CSE 584 HW1 (Yucheng Zhou)

Paper 1: Deep Bayesian Active Learning with Image Data (https://arxiv.org/abs/1703.02910)
Motivation of this paper: A major challenge in active learning is lack of scalability to high-dimensional data. For instance, data in image form, like dermoscopic lesion images for cancer diagnosis.

How it is solved: They combined recent advances in Bayesian deep learning into the active learning framework. And developed an active learning framework for high dimensional data. They utilized Bayesian CNNs where was used as a variational Bayesian approximation. First they used a subset of the training set to train the model. Then this model is used to predict the unlabeled images, and each one's entropy is recorded as the uncertainty score. Next, new training points are labeled and added according to the scores. Finally, the model is re-trained and updated until certain conditions are satisfied.

Contributions: This work extend the active learning to high-dimensional data by incorporating Bayesian deep neural networks. And it achieved better performance on MNIST dataset compared with previous works like DGN and Ladder Network. In addition, they applied this method to skin cancer diagnosis task by using a small set of lesion images and fine-tuning VGG16 network

Downsides: The model performance is assessed by resetting and re-training after each acquisition. This will require more training time. This can be further studied and possibly reduced to ensure a faster convergence. In addition, they used a naive batch acquisition technique, which may not obtain jointly informative data points.

Paper 2: BatchBALD: Efficient and Diverse Batch Acquisition for Deep Bayesian Active Learning (https://arxiv.org/abs/1906.08158)
Motivation of this paper: In a previous work, a naive batch acquisition technique was used in Deep Bayesian Active Learning (the acquisition of top several points according to the BALD acquisition score). This will acquire data points that are individually very informative but not necessarily jointly informative.

How it is solved: Naive batch acquisition technique also enumerates all possible subsets of data, thus the number of the potential subsets expands exponentially. This work proposeed BatchBALD as an extension of BALD whereby the points were jointly scored by estimating the mutual information between a joint of multiple data points and the model parameters. They developed a greedy algorithm that selects a batch in linear time, and showed that it can give very good approximations to the optimal choice.

Contributions: First, they proposed BatchBALD, a data-efficient active learning method that acquires batches of image data. Second, they gave a greedy algorithm to select a batch of points efficiently. Third, they proposed an estimator for the acquisition function that scales to larger sizes and to more complex datasets.

Downsides: While their method can significantly reduce the number of data points that need to be labelled and the number of times the model has to be retrained, it also comes with additional computation cost in the batch acquisition process. This can be further improved and the possible acquisition can be further reduced to speed up this process.

Paper 3: The Relevance of Bayesian Layer Positioning to Model Uncertainty in Deep Bayesian Active Learning (https://arxiv.org/abs/1811.12535)

Motivation of this paper: One major challenge to implementing and using Bayesian CNNs in active learning is the time and difficulty required to train them. Also, whether fully Bayesian neural networks can effectively capture the uncertainty in a problem deserve further study.

How it is solved: They varied the number and position of Bayesian layers in active learning and the weight distribution initialization in the CNNs to examine their ability to capture uncertainty on MNIST.

Contributions: They showed that it is unnecessary to use fully Bayesian CNNs in active learning for capturing model uncertainty. Their results on several Bayesian CNNs confirm that in order to represent the model uncertainty, one needs to apply the Bayesian prior on only a few last layers before the output.Then the model would enjoy the benefits of both deterministic and Bayesian CNNs.

Downsides: The network they used is small and simple. It would be better if this work can be extended to larger networks. Also, extending the experiment to examine the effect of different types of priors on Bayesian CNN performance is also a possible improvement.