DS-GA 1006 Fall 2017
Capstone Project Status Update I

Voyagers:
Yurui Mu, Sudhir Nallam
Xiaoyu Wang, Wenqing Zhu

Nov 16, 2017

---

1. **Abstract**

   a. Fixating on one dataset, we would like to continue comparing different model performances.
   b. We would like to reproduce the results in QCD paper with other datasets.
   c. Apply Gated RNN model on noisy dataset to check if performance is conditional on noise parameter.
   d. Attach adveserial loss piece to original RNN structure to train a classifier pivot against noise parameter.
   e. We don't see any foreseeable problems at this time and planning to deliver all our project goals within deadline.

2. **Details**

In the past week, we have closely read [1] and understood the ideas behind finding a decision function robust to nuisance parameters. To further explain, we have establised our goal as to train a classifer f that is invariant to nuisance factor z, probalisitically, this is equavalent to find:

$$p(f(X;\theta_f) = s|z) = p(f(X;\theta_f) = s|z') \tag{1}$$

Alternatively speaking, we are looking for a classification function f as a pivotal quantity with respect to the nuisance parameter. This implies that $f(X;\theta_f)$ and Z will be independent random variables.
Aside from our original loss term $L_f(\theta_f)$ we'll then also include an adversrial loss piece, which together will be our new objective function:

$$E(\theta_f, \theta_r) = L_f(\theta_f) - L_r(\theta_f, \theta_r) \tag{2}$$

$$\hat{\theta}_f, \hat{\theta}_r = arg \min \max E(\theta_f, \theta_r) \tag{3}$$

we'll find the minimax solution through applying stochastic gradient descent in a alternatively.

Meanwhile, to complete our roadmap in model experiments, we have submitted tasks on HPC to train Gated Recurrent Neural Network model. We then can compare the AUC score and ROC curve with our previous results. More specifically, we are going to compare the background efficiency when signal efficiency is constrained at 0.5 level with previous results.

As we have observed performance improvement by using three layers of 1x1x1 convolution, we want to further parametrize the activation function by allowing to control the number of convolution layers to use, and then optimize for best result.

References

[1] Gilles Louppe, Michael Kagan, and Kyle Cranmer. Learning to Pivot with Adversarial Networks. 2016.