**Question 2**

(a) **Maximum Likelihood Estimation (MLE)** is a method used for estimating the parameters of a statistical model. It aims to find the values of parameters that maximize the likelihood function. **Maximum A Posteriori (MAP)** estimation is similar to MLE but involves incorporating prior probabilities to find the maximum posterior probability.

(b) **MLE Application:** In the case of a Binomial distribution (e.g., coin toss), estimating the success probability (parameter *p*) involves using MLE, where the probability of observing a certain number of successes in *n* trials is maximized.

(c) **MAP Application:** Similarly, in MAP estimation, for the same example of a Binomial distribution, prior information about the success probability could be incorporated, influencing the estimation of the success probability based on both prior and observed data.

(d) **Impact of Prior Information:** MLE assumes no prior information, while MAP incorporates prior knowledge. The presence of prior information in MAP affects the estimation by combining prior beliefs with the observed data.

**Question 3**

(a) **Decision Tree Concept:** A decision tree is a tree-like model used for making decisions based on a set of conditions derived from features in the dataset.

(b) **Growing a Decision Tree:** The process involves selecting attributes based on measures like information gain, which helps decide the best splitting attribute at each node.

(c) **Overfitting in Decision Trees:** Overfitting occurs when a tree is too complex and fits the training data too closely, losing its ability to generalize. Pruning, setting minimum samples per leaf, or using validation techniques can help mitigate overfitting.

(d) **Strengths and Weaknesses:** Decision trees are interpretable, simple to understand, and handle both numerical and categorical data well. However, they can be prone to overfitting. For example, in the banking sector, decision trees can be used for loan approval, where transparent decision-making is essential.

**Question 4**

(a) **Logistic Regression Concept:** Logistic regression is a statistical model used for binary classification. It predicts the probability of a sample belonging to a particular class.

(b) **Sigmoid Function:** The sigmoid function transforms the linear combination of input features into a probability between 0 and 1. It compresses the output into a range suitable for probability estimation.

(c) **Limitations:** Logistic regression is limited to binary classification and can't handle non-linear relationships well. For multiclass problems, techniques like one-vs-rest or multinomial logistic regression can be used to extend its application.

**Question 5**

(a) **Pattern Recognition:** Pattern recognition involves identifying regularities in data through the use of algorithms. For instance, facial recognition in images is a form of pattern recognition.

(b) **Considerations in Pattern Recognition:** Key considerations include selecting appropriate features, choosing the right algorithm, handling noisy data, and addressing scalability and computational efficiency. Motivations for using pattern recognition include automation, prediction, and decision-making in various fields such as healthcare, finance, and security.