Contents

Deep Bayesian Neural Networks Technical Report

Jing Zhang

2020 Jan.

1 Introduction

2 Bayes theorem

Thomas Bayes, a British mathematician, first proposed this theorem in a paper published in 1763. The paper was published by a friend after his death. In this paper, he proposed Bayes' theorem in order to solve an "Inverse probability" problem. Before "Inverse probability problem", people can calculate "Forward probability", that is probability that an event may happen.
The Bayes' theorem formula is as follows:

$$P(A|B) = P(A) * \frac{P(B|A)}{P(B)} \tag{1}$$

A and B are two events. Generally,

- event A is the problem to be solved;
- event B is known condition.

$P(A|B)$ **is the probability of event A happens after the event B happens.**
From the formula, we know there are three parts.

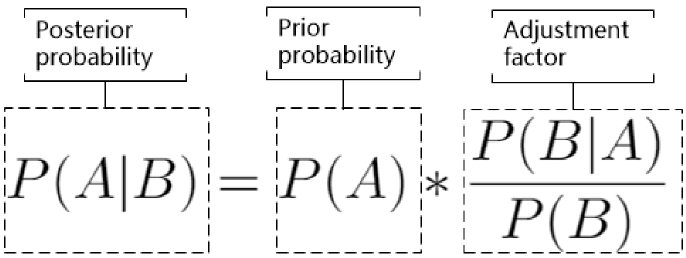

Figure 1: Composition of Bayes formula

- $P(A)$ is called Prior probability. That is, without knowing the occurrence of event B, we make a subjective judgment on the probability of occurrence of event A.
- $\frac{P(B|A)}{P(B)}$ is called Likely-hood. This is an adjustment factor, that is, the occurrence of the event B makes the prior probability closer to the true probability.
  - if Likely-hood $\frac{P(B|A)}{P(B)} > 1$, means that the "prior probability" is enhanced and the probability of occurrence of event A becomes greater.
  - if Likely-hood $\frac{P(B|A)}{P(B)} = 1$, means that event B does not help judge the likelihood of event A.
  - if Likely-hood $\frac{P(B|A)}{P(B)} < 1$, means that the "prior probability" is weakened and the probability of event A becomes smaller.
- $P(A|B)$ is called Posterior probability. That is, after the occurrence of event B, we re-evaluate the probability of event A.

The underlying idea of Bayes:
If we can grasp all the information about a thing, Absolutely we can calculate an objective probability (classical probability, forward probability). However, most of the information in decision-making in life is incomplete, and we have only limited information in our hands. Since comprehensive information is not available, we make as good a prediction as possible with limited information. That is, on the basis of subjective judgment, a value (a prior probability) can be estimated first, and then continuously modified (Likely-hood function) based on the new information observed.
In order to calculate the $P(B)$, we compliment a concept called **Law of total probability**.

$$P(B) = P(B|A)P(A) + P(B|A')P(A') \tag{2}$$

The meaning of this formula is that if event A and A' constitute the entirety of the problem (the entire sample space), then the probability of event B is equal to the the sum of probability of A multiplied by the conditional probabilities of B and A' multiplied by the conditional probabilities of B, respectively.

3 Bayesian deep learning

Having a distribution instead of a single value is a powerful thing. For one, it becomes possible to sample from a distribution many many times and see how this affects the predictions of the model. If it gives consistent predictions, sampling after sampling, then the net is said to be "confident" about its predictions[1].

$$p(w|x, y) = p(w) \cdot \frac{p(x, y|w)}{\int p(x, y|w)p(w)dw} \tag{3}$$

Three ways to approximate the formula above:

[1] https://medium.com/@joeDiHare/deep-bayesian-neural-networks-952763a9537

- Approximating the integral with MCMC.
- Using black-box variational inference (with edward).
- Using MC dropout.

4 Experiment

References