Name: Yuxuan Zhang (yuxuanz8)

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)?

Your answer:
The generator G. tries to trick discriminator by generating samples hard to distinguish.

(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)?

Your answer:
$\max_\theta \min_w$. objective function requires p_w to be convex and G_θ to be concave in θ in order to obtain the optimal

(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).

Your answer:

$$\max_\theta \min_w + \sum_{x\in D} \log(1+\exp w^T x) - \sum_{z\in Z} \log\left(1 - \frac{1}{1+\exp w^T G_\theta(z)}\right) + \frac{C}{2}\|w\|_2^2$$

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

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

Name: Yuxuan Zhang (yuxuanz8)

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

Your answer: $f(a) = \frac{C}{2}\|a\|_2^2 - a^T b$.

$\frac{\partial f(a)}{\partial a} = \frac{C}{2} \cdot 2a - b = Ca - b$

$\frac{\partial^2 f(a)}{\partial a^2} = C \geq 0. \quad \forall a.$

$C \geq 0$. then the function is convex. if and only if $f''(a) \geq 0$.

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

Your answer: $g(a) = \log(1 + \exp a^T b)$

$g'(a) = \frac{\exp(a^T b) \cdot b}{1 + \exp(a^T b)}$

$g''(a) = b^T \frac{\exp(a^T b) \cdot b(1 + \exp(a^T b)) - \exp(a^T b)[\exp(a^T b) \cdot b]}{[1 + \exp(a^T b)]^2}$

$= \frac{\exp(a^T b)(b^T b)}{[1 + \exp(a^T b)]^2} \geq 0.$

The Hessian derivative is always positive semi-definite so the function $\log(1 + \exp a^T b)$ is always convex.

(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

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

in w and why?

Your answer:

We can conclude that this function is convex because sum of convex function is still convex.

$\underbrace{\sum_{x \in D} \log(1 + \exp w^T x)}_{\text{convex}} + \underbrace{\sum_{z \in Z} \log(1 + \exp(w^T G_\theta(z)))}_{\text{convex.}} - \sum_{z \in Z} w^T G_\theta(z) + \frac{C}{2}\|w\|_2^2$

$\underset{a}{\frac{C}{2}\|w\|_2^2} - \underset{a.}{w^T}\Big[\sum_{z \in Z} G_\theta(z)\Big]^{b.}$ is also convex.

Name: Yuxuan Zhang (yuxuanz8)

(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:

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

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

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

Your answer: $\xi_x - w^T x = 0.$

$\xi_z - w^T G_\theta(z) = 0$

$$\mathcal{L} = \sum_{x\in D} \log(1+\exp\xi_x) + \sum_{z\in Z}\log(1+\exp\xi_z) - \sum_{z\in Z} w^T G_\theta(z) + \frac{C}{2}\|w\|_2^2 + \sum_{x\in D}\lambda_x(\xi_x - w^T x) + \sum_{z\in Z}\lambda_z(\xi_z - w^T G_\theta(z))$$

(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?

Your answer: $C \geq 0$. It is convex and have the minimum

$f = \frac{C}{2}\|w\|_2^2 - w^T b.$

$\frac{\partial f}{\partial w} = \frac{C}{2} 2\cdot w - b = 0.$

$w = \frac{b}{C}$

minimum $= \frac{C}{2}\frac{b^T b}{C^2} - \frac{b^T b}{C} = -\frac{b^T b}{2C}$

(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?

Your answer: $f(\xi) = \lambda\xi + \log(1+\exp\xi)$

$f'(\xi) = \lambda + \frac{\exp\xi}{1+\exp\xi} = 0.$

$f''(\xi) = \frac{e^\xi}{(1+\exp\xi)^2} > 0.$ convex. the minimum exists.

$e^\xi = -\frac{\lambda}{\lambda+1}$ $\quad \frac{\lambda}{\lambda+1} < 0.$ $\lambda \in (-1, 0)$

$\xi = \log -\frac{\lambda}{\lambda+1}$ $\quad$ Minimum $= \lambda\log(-\frac{\lambda}{\lambda+1}) + \log\frac{1}{\lambda+1}$

Name: Yuxuan Zhang (yuxuanz8)

(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.

Your answer:

$$\mathcal{L}(w, \lambda_x, \lambda_z, \xi_x, \xi_z) = \sum_{x\in D}\left(\log(1+\xi_x) + \lambda_x \xi_x\right) + \sum_{z}\left(\log(1+\xi_z) + \lambda_z \xi_z\right) + \frac{C}{2}\|w\|^2 - w^T\left(\sum_z G_\theta(z) + \sum_x \lambda_x x - \sum_z \lambda_z G_\theta(z)\right).$$

Max Dual

$$G(\lambda_x, \lambda_z) = \sum_x\left[\lambda_x \log\left(-\frac{\lambda_x}{1+\lambda_x}\right) + \log\frac{1}{1+\lambda_x}\right] + \sum_z\left[\lambda_z \log\frac{-\lambda_z}{1+\lambda_z} + \log\frac{1}{1+\lambda_z}\right] - \frac{1}{2C}\left\|\sum_z G_\theta(z) + \sum_x \lambda_x x - \sum_z \lambda_z G_\theta(z)\right\|^2$$

$$\lambda_x, \lambda_z \in (-1, 0) \quad \forall x \in D, z \in Z$$

$$\frac{\partial G}{\partial \lambda_x} = 0 \quad \frac{\partial}{\partial \lambda_x} G = \log(-\lambda_x) - \log(1+\lambda_x) + \frac{1}{C}\left(\sum_z G_\theta(z) + \sum_x \lambda_x x - \sum_z \lambda_z G_\theta(z)\right) x^T = 0.$$

$$\frac{\partial G}{\partial \lambda_z} = 0 \quad \frac{\partial G}{\partial \lambda_z} = \log(-\lambda_z) - \log(1+\lambda_z) + \frac{1}{C}\left(\sum_z G_\theta(z) + \sum_x \lambda_x x - \sum_z \lambda_z G_\theta(z)\right)(-G_\theta(z))^T = 0.$$

We can solve λ_x, λ_z as a system of equations
max min problem has changed to max max.
As the function inside is concave and we can solve GAN problem.

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

Your answer:

DLoss : loss = criterion (logit, target1)

GLoss : loss = criterion (logit, target2)

E:0 B:0 DLoss: 0.686484

E:0 B:0 GLoss: 0.694809.

(Run time too slow. Colab doesn't work, buffer issue)