Your classifier is secretly an Energy Based Model and you should treat it like one

Prior work in generative models has focussed on their adaptation towards downstream tasks. Recent work, on other hand, emphasizes on sample quality and likelihood of validation sets. This results in a performance gap between best generative and conventional discriminative frameworks. To address these challenges, the work proposes to reinterpret standard discriminative classifiers of the form $p(y|x)$ as Energy-Based Models (EBMs) which approximate the joint distribution $p(x, y)$. This elegant reinterpretation of models provides a generalized framework wherein discriminative classifiers can be incorporated in conjunction with unlabeled data. The Joint Energy-based Model (JEM) scheme demonstrates improved calibration, adversarial robustness and Out-Of-Distribution (OOD) detection.

Potential of EBMs towards downstream discriminative problems is an active area of work. The paper advances in this direction by realizing discriminative models as a joint framework for modelling labels and data. One can think of $p(x, y)$ as an energy-based formulation consisting of the energy function $-f(x)[y]$ and partition function $Z(\theta)$. Moreover, the discriminative framework $p(y|x)$ is an energy-based scheme consisting of $f(x)[y]$ as the energy function and $Z(\theta)$ as the partition function. Upon marginalizing the labels in the joint model $p(x, y)$ we obtain $p(x)$ wherein the LogSumExp operator can be used to characterize the energy of a data point x. This allows the joint energy-based scheme to make use of one extra degree of freedom hidden with the logits $f(x)$. Moroever, JEM is a strict generalization of the discriminative framework which can be obtained by dividing $p(x, y)$ with $p(x)$. JEM framework, in practice leverages this insight and constructs an optimization objective consisting of likelihoods $\log p(x)$ and $\log p(y|x)$ with the former is approximated using SGLD and the latter using standard cross-entropy.

The JEM framework demonstrates suitability for a wide variety of tasks. Firstly, JEM is found suitable for hybrid modeling wherein the approach presents improved accuracy and inception scores in comparison to discriminative and generative models. Secondly, JEM highlights an apt neccessity for calibration in classifiers based on model's confidence levels. Lastly, JEM presents suitable detection of OOD data by assigning accurate likelihood estimates to OOD samples in the joint data distribution. Additionally, adversarial robustness of JEM is validated by carrying out a number of black-box and white-box attacks wherein its performance is found comparable to state-of-the-art methods. While JEM is a suitable mechanism for reinterpreting classifiers as generative models, its training presents a number of caveats given its current energy-based form. JEM, being an energy-based scheme, requires many steps of SGLD updates which acts as a computational bottleneck. JEM is also found to be unstable during training wherein the setup needs to restart multiple times. Additionally, experimental setup provides a novel evaluation of calibration in discriminative classifiers but does not throw light on its adaptation during training and testing phases. For instance, how can calibration be adapted to unsupervised scenarios which constitute the majority of downstream tasks for EBMs?

The joint framework provides a myriad of new directions for future work. To list two of them, JEM paves the way for application of EBMs as discriminative models making use of additional degrees of fredom. Lastly, JEM highlights the importance of stabilizing EBM training for downstream tasks under computational constraints.