Why is optimizing G in a situation where D is optimized the same as minimizing Jensen-Shannon Divergence?

Author: Sehwan Park

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

$E_{z\sim p_z(z)}[\log(1-D(G(z)))] = E_{y\sim p_g(y)}[\log(1-D(y))]$

$V(G,D) = E_{x\sim p_{data}(x)}[\log(D(x))] + E_{y\sim p_g(y)}[\log(1-D(y))]$

$= \int_x P_{data}(x)\log D(x)dx + \int_x P_g(x)\log(1-D(x))dx$

$= \int_x \left[P_{data}(x)\log D(x) + P_g(x)\log(1-D(x))\right]dx$

$Let, a = P_{data}(x)\ and\ b = P_g(x)$

$f(x) = a\log(x) + b\log(1-x)$

$f(x)\ is\ a\ upper\ convex\ function\ so\ differentiate\ f(x)\ to\ find\ optimal\ D$

$\frac{f(x)}{dx} = \frac{a}{x} - \frac{b}{1-x} = 0$

$a(x-1) + bx = 0$

$(b+a)x - a = 0$

$x = \frac{a}{a+b}$

$optimal D^* = \frac{P_{data}(x)}{P_{data}(x)+P_G(x)}$

$Let's\ optimize\ G\ when\ D\ is\ optimization$

$\max_D V(G,D) = V(G,D^*)$

$= \int_x \left[P_{data}(x)\log D^*(x) + P_G(x)\log(1-D^*(x))\right]dx$

$= E_{x\sim p_{data}(x)}[\log(D^*(x))] + E_{x\sim p_g(x)}[\log(1-D^*(x))]$

$= E_{x\sim p_{data}(x)}[\log\frac{p_{data}(x)}{p_{data}(x)+p_g(x)}] + E_{x\sim p_g(x)}[\log\frac{p_g(x)}{p_{data}(x)+p_g(x)}]$

$= E_{x\sim p_{data}(x)}[\log(p_{data}(x)) - \log(\frac{p_{data}(x)+p_g(x)}{2})] + E_{x\sim p_g(x)}[\log(p_g(x)) - \log(\frac{p_{data}(x)+p_g(x)}{2})]$

$-\log(4)$

$= KL(p_{data}||\frac{p_{data}+p_g}{2}) + KL(p_g||\frac{p_{data}+p_g}{2}) - \log(4)$

$= 2\times JSD(p_{data}||p_g) - \log(4)$

$KL(P \parallel Q) = \sum_x P(x) log \frac{P(x)}{Q(x)}$

$JSD(P \| Q) = \frac{1}{2} KL(P \| M) + \frac{1}{2} KL(Q \| M)$

$M = \frac{1}{2}(P + Q)$

$So,\ Optimization\, G\ is\ equal\ with\ minimizing\ Jenson{-}Shannon\, Divergence.$