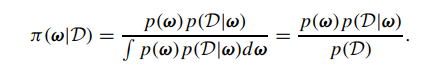
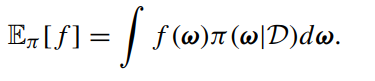
2.5.1 先进神经网络

与传统的神经网络不同之处在于，贝叶斯神经网络中的权值不是固定不变的。该模型的思想是将神经网络和随机建模的优势结合起来，可以为预测提供概率上的保证[文献]。这里的权值是基于我们观察到的来推断的，是一个逆概率问题，需要通过贝叶斯定理来解决[文献]。这里权重w代表的是无法直接观察分布的不可见变量，无法直接观察真实分布，贝叶斯允许我们得到模型参数在已观察到的数据条件下的分布p(ω|D)，即后验分布。

同时，p(D|w) 似然function在多分类问题里将神经网络的预测（经过Log Softmax转换）与实际观测类别标签相结合，来量化他们之间的一致性。在确定似然项以及p(w)先验分布后，便可根据贝叶斯理论计算后验分布，



得到后验分布后，将使用边缘化方法进行预测：



2.5.2 推理方法：

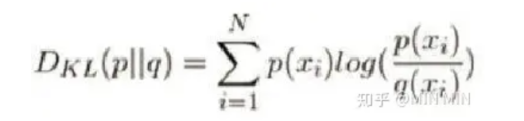
推理方法适用于解决这个归一化因子即证据项在高维中难以处理的问题。因此需要采用近似技术来获得后验分布。除此以外，贝叶斯推理方法还可以根据模型预测的不确定性来寻找最有可能提高模型性能的样本。主流的BNN有有两种方式，一种是基于采样的方式（马尔可夫链蒙特卡洛），或者是变分推断[文献]。

Markov Chain Monte Carlo (MCMC)

MCMC是目前最流行的基于采样来估计后验分布的方式，主要思想是构建前后样本的依赖关系[78]，进而模拟分布。但是由于连续样本可能出现自相关的情况，则需要更加庞大的样本集来抽样去获得近似独立样本，这样做会增加计算成本。因此，对于BNN来说，Metropolis-Hasting算法更加有效[文献]。通过与分布成比例的函数，省去获取精确概率分布的过程，并使用接受-拒绝机制来完成对候选样本的选取。如果样本比前一个数据点出现概率低则拒绝样本。除此以外，汉密尔顿蒙特卡罗算法（HMC）[81]是更加改进的方式，通过更新提议来拒绝少量样本的同时减少相关性来缩短burn-in时间。

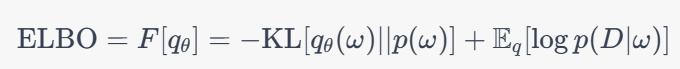
Variational Inference

变分推理是相较于MCMC的一种提高计算速度的方式，通过优化参数，在VI定义的参数化数族中找到最优近似分布。主要思想是结合KL散度来定义分布的差距，当两个分布的差异越大时，KL散度就越大：



P(x)是目标分布，q(x)是去匹配的分布，如果这两个分布完全匹配则。

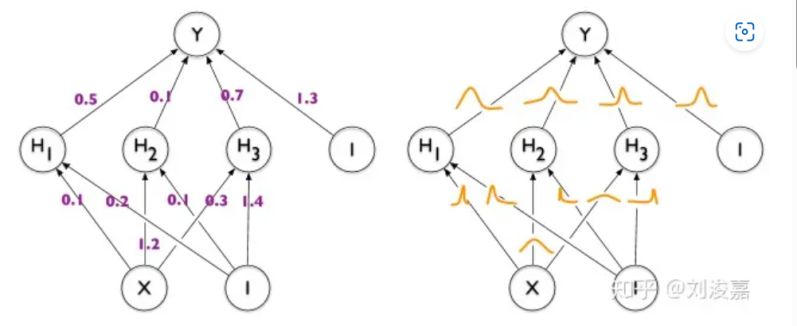
但是，真实后验分布是未知的，直接优化KL散度并不简单，在此论文中将KL公式进行转化为计算Evidence Lower Bound（ELBO），最小化KL散度等于最大化ELBO[文献]。



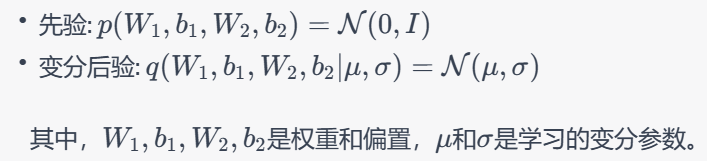
这里的ELBO（证据下界）包括：变分分布和真实先验之间的KL散度的负值，以及给定模型参数的数据似然的期望。

贝叶斯神经网络

发现贝斯斯网络在MNIST数据集上表现良好，适合对有略微扭曲的数据集进行预测，证明了极强的抵抗力[文献i don’t know]，因此此实验将尝试使用贝叶斯神经网络增加模型的鲁棒性。在模型结构上，与普通神经网络结构相同的是他们有相同数量的隐藏层，以及激活函数，并具有大小为’input\_size’的输入。不同之处是贝叶斯神经网络使用Dropout正则化，在训练期间用’dropout\_prob’的概率来随机关闭部分输入。不同于传统神经网络的权重，在这里权重被视为有概率分布的随机变量，如下图所示。



结合第二章的描述，贝叶斯神经网络通过‘model’方式定义权重和偏置的先验，并描述了似然分布，公式如下所示。通过‘guide’方式找到逼近真实后验分布的变分分布。



在训练过程中，我们采用了与普通神经网络不同的方法。通过使用Pyro库和随机变分推断（SVI），我们合并了优化的各个步骤，而不是分开执行。SVI的目标是通过迭代最大化证据下界（ELBO）来逼近真实的后验分布。

在对验证集和测试集的预测方面，贝叶斯神经网络与传统网络的工作方式不同。它通过多次采样来生成多个预测结果，而不是仅依赖一个固定模型。这种方式增加了预测的多样性，并反映了模型的不确定性。最后，在确定最可能的类别标签时，我们采取了一种整合方法，即取所有预测结果的均值，然后使用`np.argmax`。这种方法可能提供更鲁棒的预测。

（第二章bert那里）Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In NAACL-HLT.

BALD介绍的第一个文献：Gal, Y., Islam, R., & Ghahramani, Z. (2017, July). Deep bayesian active learning with image data. In *International conference on machine learning* (pp. 1183-1192). PMLR.

@article{mullachery2018bayesian,

title={Bayesian neural networks},

author={Mullachery, Vikram and Khera, Aniruddh and Husain, Amir},

journal={arXiv preprint arXiv:1801.07710},

year={2018}

}

@article{goan2020bayesian,

title={Bayesian neural networks: An introduction and survey},

author={Goan, Ethan and Fookes, Clinton},

journal={Case Studies in Applied Bayesian Data Science: CIRM Jean-Morlet Chair, Fall 2018},

pages={45--87},

year={2020},

publisher={Springer}

}

@inproceedings{zhao2018empirical,

title={An empirical evaluation of Bayesian inference methods for Bayesian neural networks},

author={Zhao, Rui and Ji, Qiang},

booktitle={NIPS Bayesian Deep Learning (BDL) Workshop},

year={2018}

}

@article{valentin2020hands,

title={Hands-on Bayesian neural networks--a tutorial for deep learning users},

author={Valentin Jospin, Laurent and Buntine, Wray and Boussaid, Farid and Laga, Hamid and Bennamoun, Mohammed},

journal={arXiv e-prints},

pages={arXiv--2007},

year={2020}

}

1. R. Bardenet, A. Doucet, and C. Holmes, “On Markov Chain Monte Carlo methods for tall data,” J. Mach. Learn. Res., vol. 18, no. 1, pp. 1515–1557, Jan. 2017.

@article{chib1995understanding,

title={Understanding the metropolis-hastings algorithm},

author={Chib, Siddhartha and Greenberg, Edward},

journal={The american statistician},

volume={49},

number={4},

pages={327--335},

year={1995},

publisher={Taylor \& Francis}

}

@article{thomas2021learning,

title={Learning hamiltonian monte carlo in r},

author={Thomas, Samuel and Tu, Wanzhu},

journal={The American Statistician},

volume={75},

number={4},

pages={403--413},

year={2021},

publisher={Taylor \& Francis}

}

Hoffman, M. D., & Johnson, M. J. (2016, December). Elbo surgery: yet another way to carve up the variational evidence lower bound. In Workshop in Advances in Approximate Bayesian Inference, NIPS (Vol. 1, No. 2).

Myojin, T., Hashimoto, S., & Ishihama, N. (2020). Detecting uncertain BNN outputs on FPGA using Monte Carlo dropout sampling. In Artificial Neural Networks and Machine Learning–ICANN 2020: 29th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 15–18, 2020, Proceedings, Part II 29 (pp. 27-38). Springer International Publishing.

Tsymbalov, E., Panov, M., & Shapeev, A. (2018). Dropout-based active learning for regression. In Analysis of Images, Social Networks and Texts: 7th International Conference, AIST 2018, Moscow, Russia, July 5–7, 2018, Revised Selected Papers 7 (pp. 247-258). Springer International Publishing.

Grezes, F. (2022). Finite Gaussian Neurons: Defending Against Adversarial Attacks by Making Neural Networks Say" I Don’t Know" (Doctoral dissertation, City University of New York).

S

