Name: ____________________

University of Illinois Spring 2020

CS 446/ECE 449 Machine Learning

Homework 9: Generative Adversarial Nets (GANs)

Due on Tuesday April 28 2020, noon Central Time

1. **[24 points]** Generative Adversarial Nets (GANs) and Duality

   Consider the following program for a dataset $\mathcal{D} = \{(x)\}$ of points:

$$\max_{\theta} \min_{w} -\sum_{x\in\mathcal{D}} \log p_w(y=1|x) - \sum_{z\in\mathcal{Z}} \log(1 - p_w(y=1|G_\theta(z))) + \frac{C}{2}\|w\|_2^2. \quad (1)$$

   Hereby θ denotes the parameters of the generator $G_\theta(z)$, which transforms 'perturbations' $z \in \mathcal{Z}$ into artificial data, w refers to the parameters of the discriminator model $p_w(y|x)$, $y \in \{0,1\}$ denotes artificial or real data, and $C \geq 0$ is a fixed hyper-parameter.

   (a) (1 point) What is the original motivation (the one used in Goodfellow *et al.* (NIPS'14)) underlying generative adversarial nets (GANs)?

   Solution:
   Two-player game with generator producing artificial samples and discriminator trying to tell apart artificial samples from real data.

   (b) (1 point) Without restrictions on the generator model G_θ and the discriminator model p_w, what are challenges in solving the program given in Eq. (1)?

   Solution: Saddle point objective which is neither convex in w nor concave in θ.

   (c) (2 points) We now restrict the discriminator as follows:

$$p_w(y=1|x) = \frac{1}{1+\exp w^\top x}.$$

   Using this discriminator, write down the resulting cost function for the program given in Eq. (1).

   Solution:

$$\sum_{x\in\mathcal{D}} \log(1+\exp w^\top x) + \sum_{z\in\mathcal{Z}} \log(1+\exp(w^\top G_\theta(z))) - \sum_{z\in\mathcal{Z}} w^\top G_\theta(z) + \frac{C}{2}\|w\|_2^2$$

   (d) (2 points) When is the function $\frac{C}{2}\|a\|_2^2 - a^\top b$ convex in a? Why?

   Solution: Hessian is CI which is positive semi-definite if $C \geq 0$.

   (e) (2 points) When is the function $\log(1+\exp a^\top b)$ convex in a? Why?

   Solution: Hessian is

$$\frac{\exp a^\top b}{(1+\exp a^\top b)^2} bb^\top$$

   which is always positive semi-definite.

   (f) (2 points) Assume we restrict ourselves to the domain (if any) where $\frac{C}{2}\|a\|_2^2 - a^\top b$ and $\log(1+\exp a^\top b)$ are convex in a, what can we conclude about convexity of the function

$$\sum_{x\in\mathcal{D}} \log(1+\exp w^\top x) + \sum_{z\in\mathcal{Z}} \log(1+\exp(w^\top G_\theta(z))) - \sum_{z\in\mathcal{Z}} w^\top G_\theta(z) + \frac{C}{2}\|w\|_2^2$$

Name: ___________________________

in w and why?

Solution: Function is convex in w since a sum of convex functions is convex.

(g) (2 points) Let us introduce variables $\xi_x = w^\top x$ and $\xi_z = w^\top G_\theta(z)$ and let us consider the following program:

$$\min_w \quad \sum_{x\in\mathcal{D}} \log(1+\exp\xi_x) + \sum_{z\in\mathcal{Z}} \log(1+\exp\xi_z) - \sum_{z\in\mathcal{Z}} w^\top G_\theta(z) + \frac{C}{2}\|w\|_2^2 \qquad (2)$$

$$\text{s.t.} \quad \left\{\begin{array}{lcllc} \xi_x & = & w^\top x & \forall x\in\mathcal{D} & (\text{C1}) \\ \xi_z & = & w^\top G_\theta(z) & \forall z\in\mathcal{Z} & (\text{C2}) \end{array}\right. .$$

What is the Lagrangian for this program? Use the Lagrange multipliers λ_x and λ_z for the constraints (C1) and (C2) respectively.

Solution:

$$\sum_{x\in\mathcal{D}} [\lambda_x\xi_x + \log(1+\exp\xi_x)] + \sum_{z\in\mathcal{Z}} [\lambda_z\xi_z + \log(1+\exp\xi_z)]$$
$$-w^\top \left[\sum_{x\in\mathcal{D}} \lambda_x x + \sum_{z\in\mathcal{Z}} (1+\lambda_z) G_\theta(z)\right] + \frac{C}{2}\|w\|_2^2$$

(h) (2 points) What is the value of

$$\min_w \frac{C}{2}\|w\|_2^2 - w^\top b$$

in terms of b and C?

Solution:

$$-\frac{1}{2C}\|b\|_2^2$$

(i) (2 points) What is the value of

$$\min_\xi \lambda\xi + \log(1+\exp\xi)$$

in terms of λ? What is the valid domain for λ?

Solution:

$$\xi = \log\frac{-\lambda}{1+\lambda} \qquad -1 \le \lambda \le 0$$

Value is:

$$\lambda\log(-\lambda) - (1+\lambda)\log(1+\lambda)$$

(j) (6 points) Combine your results from the previous two sub-problems to derive the dual function of the program given in Eq. (2). Also state the dual program and clearly differentiate it from the dual function. State how this dual program can help to address a challenge in GAN training.

Name:

Solution: Dual function:

$$g(\lambda) = \frac{-1}{2C} \| \sum_{x \in \mathcal{D}} \lambda_x x + \sum_{z \in \mathcal{Z}} (1+\lambda_z) G_\theta(z) \|_2^2 + \sum_{x \in \mathcal{D}} H(\lambda_x) + \sum_{z \in \mathcal{Z}} H(\lambda_z)$$

where $H(\lambda) = \lambda \log(-\lambda) - (1+\lambda)\log(1+\lambda)$.
Dual program:

$$\max_{-1 \le \lambda_x \le 0, -1 \le \lambda_z \le 0} g(\lambda)$$

GAN training:

$$\max_{\theta} \max_{-1 \le \lambda_x \le 0, -1 \le \lambda_z \le 0} g(\lambda)$$

searches for a global maximizer rather than a saddle-point.

(k) (2 points) Implement and provide the loss for the discriminator and the generator when using the '-log-D' trick in A9_GAN.py.

Solution:
Discriminator loss:

```
loss = criterion(logit, target1)
```

Generator loss:

```
loss = criterion(logit, target2)
```