模式识别与机器学习 **30250293-0 (2022 春)**

第 15 章 (深度生成式模型) 作业

1 极大似然与 KL 散度

1.1 证明极大化“样本集在模型分布上的似然”与极小化“数据分布与模型分布的 KL 散度”的等价性

$$
\begin{aligned}
\min_{\boldsymbol{\theta}} D_{\mathrm{KL}}(p_{\mathrm{data}}||p_{\boldsymbol{\theta}}) &\Leftrightarrow \min_{\boldsymbol{\theta}} \int_{\mathcal{X}} p_{\mathrm{data}}(x) \log\left(\frac{p_{\mathrm{data}}(x)}{p_{\boldsymbol{\theta}}(x)}\right) \mathrm{d}x \\
&\Leftrightarrow \min_{\boldsymbol{\theta}} \int_{\mathcal{X}} p_{\mathrm{data}}(x) \log(p_{\mathrm{data}}(x)) - p_{\mathrm{data}}(x) \log(p_{\boldsymbol{\theta}}(x)) \mathrm{d}x \\
&\Leftrightarrow \min_{\boldsymbol{\theta}} \int_{\mathcal{X}} -p_{\mathrm{data}}(x) \log(p_{\boldsymbol{\theta}}(x)) \mathrm{d}x \\
&\Leftrightarrow \max_{\boldsymbol{\theta}} \int_{\mathcal{X}} p_{\mathrm{data}}(x) \log(p_{\boldsymbol{\theta}}(x)) \mathrm{d}x \\
&\Leftrightarrow \max_{\boldsymbol{\theta}} \mathbb{E}_{x\sim p_{\mathrm{data}}(x)} \left[\log(p_{\boldsymbol{\theta}}(x))\right] \mathrm{d}x
\end{aligned}
$$

1.2 对比说明优化目标选择的合理性

p_{data} 是未知的分布选择 MSE 需要知道样本点 x 上 $p_{\mathrm{data}}(x)$ 的取值才能计算，而这是不可知的。而最小化 KL 散度可以在等价转换可以进行蒙特卡洛近似，消除积分式内的 p_{data} 项，可以在只知道样本点信息不知道样本点上的样本点 x 上 $p_{\mathrm{data}}(x)$ 就能计算