Homework 6

Haopeng Zhang - 2023010770

2025 6 - 7

Lecture on BFNs and AI4Science

Question 1:

Suppose our goal is to estimate the unknown y given some observations x. Prove that the optimal solution $\hat{y}$ to the MSE loss $L(\hat{y}) = \mathbb{E}_{y \sim p(y|x)}[(y - \hat{y})^2]$ is the posterior expectation $\hat{y}^* = \mathbb{E}[y|x]$, and explain the role of $p(y|x)$ in the context of Bayesian inference.

Answer:

$$
\begin{aligned}
\hat{y}^* &= \arg\min_{\hat{y}} \ \mathbb{E}_{y \sim p(y|x)}[(y - \hat{y})^2] \\
&= \arg\min_{\hat{y}} \ \mathbb{E}_{y \sim p(y|x)}[y^2 - 2y\hat{y} + \hat{y}^2] \\
&= \arg\min_{\hat{y}} \ \mathbb{E}_{y \sim p(y|x)}[y^2] - 2\hat{y} \ \mathbb{E}_{y \sim p(y|x)}[y] + \hat{y}^2
\end{aligned}
$$

This quadratic function in $\hat{y}$ reaches its minimum at the vertex of the parabola, which gives $\hat{y}^* = \mathbb{E}_{y \sim p(y|x)}[y] = \mathbb{E}[y|x]$. As for the role of $p(y|x)$ in the context of Bayesian inference, it represents how we should update our belief about y after observing x as evidence, which is essential in predicting and reasoning under uncertainty. In Bayesian Flow Networks, a continuous process over time, which transforms a prior distribution into an approximation of the posterior, is applied to learn a conditional distribution $p_\theta(y|x)$ that matches the true distribution.