Week 9 – Tutorial

Section A: Lecture 9 Review

Review Lecture 9 and provide answers to the following questions in Week9.md in your Github Repository:

1. What is the main goal of Variational Inference and explain why Variational Inference is often preferred over Markov Chain Monte Carlo (MCMC) methods.
2. What is the variational distribution $q(w)$ in Variational Inference, and why is it used?
3. What is KL divergence, and how is it used in Variational Inference?
4. What are the two main components of the ELBO in Variational Inference?

Section B: Multiple Choice Questions

1. Which of the following best describes Bayes' Theorem?

(A) It calculates the likelihood of an event.
(B) It updates the probability of a hypothesis as more evidence becomes available.
(C) It measures the mean of a distribution.
(D) It calculates the variance of a data set.

2. Bayes' Theorem can be expressed as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

What does $P(A\backslash B)$ represent?

(A) Prior probability
(B) Likelihood
(C) Posterior probability
(D) Marginal likelihood

3. In the context of Bayes' Inference in Machine Learning models, what is the prior probability?

(A) The likelihood of the data given the parameters.
(B) The probability distribution over the parameters before any data is observed.
(C) The evidence or marginal likelihood.
(D) The probability of observing new data given the posterior.

4. Bayes' Theorem is used in machine learning to:

(A) Compute posterior distributions given observed data.
(B) Maximize the likelihood of a model.
(C) Minimize the variance of model parameters.
(D) Calculate the exact probability of future events.

5. What does the marginal likelihood in Bayes' Theorem represent?

(A) The likelihood of data given the parameters.
(B) The total probability of observing the data, summed over all possible values of the parameters.
(C) The posterior probability distribution.
(D) The likelihood of the parameters given the data.

6. Which of the following is not a component of Bayes' Theorem?

(A) Posterior probability
(B) Prior probability
(C) Maximum likelihood estimation
(D) Evidence (Marginal likelihood)

7. What is the key difference between a Bayesian Neural Network (BNN) and a traditional neural network?

(A) BNN uses a different activation function.
(B) BNN assigns distributions over weights instead of fixed values.
(C) BNN can only be used for regression tasks.
(D) BNN does not use backpropagation.

8. In Bayesian Neural Networks, what does the posterior distribution represent?

(A) The likelihood of the data given the model weights.
(B) The prior belief about the model weights.
(C) The distribution of the input data.
(D) The updated belief about the model parameters after observing data.

9. Which of the following is used to approximate the posterior distribution in Bayesian Neural Networks?

(A) Backpropagation
(B) Variational Inference
(C) Maximum Likelihood Estimation
(D) Random Sampling

10. In a Bayesian Neural Network, what does it mean when a weight has a narrow posterior distribution?

(A) High uncertainty about the weight's value.
(B) The weight is irrelevant to the model.
(C) The weight has no effect on the network's output.
(D) Low uncertainty about the weight's value.

11. The prior distribution in Bayesian Neural Networks is:

(A) The distribution of the model's predictions.
(B) The probability distribution over the model parameters before observing data.
(C) The likelihood of the model given the parameters.
(D) The posterior distribution after observing data.

12. What is the main reason for using Bayesian Neural Networks?

(A) To increase the speed of training.
(B) To introduce uncertainty in model predictions.
(C) To reduce overfitting by using more parameters.
(D) To make the model fully deterministic.

13. Which of the following statements is true about the relationship between the prior and posterior distributions in Bayesian inference?

(A) The posterior is always narrower than the prior.
(B) The posterior is independent of the prior.
(C) The posterior combines prior beliefs with observed data to update the distribution of the parameters.
(D) The prior has no effect on the posterior distribution.

14. What does the evidence or marginal likelihood in Bayes' Theorem represent?

(A) The probability of the data given a specific model.
(B) The prior probability of the data.
(C) The total probability of the observed data, integrating over all possible parameter values.
(D) The likelihood of the model parameters.

15. In Bayesian inference, the term P(D|w) refers to:

(A) The prior distribution of the data.
(B) The probability of observing the data given the model parameters w.
(C) The posterior probability of the parameters given the data.
(D) The marginal likelihood of the data.

16. In Bayesian inference, what role does the likelihood function play?

(A) It describes the probability of the observed data given the model parameters.
(B) It serves as the prior distribution for the model parameters.
(C) It is used to approximate the posterior distribution.
(D) It measures the uncertainty of the model parameters.

17. How does a Bayesian Neural Network (BNN) quantify uncertainty in predictions?

(A) By using dropout regularization during training.
(B) By maintaining distributions over weights rather than point estimates.
(C) By increasing the number of layers in the network.
(D) By maximizing the likelihood of the data.

18. Which of the following best describes epistemic uncertainty in a Bayesian Neural Network?

(A) Uncertainty due to noise inherent in the data.
(B) Uncertainty that arises due to a lack of knowledge in the model.
(C) Uncertainty in the initialization of network weights.
(D) Uncertainty caused by errors in data labeling.

19. A Bayesian Neural Network provides a posterior distribution over the model weights. This distribution allows the model to:

(A) Minimize the loss function faster.
(B) Quantify both the uncertainty in the model and the uncertainty in its predictions.
(C) Remove overfitting entirely.
(D) Make deterministic predictions.

20. Which type of uncertainty can Bayesian Neural Networks help mitigate by modeling it explicitly?

(A) Aleatoric uncertainty (inherent data noise)
(B) Epistemic uncertainty (uncertainty due to lack of knowledge about the model)
(C) Structural uncertainty (uncertainty in model architecture)
(D) All types of uncertainty

21. In the context of predictions made by a Bayesian Neural Network, aleatoric uncertainty refers to:

(A) Uncertainty inherent in the observations or data itself.
(B) Uncertainty due to lack of data or information about the model.
(C) Uncertainty caused by varying learning rates.
(D) Uncertainty in the activation functions used in the network.

22. When a Bayesian Neural Network predicts a distribution rather than a single point value, it allows for:

(A) Estimation of overfitting tendencies in the network.
(B) Evaluation of how certain the network is about its predictions.
(C) Minimization of the test set error rate.
(D) Reducing model complexity.

23. What is the objective of Variational Inference in probabilistic models?

(A) To approximate the posterior distribution by maximizing the evidence lower bound (ELBO) and minimizing KL divergence.
(B) To minimize the model loss function.
(C) To calculate the exact posterior distribution.
(D) To approximate the prior distribution by minimizing the evidence lower bound (ELBO) and maximizing KL divergence.

24. Which of the following measures the difference between the approximate posterior and the true posterior in Variational Inference?

(A) Evidence Lower Bound (ELBO)
(B) Maximum Likelihood Estimation
(C) KL Divergence
(D) Log Likelihood

25. In Variational Inference, what is the role of the ELBO?

(A) It provides an upper bound on the likelihood of the model.
(B) It approximates the log marginal likelihood by optimizing a variational distribution.
(C) It computes the exact probability of the model's parameters.
(D) It minimizes the model's complexity.

26. Why is Variational Inference preferred over other methods, such as MCMC, in certain scenarios?

(A) It is an exact method for posterior computation.
(B) It does not require the use of gradient-based optimization.
(C) It always provides more accurate results.
(D) It converges faster and is more computationally efficient.

27. What does minimizing KL divergence in Variational Inference accomplish?

(A) It minimizes the prior distribution.
(B) It makes the variational approximation closer to the true posterior.
(C) It maximizes the evidence lower bound (ELBO).
(D) It increases the likelihood of observing the data.

28. Which of the following techniques is commonly used for optimizing the ELBO in Variational Inference?

(A) Backpropagation
(B) Feedforwarding
(C) Stochastic Gradient Descent
(D) Maximum Likelihood Estimation

29. A manufacturing plant produces 80% high-quality parts and 20% defective parts. If a quality control test is performed, it correctly identifies high-quality parts 90% of the time and incorrectly labels defective parts as high quality 15% of the time. What is the probability that a randomly selected part is high quality, given that it passed the quality control test?

(A) 0.871
(B) 0.352
(C) 0.960
(D) 0.765