1. **What is the definition of a target function? In the sense of a real-life example, express the target function. How is a target function's fitness assessed?**

A target function, in machine learning, is a method for solving a problem that an AI algorithm parses its training data to find. Once an algorithm finds its target function, that function can be used to predict results (predictive analysis).

1. **What are predictive models, and how do they work?**

Predictive modelling is a statistical technique using machine learning and data mining to predict and forecast likely future outcomes with the aid of historical and existing data. It works by analyzing current and historical data and projecting what it learns on a model generated to forecast likely outcomes

**What are descriptive types, and how do you use them? Examples of both types of models should be provided. Distinguish between these two forms of models.**

Descriptive analysis can be categorized into four types which are measures of frequency, central tendency, dispersion or variation, and position. These methods are optimal for a single variable at a time

1. **Describe the method of assessing a classification model's efficiency in detail. Describe the various measurement parameters.**

We’ve :

Confusion Matrix:. Terms used in defining a confusion matrix are TP, TN, FP, and FN.

* **Accuracy**: Accuracy = (TP + TN) / (TP + FP +TN + FN)

This term tells us how many right classifications were made out of all the classifications. In other words, how many TPs and TNs were done out of TP + TN + FP + FNs. It tells the ratio of “True”s to the sum of “True”s and “False”s.

* **Precision:** Precision = TP / (TP + FP)
* **Recall or Sensitivity:**

Recall = TP/ (TN + FN)

* F1-Score:

F1 score = 2\* (Precision \* Recall) / (Precision + Recall)

* Area Under Curve (AUC) and ROC Curve:

4.

**i. In the sense of machine learning models, what is underfitting? What is the most common reason for underfitting?**

Underfitting refers to a model that can neither model the training data nor generalize to new data.

It when a model is too simple — informed by too few features or regularized too much

ii. What does it mean to overfit? When is it going to happen?

Overfitting is a modeling error in statistics that occurs when a function is too closely aligned to a limited set of data points. As a result, the model is useful in reference only to its initial data set, and not to any other data sets.

**iii. In the sense of model fitting, explain the bias-variance trade-off.**

In statistics and machine learning, the bias–variance tradeoff is the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters.

1. **Is it possible to boost the efficiency of a learning model? If so, please clarify how.**

Yes. Presence of more data results in better and accurate models

1. **How would you rate an unsupervised learning model's success?**

Twin sample validation can be used to validate results of unsupervised learning. It should be used in combination with internal validation. It can prove to be highly useful in case of time-series data where we want to ensure that our results remain same across time

**What are the most common success indicators for an unsupervised learning model?**

**7. Is it possible to use a classification model for numerical data or a regression model for categorical data with a classification model? Explain your answer.**

8. Describe the predictive modeling method for numerical values. What distinguishes it from categorical predictive modeling?

9. The following data were collected when using a classification model to predict the malignancy of a group of patients' tumors:

i. Accurate estimates – 15 cancerous, 75 benign

ii. Wrong predictions – 3 cancerous, 7 benign

Determine the model's error rate, Kappa value, sensitivity, precision, and F-measure.

10. Make quick notes on:

1. The process of holding out

2. Cross-validation by tenfold

3. Adjusting the parameters

11. Define the following terms:

1. Purity vs. Silhouette width

2. Boosting vs. Bagging

3. The eager learner vs. the lazy learner