Theory Questions:

1. What is a Support Vector Machine (SVM)?
SVM is a supervised learning algorithm used for classification and regression tasks. It works by
finding the optimal hyperplane that maximizes the margin between different classes in a dataset.
SVM is effective for high-dimensional spaces and is robust against overfitting.

2. What is the difference between Hard Margin and Soft Margin SVM?
Hard Margin SVM: Uses a strict boundary where no data points are allowed inside the margin. It works well with linearly separable data but is sensitive to noise.
Soft Margin SVM: Allows some misclassification by introducing a slack variable, making it more flexible for real-world noisy datasets.

3. What is the mathematical intuition behind SVM?
SVM optimizes the function:
Min ½ ||w||^2
subject to
yi(w·xi+b)≥1

4. What is the role of Lagrange Multipliers in SVM? Lagrange multipliers transform constrained optimization into an unconstrained problem. In
SVM, they help handle the constraints efficiently using the Lagrange function, leading to the
dual formulation for computational efficiency.

5. What are Support Vectors in SVM?
Support vectors are data points that lie closest to the hyperplane and influence its position. They
define the margin and are crucial in determining the decision boundary.

6. What is a Support Vector Classifier (SVC)?
SVC is a classification algorithm based on SVM that finds the best hyperplane to separate
different classes. It extends to soft margins for handling non-linearly separable data.

7. What is a Support Vector Regressor (SVR)?
SVR applies the principles of SVM to regression problems. Instead of maximizing the margin, it

tries to fit data within a margin of tolerance while minimizing the error.

8. What is the Kernel Trick in SVM?
The Kernel Trick allows SVM to operate in higher-dimensional space without explicitly transforming the data. It enables SVM to classify non-linearly separable data by using functions
like the polynomial or RBF kernel.

9. Compare Linear Kernel, Polynomial Kernel, and RBF Kernel.
Linear Kernel: Best for linearly separable data.
Polynomial Kernel: Maps input data to higher dimensions using a polynomial transformation.
RBF Kernel: Uses a Gaussian function to create flexible decision boundaries, effective for complex data distributions.

10. What is the effect of the C parameter in SVM?
C controls the trade-off between maximizing margin and minimizing classification error:
High C: Less margin, fewer misclassifications, but can overfit.
Low C: More margin, higher misclassification tolerance, prevents overfitting.

11. What is the role of the Gamma parameter in RBF Kernel SVM?
Gamma defines how far the influence of a training point extends:
High Gamma: Points have a narrow influence, leading to overfitting.
Low Gamma: Points have a broader influence, leading to underfitting.

12. What is the Naïve Bayes classifier, and why is it called "Naïve"?
Naïve Bayes is a probabilistic classifier based on Bayes' theorem. It assumes that features are
independent given the class, which is often unrealistic, hence the term "naïve."

13. What is Bayes' Theorem?
Bayes' theorem is:
$P(A|B)= P(B|A)P(A) / P(B)$
where:
$P(A|B)P(A|B)P(A|B)$ is the posterior probability,
$P(B|A)P(B|A)P(B|A)$ is the likelihood,
$P(A)P(A)P(A)$ is the prior probability, and
$P(B)P(B)P(B)$ is the evidence.

14. Explain the differences between Gaussian Naïve Bayes, Multinomial Naïve Bayes, and
Bernoulli Naïve Bayes.

Gaussian Naïve Bayes: Assumes features follow a normal distribution, suitable for continuous data.
Multinomial Naïve Bayes: Used for text classification, where features represent word counts.
Bernoulli Naïve Bayes: Deals with binary feature data (presence/absence of words in text).

15. When should you use Gaussian Naïve Bayes over other variants?

Gaussian Naïve Bayes is best for datasets where features follow a normal distribution, such as
continuous numerical datasets like weather prediction or medical diagnosis.

16. What are the key assumptions made by Naïve Bayes?

1. **Feature independence**: Each feature contributes independently to the probability.
2. **Equal importance of features**: All features are equally relevant for classification.
3. **Conditional probability is based on past occurrences**.

17. What are the advantages and disadvantages of Naïve Bayes?

Advantages:
Fast and efficient for large datasets.
Works well with small datasets and noisy data.
Performs well in text classification tasks.**Disadvantages**:
Strong independence assumption may not hold.
Cannot handle correlated features well.
Struggles with very complex relationships between variables.

18. Why is Naïve Bayes a good choice for text classification?

Works well with high-dimensional data (like text).
Handles missing data effectively.
Performs well even with a small dataset.
Computationally efficient.

19. Compare SVM and Naïve Bayes for classification tasks.

Feature
SVM
Naïve Bayes
Type

Discriminative
Generative
Performance
Good with fewer features
Good for high-dimensional data
Assumptions
No independence assumption Assumes feature independence
Training Speed
Slower for large datasets
Fast
Text Classification Works but requires tuning
Excellent

20. How does Laplace Smoothing help in Naïve Bayes?
Laplace Smoothing prevents zero probabilities for unseen words or features by adding a small
constant α\alphaα to the numerator and adjusting the denominator accordingly:
P(word|class)= count(word)+α / total_words+α×unique_words
This ensures robustness in text classification where some words may not appear in training data.