argument for that, and we even put into question if the B in BNN is doing much.

A Sober Look at Bayesian Neural Networks
by Carles Gelada and Jacob Buckman WARNING: This is an old version of this blogpost, and if you are a Bayesian, it might ...
jacobbuckman.com

11:53 PM · Jan 17, 2020 · Twitter Web App

54 Retweets **295** Likes

iandanforth @iandanforth · Jan 18
Replying to @carlesgelada and @jacobmbuckman
Bayesian language is so obfuscating. If I said my first "guess" doesn't matter, or that my "hypothesis" doesn't matter, it would sound absurd, but call it a "prior" and people start nodding along ...

2 2 11

Charles Sutton @RandomlyWalking · Jan 19
Whereas deep learning has no jargon, hmmm?

Let's not be catty or closed minded here. My current work is not Bayesian, but there is much more to it than you suggest. For example, hierarchical priors are incredibly powerful and worth knowing about.

Bayesian Deep Learning

NeurIPS 2019 Workshop

Friday, December 13, 2019

West Exhibition Hall C, Vancouver Convention Center, Vancouver, Canada

References:

1. http://mlg.eng.cam.ac.uk/yarin/PDFs/2015_UCL_Bayesian_Deep_Learning_talk.pdf
2. https://cims.nyu.edu/~andrewgw/caseforbdl/
3. https://jacobbuckman.com/2020-01-22-bayesian-neural-networks-need-not-concentrate/

Thanks!