Machine learning

1) What - is Machine learning? Explain selecting a Model in ML?
2) Explain Model Representation and interpretability
3) Explain evaluating performance of a model
4) Explain feature transformation in ML?

Answers

1.) Machine learning is a branch of Artificial intelligence

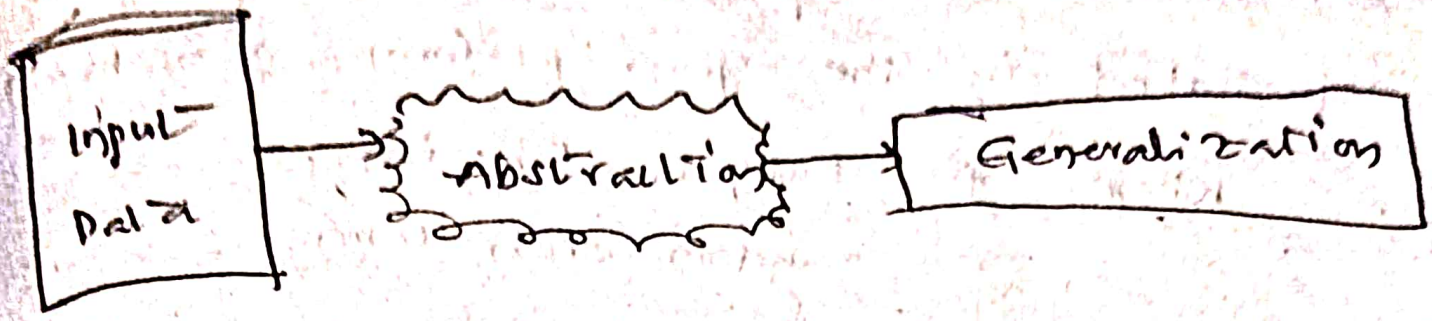
⇒ It is mainly concerned with development of algorithms which allow computer to learn from data & past experiences on their own.

⇒ introduced by Arthor samuel in 1959

⇒ "ML enables a machine to automatically learn from data, improve performance from experience and predict things without being explicity programmed.

⇒ A machine learning contains 3 parts

1. Data input
2. Abstraction
3. Generalization



Machine Learning Models :-

⇒ It is defined as the Mathematical Representation of the output of the training process

⇒ A Machine learning model is similar to computer software designed to recognize patterns or behaviours based on previous experience or data

⇒ In machine learning paradigm, the input variables or features are very important for model selection

⇒ The input variables are also called as predictors, attributes, features, independent variables or simply variables

⇒ Input variables can be denoted by X, while individual input variables are represented by $x_1, x_2, x_3, \ldots x_n$ and output variable by symbol Y.

⇒ The relationship between x and y are represented in the general form: $Y = f(X) + e$

Where f = target function
e = random error term

Classification of Machine Learning Models —

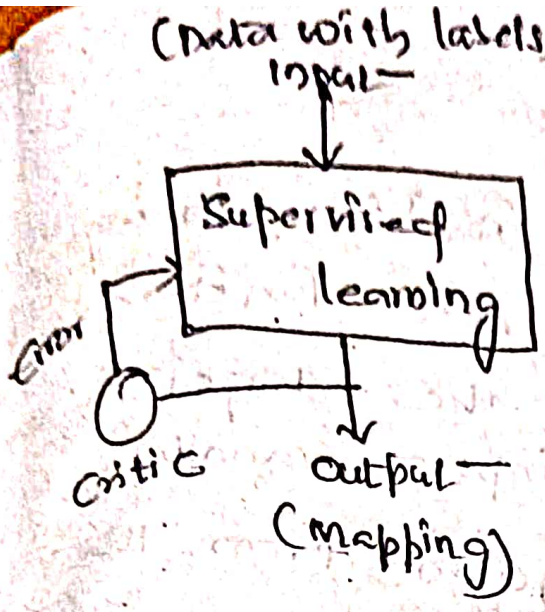
⇒ Based on different business goals and data sets, there are three learning models for algorithms.

Each Machine learning algorithm settles into one of the three models

1. Supervised — classification, Regression
2. Un supervised — clustering, Association Rule
3. Reinforcement — Dimensionality Reduction

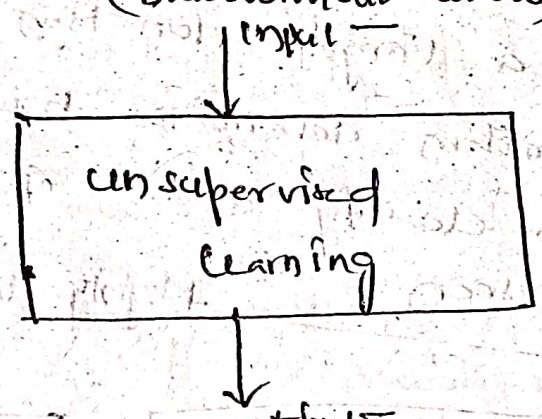
Supervised learning :—

Supervised learning is the simplest Machine learning model to understand in which input data is training data and has a known label (or result).



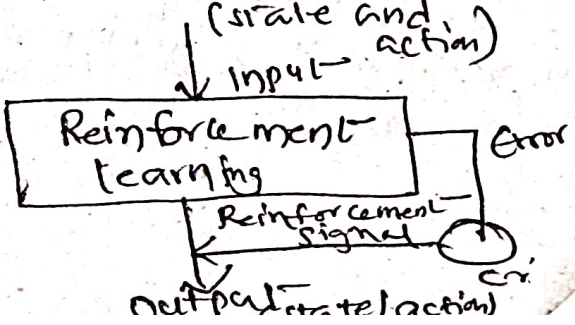
unsupervised learning :-

→ It is opposite to supervised learning, it enables the model to learn from the unlabelled training data set.



Reinforcement learning :-

→ In reinforcement learning, the algorithm learns actions for a given set of states that lead to a goal state.



2) MODEL Representation :-

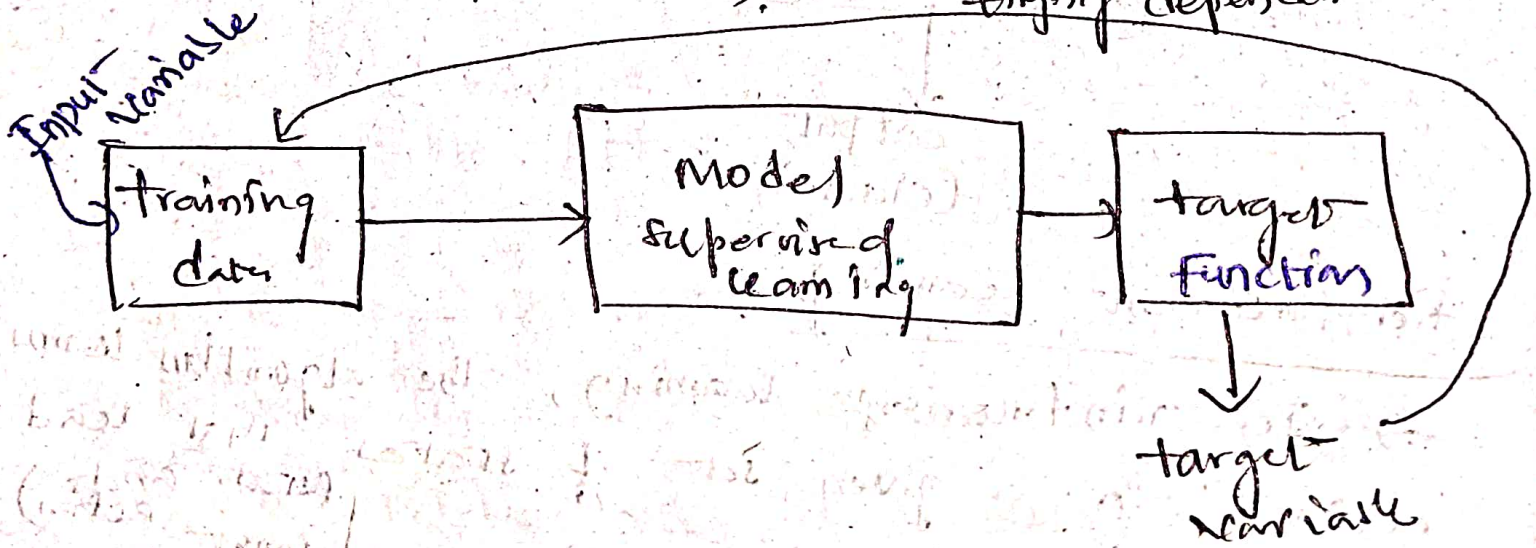
⇒ The goal of supervised machine learning is to learn (or derive a target function which can best determine the target variable from the set of input variables

⇒ learning the target function from the training data to the extent of generalization

→ Input data is limited, and specific

→ The new, unknown data is the test data set may be differing from the training data

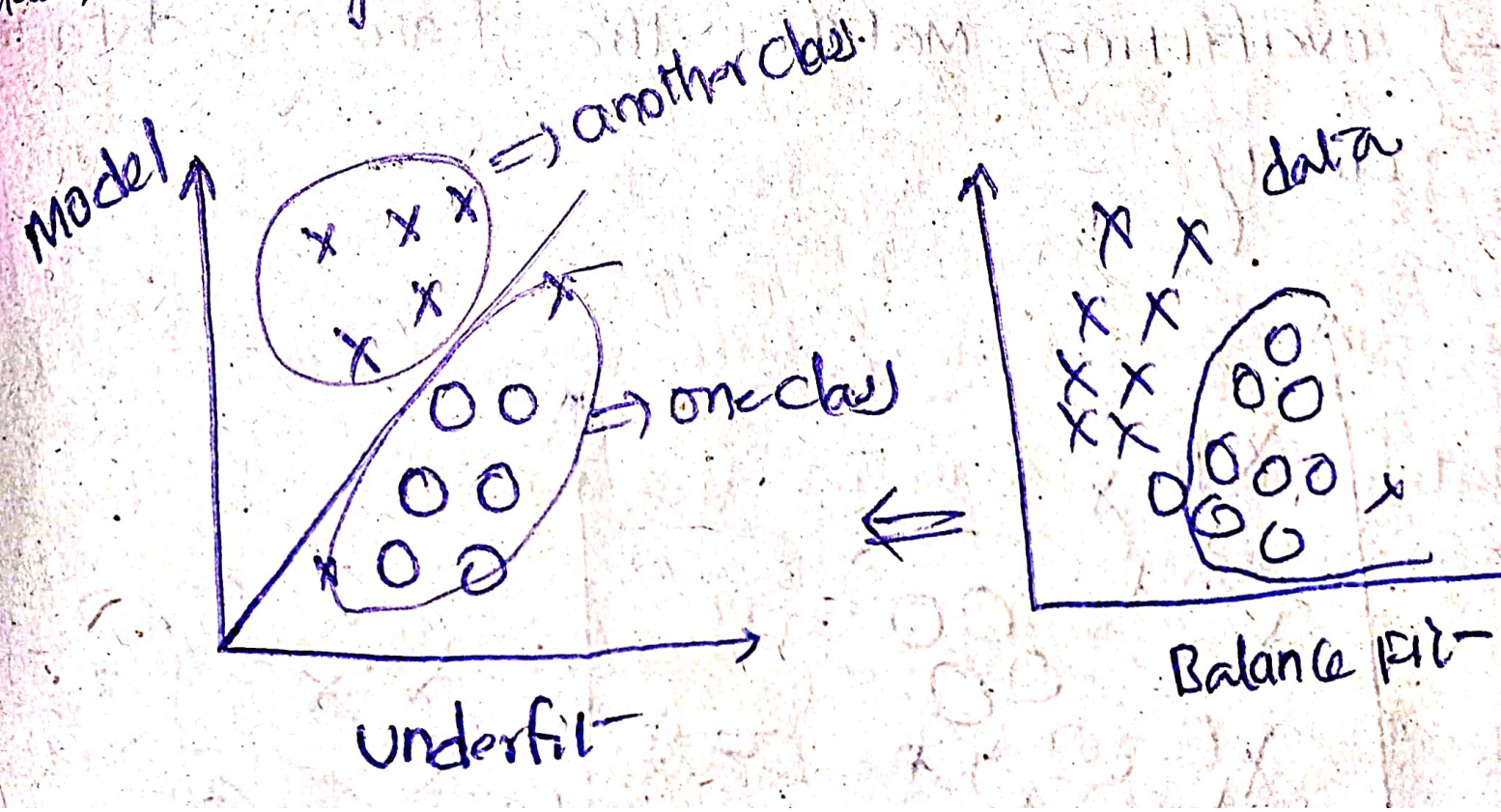
⇒ Fitness of a target function approximated by learning algorithm determines how correctly it is able to classify a set of data, that it has never seen.



Underfitting :-

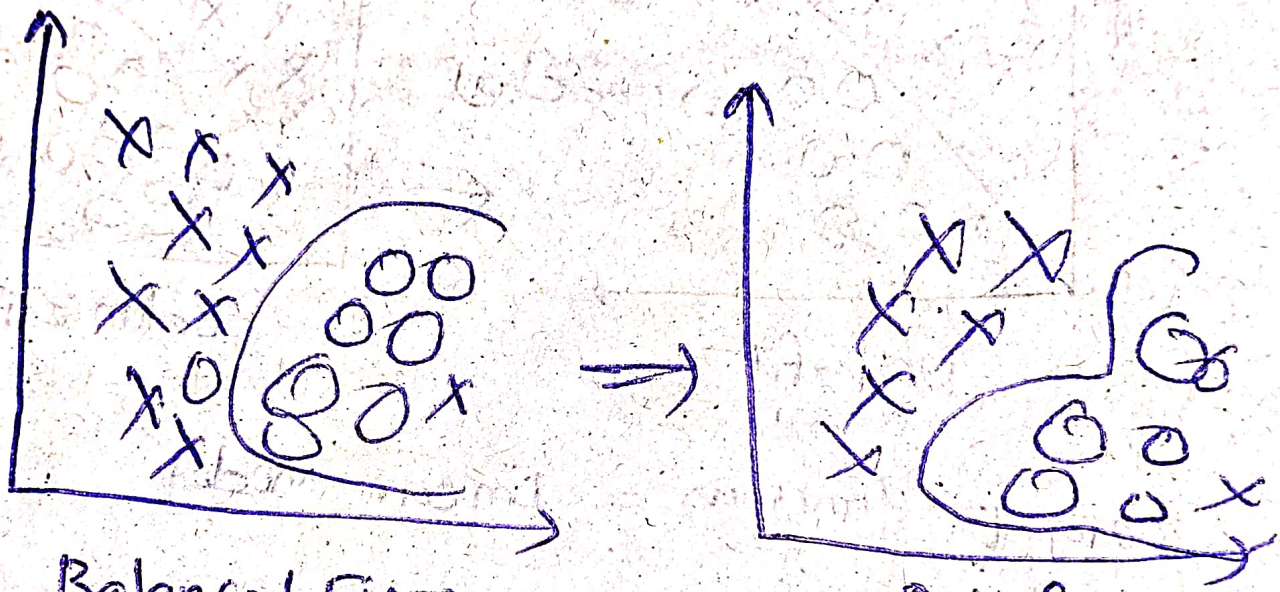
→ if the target function is kept too simple, it may not able to capture the essential output and represent the underlying data well.

→ underfitting may occur when trying to represent a non-linear data with a linear model as demonstrated by both cases of underfitting as shown in figure



The drawbacks of underfit will done by overfitting.

⇒ overfitting matches the training data too closely



Balanced Fit

Overfit

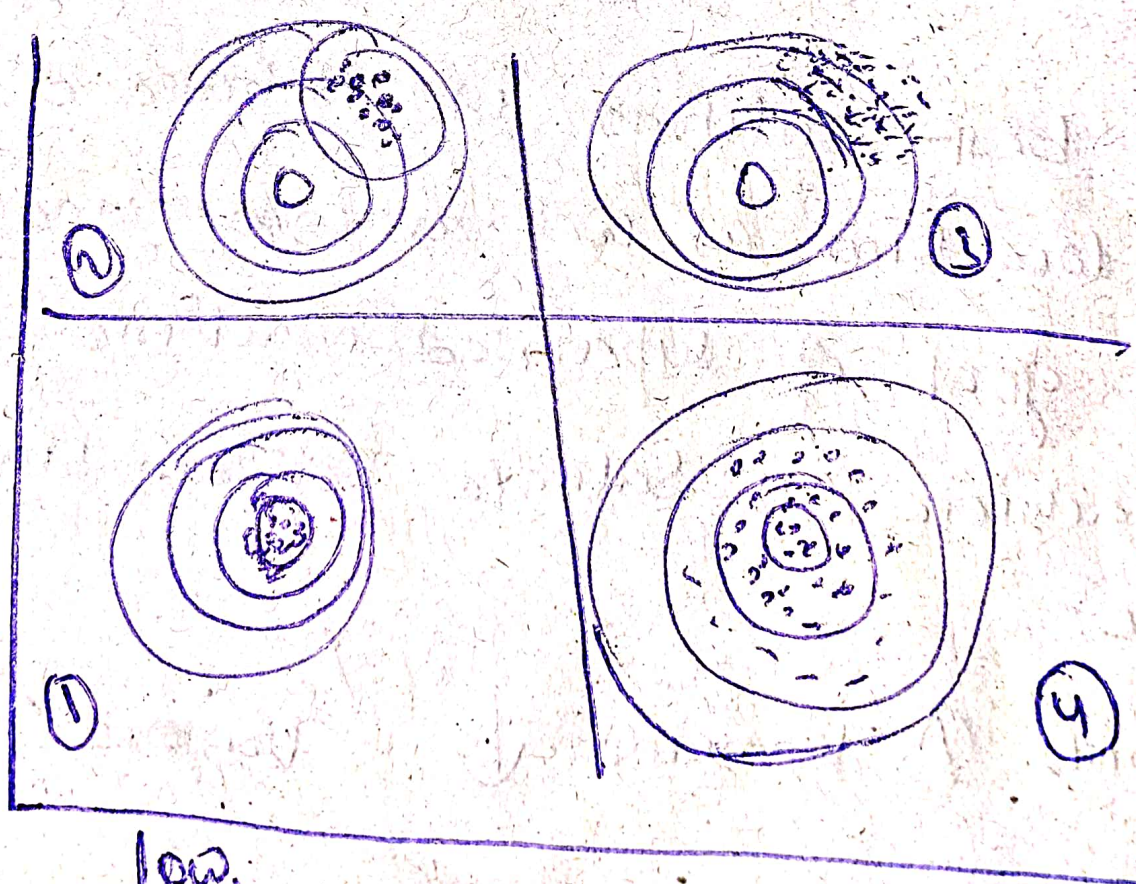
⇒ If any specific deviation in the training data, like noise (or) outliers, gets embeded in the model then performance automatically reduces.

Bias - variance trade-off :-

In supervised learning, the class value assigned by the learning model built based on the training data may differ from the actual class valu

→ The Error in learning can be of two type

- Error due to 'bias' and
- Error due to 'variance'.



Error due to bias :-

- Error due to bias arise from simplifying assumptions made by the model to make the target function less complex (or easier to learn)

⇒ It is due to underfitting.

⇒ underfitting results in high bias

Error due to variance :-

⇒ It occurs from difference in training data sets used to train the model.

⇒ In overfitting, since the model closely matches the training data

⇒ Even a small difference in training data gets magnified in the model.