DGM HW3 Written part

ALT_PDG

2025 年 11 月 5 日

Consider a GAN model where G is the generator and D is the discriminator. Let $p_{\text{data}}(x)$ represent the probability distribution of the real data, and $p_g(x)$ represent the distribution generated by the generator G. D and G play the following two-player minimax game with value function $V(G, D)$:

$$\min_G \max_D V(D, G) = \min_G \max_D E_{x\sim p_{\text{data}}(x)} [\log D(x)] + E_{z\sim p_z(z)} [\log (1 - D(G(z)))] \quad (1)$$

For a fixed generator G, the optimal discriminator D is given by:

$$D_G^*(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_g(x)}$$

Problem 4(a)

Provide a detailed derivation for this optimal discriminator formula. You may start by considering the form of the GAN objective function and use calculus to derive the optimality condition for D.

Proof: Indeed

$$E_{x\sim p_{\text{data}}(x)} [\log D(x)] + E_{z\sim p_z(z)} [\log (1 - D(G(z)))] = \int_{R^N} logD(x)p_{data}(x)\,dx + \int_{R^N} log(1 - D(G(z)))p_z(z)\,dz$$

For a fixed G, it sufficies to find

$$\arg\max_D \int_{R^N} logD(x)p_{data}(x)\,dx + \int_{R^d} log(1 - D(G(z)))p_z(z)\,dz$$

Intuitively $\int f(G(z))p_z(z)\,dz$ equals $\int f(y)p_g(y)\,dy$ since fake data y is generated through G(z). But the following derivation is not rigorous since G may not have a inverse

If we set $y = G(z)$, then $z = G^{-1}(y)$ and

$$P(y \in A) = P(z \in G^{-1}(A)) = \int_{G^{-1}(A)} p_z(z)\,dz = \int_A p_z(G^{-1}(t))|det\frac{\partial G^{-1}(t)}{\partial t}|, dt$$

Hence $p_g(y) = p_z(G^{-1}(y))|det\frac{\partial G^{-1}(y)}{\partial y}|$, and

$$\int_{R^d} log(1-D(G(z)))p_z(z)\,dz = \int_{R^N} log(1-D(y))p_z(G^{-1}(y))|det\frac{\partial G^{-1}(y)}{\partial y}|\,dy = \int_{R^N} log(1-D(y))p_g(y)\,dy$$

hence

$$\int_{R^N} logD(x)p_{data}(x)\,dx + \int_{R^d} log(1 - D(G(z)))p_z(z)\,dz = \int_{R^N} logD(x)p_{data}(x)\,dx + \int_{R^N} log(1 - D(x))p_g(x)\,dx$$

To obtain the optimal D, we consider the critical point of function $f(a) = uloga + vlog(1-a)$, a direct calculation yiels

$$f'(a) = \frac{u}{a} + \frac{v}{a-1}$$

and f has attains it's maximum when $\frac{u}{a} = \frac{v}{1-a} \Rightarrow u - ua = va \Rightarrow a = \frac{u}{u+v}$, let $a = D(x), u = p_{data}, v = p_g$, then the optimal D would be

$$D_G^* = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}$$

Problem 4(b)

The global minimum of the virtual training criterion $C(G) = \max_D V(D, G)$ is achieved if and only if $p_g = p_{\text{data}}$. At this point, $C(G)$ achieves the value $-\log 4$.
Provide a detailed derivation of this result. Show that the global minimum of the criterion occurs when $p_g = p_{data}$, and calculate the corresponding value of C(G).

Proof:From the previous problem we know

$$V(D, G) = \int_{R^N} logD(x)p_{data}(x)\,dx + \int_{R^N} log(1 - D(x))p_g(x)\,dx$$

hence

$$V(D_G^*, G) = \int_{R^N} log\frac{p_{data}}{p_g(x) + p_{data}(x)}p_{data}(x)\,dx + \int_{R^N} log\frac{p_g(x)}{p_g(x) + p_{data}(x)}p_g(x)\,dx$$

$$= \int_{R^N} log\frac{p_{data}}{p_g(x) + p_{data}(x)}p_{data}(x) + log\frac{p_g(x)}{p_g(x) + p_{data}(x)}p_g(x)\,dx$$

$$= \int_{R^N} log\frac{2p_{data}}{p_g(x) + p_{data}(x)}p_{data}(x) + log\frac{2p_g(x)}{p_g(x) + p_{data}(x)}p_g(x)\,dx - log2\int_{R^N} p_{data}(x)\,dx - log2\int_{R^N} p_g(x)\,dx$$

$$= \int_{R^N} log\frac{2p_{data}}{p_g(x) + p_{data}(x)}p_{data}(x) + log\frac{2p_g(x)}{p_g(x) + p_{data}(x)}p_g(x)\,dx - log4$$

To find the disired $p_g(x)$, we consider the function

$$f(x) = ylog\frac{2y}{x+y} + xlog\frac{2x}{x+y} = ylog2y - ylog(x+y) + xlog2x - xlog(x+y)$$

$$f'(x) = -\frac{y}{x+y} + 1 + log2x - log(x+y) - \frac{x}{x+y} = log2x - log(x+y) = log\frac{2x}{x+y} \qquad x, y > 0$$

Hence f attains it's minimum when $x = y$, and $f(x) \geq f(y) = 0$, hence

$$V(D_G^*, G) \geq 0 - log4 = -log4$$

The equality holds if and only if $log\frac{2p_{data}}{p_g(x)+p_{data}(x)}p_{data}(x) + log\frac{2p_g(x)}{p_g(x)+p_{data}(x)}p_g(x) = 0$ for any x, if and only if $p_g = p_{data}$

Problem 5

Consider two finite sets $X = \{1, 2, \ldots, k_1\}$ and $Y = \{1, 2, \ldots, k_2\}$, and two discrete distributions $v \in R^{k_1}$ and $w \in R^{k_2}$, such that:

$$\sum_{i=1}^{k_1} v_i = \sum_{j=1}^{k_2} w_j = 1.$$

Let $C = [c_{ij}]$ be a cost matrix, where c_{ij} represents the cost of transporting mass from element $i \in X$ to element $j \in Y$.

The Optimal Transport (OT) problem can be written as the following linear program:

$$\mathrm{OT}(v, w; C) = \min_{T \in R^{k_1 \times k_2}} \sum_{i,j} T_{ij} c_{ij}$$

$$\text{s.t.} \quad T_{ij} \geq 0, \quad \sum_j T_{ij} = v_i \quad \forall i \in \{1, \ldots, k_1\},$$

$$\sum_i T_{ij} = w_j \quad \forall j \in \{1, \ldots, k_2\}.$$

The Kantorovich-Rubinstein duality theorem provides a way to express the Optimal Transport problem in its dual form, transforming it into a simpler optimization problem over dual variables.

$$\mathrm{OT}(v, w; C) = \max_{\phi, \psi} \sum_{i=1}^{k_1} v_i \phi_i + \sum_{j=1}^{k_2} w_j \psi_j$$

$$\text{s.t.} \quad \phi_i + \psi_j \leq c_{ij}, \quad \forall i \in \{1, \ldots, k_1\},\, j \in \{1, \ldots, k_2\}.$$

Here, $\phi \in R^{k_1}$ and $\psi \in R^{k_2}$ are dual potential functions, and the dual problem seeks to maximize their sum subject to the coupling constraint.

Derive the dual form of the discrete Optimal Transport problem above. To help you with the derivation, recall that the dual of a convex optimization problem can often be derived by introducing Lagrange multipliers for the equality constraints in the primal problem and maximizing the resulting Lagrangian with respect to the dual variables.

Proof: Assume that $g_{ij}(x) = -x_{ij}$, then the inequality constrains can be rewritten as

$$g_{ij}(T) \leq 0$$

Let

$$h_i(T) = \sum_{j=1}^{k_2} T_{ij} - v_i \qquad h_{k_1+j}(T) = \sum_{i=1}^{k_1} T_{ij} - w_j$$

Then by KKT condition, there the optimal transport problem is equivalent with

$$\max_{\lambda,\mu,T} \mathcal{L}(\lambda, \mu, T) = \max_{\lambda,\mu,T} \sum_{i,j} T_{ij} c_{ij} + \sum_{i,j=1}^{k_1,k_2} \mu_{ij} g_{ij}(T) + \sum_i^{k_1} \lambda_i h_i(T) + \sum_j^{k_2} \lambda_{k_1+j} h_j(T)$$

For simplicity, we can assume $(\lambda_{k_1+j}) = z_j$, then we can express $\mathcal{L}(\lambda, \mu, T) = \mathcal{L}(\lambda, z, \mu, T)$, and

$$\max_{\lambda,\mu,T} \mathcal{L}(\lambda, \mu, T) = \sum_{i,j} T_{ij} c_{ij} - \sum_{i,j} \mu_{ij} T_{ij} + \sum_{i,j} \lambda_i T_{ij} - \sum_i \lambda_i v_i + \sum_{i,j} z_j T_{ij} - \sum_j z_j w_j$$

$$\frac{\partial \mathcal{L}}{\partial T_{ij}} = c_{ij} - \mu_{ij} + \lambda_i + z_j$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = \sum_j T_{ij} - v_i = h_i(T) = 0$$

$$\frac{\partial \mathcal{L}}{\partial z_j} = h_{k_1+j}(T) = 0$$

$$\frac{\partial \mathcal{L}}{\partial \mu_{ij}} = -T_{ij}$$

By the necessity condition, we know $\frac{\partial \mathcal{L}}{\partial T_{ij}}$ must =0 for any i,j. Hence

$$\lambda_i + z_j + c_{ij} = \mu_{ij} \qquad (1)$$

($\mu_{ij} \geq 0$ is the necessity condition of KKT)
And by (1) we have

$$\max_{\lambda,\mu,T} \mathcal{L}(\lambda, z, \mu, T) = \max_{\lambda,\mu,T} \sum_{i,j} T_{ij}(c_{ij} - \mu_{ij} + \lambda_i + z_j) - \sum_i \lambda_i v_i - \sum_j z_j w_j = \max_{\lambda,\mu,T} - \sum_i \lambda_i v_i - \sum_j z_j w_j$$

Notice that the last term is irrelevant with T, hence

$$\max_{\lambda,\mu,T} \mathcal{L}(\lambda, z, \mu, T) = \max_{\lambda,\mu} - \sum_i \lambda_i v_i - \sum_j z_j w_j$$

let $\lambda_i = -\phi_i, z_j = -\psi_j$, then the optimization problem is equivalent with

$$\max_{\phi,\psi} \sum_i \phi_i v_i + \sum_j^{k_2} \psi_j w_j$$

Don't forget (1), since the KKT condition requires $\mu_{ij} \geq 0$, we have

$$\phi_i + \psi_j \leq c_{ij}$$

So we get the desired result :

$$OT = \max_{\phi,\psi} \sum_i \phi_i v_i + \sum_j^{k_2} \psi_j w_j \qquad \phi_i + \psi_j \leq c_{ij}$$