MathDNN HW 12

2018-13260 Jaehyeon Cha

1 Problem 2

For simplicity, denote p_{true} as p, and p_θ as q. Also, if $h = h(x)$, denote shortcut-notation

$$\int_{x\in\mathbb{R}} h(x)dx = \int h.$$

First of all, to remove the ϕ-term, let us determine whenever D_ϕ obtains its maximum. Focus on the expression of loss function

$$\mathbb{E}_{X\sim p}[\log D_\phi(X)] + \lambda\mathbb{E}_{\bar{X}\sim q}[\log(1 - D_\phi(\bar{X})] \tag{1}$$

$$= \int \left(p\log D_\phi + \lambda q \log(1 - D_\phi)\right). \tag{2}$$

Differentiating w.r.t. D_ϕ, (1) has its minimum at

$$D_\phi = \frac{p}{p+\lambda q},$$

which is possible since D_ϕ can be approximated to any function. Thus our loss function (1) becomes

$$\int \left(p\log\frac{p}{p+\lambda q} + \lambda q\log\frac{\lambda q}{p+\lambda q}\right). \tag{3}$$

Now, take given f and calculate divergence. We have

$$\begin{aligned}
D_f(p||q) &= \int q\left(\frac{p}{q}\log\frac{p}{p+\lambda q} + \lambda\log\frac{\lambda q}{p+\lambda q} + (1+\lambda)\log(1+\lambda) - \lambda\log\lambda\right)\\
&= \int \left(p\log\frac{p}{p+\lambda q} + \lambda q\log\frac{\lambda q}{p+\lambda q}\right) + \left((1+\lambda)\log(1+\lambda) - \lambda\log\lambda\right)\int q\\
&= \int \left(p\log\frac{p}{p+\lambda q} + \lambda q\log\frac{\lambda q}{p+\lambda q}\right) + (1+\lambda)\log(1+\lambda) - \lambda\log\lambda.
\end{aligned}$$

Since λ is a constant, minimizing $D_f(p||q)$ is equivalent to minimizing (3).