NNDL Exercise set 5 – Mathematical exercises

1.

Define the set $U \subset \mathbb{R}$ as $U = [0, 2\pi]$. Then the function f can be defined as $f(x) = (\cos x, \sin x)$. The image of the function $f : U \to \mathbb{R}^2$ is then $f(U) = \{(x, y) \in \mathbb{R}^2 |\ x^2 + y^2 = 1\}$, i.e. the unit sphere. Since $U \subset \mathbb{R}$ the dimension of the manifold is 1.

2.

The orthogonal group $O(n)$ is formed by $n \times n$ matrices Q for which $Q^T Q = QQ^T = I$. Define the function $g(Q) = Q^T Q - I$. The manifold is then $\{Q \in \mathbb{R}^{n\times n} | g(Q) = 0\}$.

3.

We have

$$p(\mathbf{x}; \mathbf{M}) = \frac{1}{\mathbf{Z}} \exp\left(-\frac{1}{2}\mathbf{x}^T \mathbf{M}\mathbf{x}\right), \tag{1}$$

where $\mathbf{M} = \mathbf{\Sigma}^{-1}$ is the precision matrix. The score function is

$$\begin{aligned} \phi(\mathbf{x}; \mathbf{M}) &= \nabla_{\mathbf{x}} \log p(\mathbf{x}; \mathbf{M}) \\ &= -\nabla_{\mathbf{x}} \frac{1}{2}\mathbf{x}^T \mathbf{M}\mathbf{x} \\ &= -\frac{1}{2}(\mathbf{M} + \mathbf{M}^T)\mathbf{x} \\ &= -\mathbf{M}\mathbf{x}, \end{aligned} \tag{2}$$

where we used the fact that $\mathbf{M}$ is symmetric. Then

$$\phi(\mathbf{x}; \mathbf{M})^2 = \mathbf{x}^T \mathbf{M}^T \mathbf{M}\mathbf{x} = \mathbf{x}^T \mathbf{M}\mathbf{M}\mathbf{x}. \tag{3}$$

The score matching objective for N observations is

$$\tilde{J}(\mathbf{M}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{n}\left[\partial_j \phi_j(\mathbf{x}_i; \mathbf{M}) + \frac{1}{2}\phi_j(\mathbf{x}_i; \mathbf{M})^2\right] \tag{4}$$

By noting that

$$\sum_{j=1}^{n} \partial_j \phi_j(\mathbf{x}; \mathbf{M}) = \sum_{j=1}^{n} -M_{jj} = -\mathrm{Tr}(\mathbf{M}), \tag{5}$$

we find

$$\tilde{J}(\mathbf{M}) = \frac{1}{N} \sum_{i=1}^{N} \left[-\mathrm{Tr}(\mathbf{M}) + \frac{1}{2} \mathbf{x}_i^T \mathbf{M} \mathbf{M} \mathbf{x}_i \right]. \tag{6}$$

4.

We have

$$\log p_{\mathrm{un}}(x; \theta) = f(x) g(\theta) \tag{7}$$

for some smooth functions $f, g : \mathbb{R} \to \mathbb{R}$ and $x, \theta \in \mathbb{R}$.

a)

The score function is

$$\phi(x; \theta) = \frac{\partial}{\partial x} \log p_{\mathrm{un}}(x; \theta) = f'(x) g(\theta). \tag{8}$$

b)

Now

$$\frac{\partial}{\partial x} \phi(x; \theta) = f''(x) g(\theta), \tag{9}$$

so the score matching objective for N observations is

$$\begin{aligned} \tilde{J}(\theta) &= \frac{1}{N} \sum_{i=1}^{N} \left[\frac{\partial}{\partial x} \phi(x_i; \theta) + \frac{1}{2} \phi(x_i; \theta)^2 \right] \\ &= \frac{1}{N} \sum_{i=1}^{N} \left[f''(x_i) g(\theta) + \frac{1}{2} f'(x_i)^2 g(\theta)^2 \right]. \end{aligned} \tag{10}$$

c)

We can find $\hat{\theta}$ by minimizing $\tilde{J}(\theta)$ w.r.t. $g(\theta)$.

$$
\begin{aligned}
&\frac{\partial}{\partial g(\theta)}\tilde{J}(\theta) = 0 \\
\implies &\frac{1}{N}\sum_{i=1}^{N}\left[f''(x_i) + f'(x_i)^2 g(\theta)\right] = 0 \\
\implies &g(\hat{\theta}) = -\frac{\sum_{i=1}^{N} f''(x_i)}{\sum_{i=1}^{N} f'(x_i)^2} \\
\implies &\hat{\theta} = g^{-1}\left(-\frac{\sum_{i=1}^{N} f''(x_i)}{\sum_{i=1}^{N} f'(x_i)^2}\right)
\end{aligned}
\tag{11}
$$

Not sure if the exercise asking for maximization instead of minimization is intended, but in case it is, we can define a new score matching objective as $\tilde{J}(\theta) \equiv -\tilde{J}(\theta)$ and proceed by minimizing it, in which case the result is the same.

5.

If the datasets have equal distributions and equal number of data points approaching infinity, then the optimal regression function should be a constant

$$
\sigma(0) = \frac{1}{2}, \tag{12}
$$

since the two classes are indistinguishable.

6.

The proposal will not work because uniform noise in $\mathbb{R}^n$ is not normalizable. This would cause the objective of the logistic regression to be ill-defined and thus causing learning to fail.