Types of Machine Learning

TIET, PATIALA

Types of Machine Learning

- Machine Learning can be classified broadly into three types:
  - Supervised Learning
  - Unsupervised Learning
  - Reinforcement Learning
- The above mentioned methods can be combined in different ways and hence we have some hybrid learning problems :
  - Semi-supervised Learning
  - Self-supervised Learning
  - Multi-Instance Learning

Supervised Learning

- Supervised Learning occurs when an algorithm learns from examples data and associated target responses (which can be numeric values or string labels).
- Supervised learning is where we have input variables (x) and an output variable (Y) and use an algorithm to learn the mapping function from the input to the output **Y = f(x)** .
- The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.
- It is similar to human learning under the supervision of a teacher.
- The teacher provides good examples for the student to memorize, and the student then derives general rules from these specific examples.

Supervised Learning Problems

- In Machine Learning, we deal with two types of supervised learning problems:

1. *Classification*:

- Classification models are trained to classify data into categories.
- The categories are discrete in nature (such as yes/no, 0/1, small/medium/large, etc.)
- Binary Classification has two class labels where as a Multilabel classification has more than two labels.

2. *Regression:*

- In regression problems, the input variables are mapped to a continuous function.
- In other words, we try to predict result within a continuous output.

Supervised Learning Problems Contd….

- Example:

Credit Scoring: Given information of customers income, savings, age, profession, past financial history.

Classification: Predict whether a new customer is at low-risk or at high-risk in returning the credit.

Regression: Predict what is the probability that a new customer is at high-risk or low-risk.

Examples-Supervised Learning Problems

Classification:

- Customer Retention
- Fraud Detection
- Image Classification
- Image Segmentation
- Medical Diagnosis
- Anomaly Detection
- Handwriting Recognition /OCR
- Text Classification
- Email Classification
- Sentiment Analysis
- Opinion Mining
- Text Summarization

Regression:

- Population Growth Prediction
- Estimating Life Expectancy
- Market Forecasting (Stock Market, Bit Coin)
- Credit Policing
- Weather Forecasting
- Advertising Popularity Prediction/ New Insights
- House Price Prediction
- Process Optimization

Classification Algorithms

Bayesian

- **Naïve Bayes**
- **Gaussian Naïve Bayes**
- **Multinomial Naïve Bayes**
- **Bayesian Belief Networks**
- **Bayesian Network**

Decision Tree

- **Classification and Regression Tree (CART)**
- **Iterative Dichotomiser 3 (ID3)**
- C4.5
- C5.0
- M5
- Conditional

Instance-Based

- **K-Nearest Neighbor**
- Learning Vector Quantization (LVQ)
- Self-Organizing Map (SOM)
- Locally Weighted Learning (LWL)

Kernel-Based

- **Support Vector Machines**
  - Linear
  - Non Linear
  - Polynomial Kernel
  - Gaussian Kernel
  - Sigmoid
  - Hyperbolic

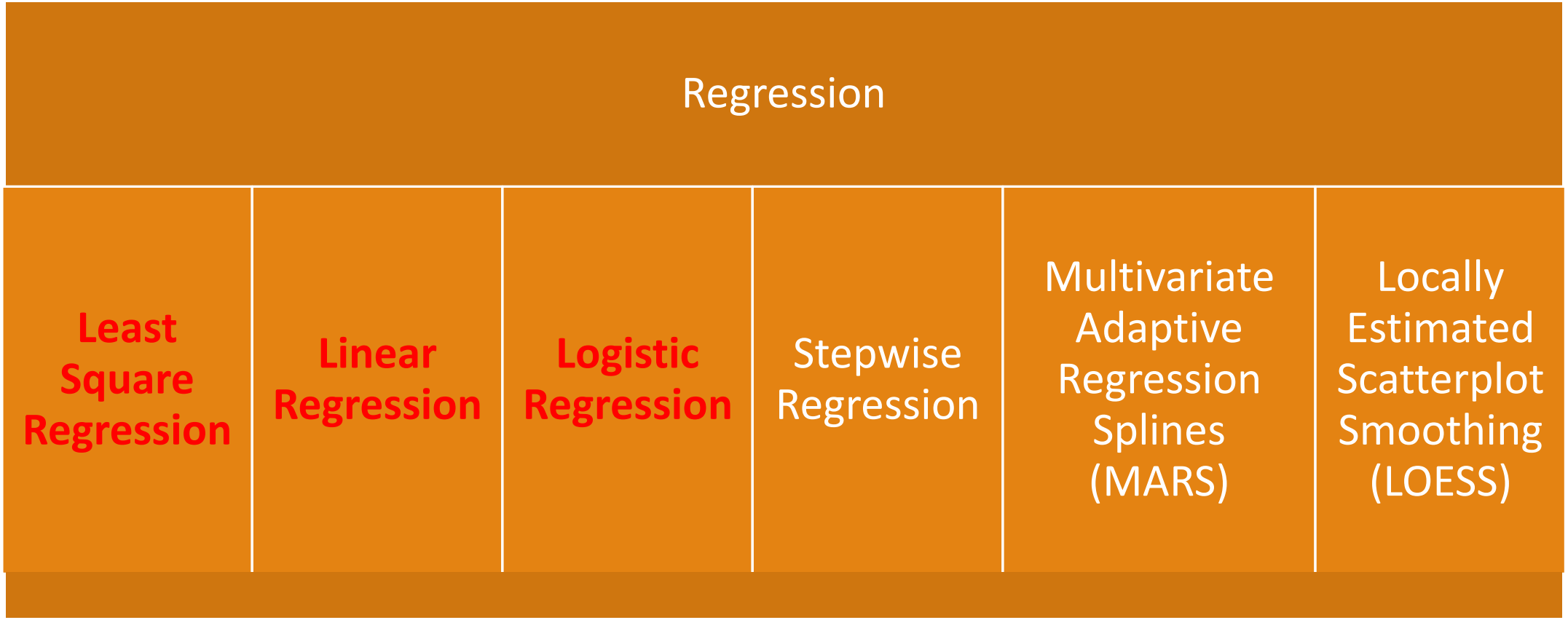
Neural Network – Based

- **Artificial Neural Network**
- **Deep Neural Network**
- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks (RNN)

Miscellaneous

- **Logistic Regression**
- **Random Forest**

Regression Algorithms



Unsupervised Learning

- Unsupervised Learning occurs when an algorithm learns from plain examples without any associated response.
- These type of algorithms determine the data patterns on its own and restructures the data into something else such as new features that may represent a class or new series of uncorrelated values.
- The goal of unsupervised learning is to discover patterns from data.
- As unlabeled data is more abundant than labeled data, machine learning methods that facilitate unsupervised learning are more valuable.

Unsupervised Learning Problems

- Unsupervised Learning methods are used in the following three types of problems:
  - Clustering
  - Learning Associations
  - Dimensionality Reduction

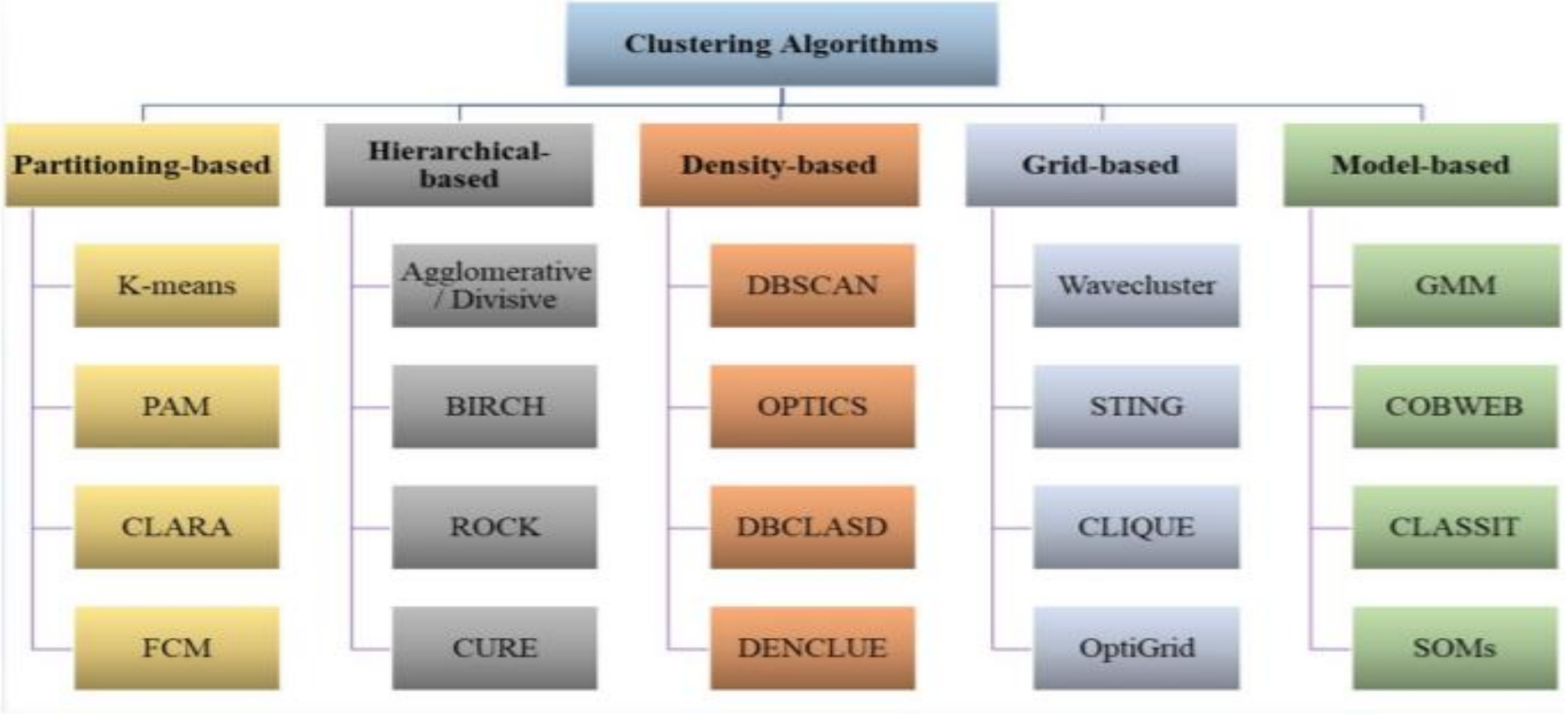
Clustering

- **Clustering** is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups.
- In simple words, the aim is to segregate groups with similar traits and assign them into clusters.

Clustering Examples

- Clustering algorithms are widely used in a number of applications such as:
  - Market Segmentation / Targeted Marketing / Recommender Systems
  - Document / News / Article Clustering
  - Biology / Genome Clustering
  - City Planning
  - Speech Recognition
  - Social Network Analysis
  - Organize Computing Clusters
  - Astronomical Data Analysis

Clustering Algorithms



Learning Associations

- Association rule learning is a type of unsupervised learning technique that checks for the dependency of one data item on another data item and maps accordingly so that it can be more profitable.
- It tries to find some interesting relations or associations among the variables of dataset.
- It is based on different rules to discover the interesting relations between variables in the database.
- It is employed in Market Basket analysis, Web usage mining, continuous production, medical diagnosis, protein sequencing, catalog design, etc.

Learning Association Algorithms

- Following three algorithms are used for learning associations:

1. **Apriori Algorithm**
2. **FP Growth Algorithm**
3. EClaT (Equivalence Class Transformation) Algorithm.

Dimensionality Reduction

- Large number of dimensions in the feature space can dramatically impact the performance of machine learning algorithms, generally referred to as the "**curse of dimensionality**."
- Therefore, it is often desirable to reduce the number of input features.
- **Dimensionality reduction**, or **dimension reduction**, is the transformation of data from a high-**dimensional** space into a low-**dimensional** space so that the low-**dimensional** representation retains some meaningful properties of the original data.
- Dimensionality Reduction is employed in a number of applications such as Structure Discovery, Big Data Visualization, noise reduction, signal processing, speech recognition.

Dimensionality Reduction Algorithms

Feature Selection

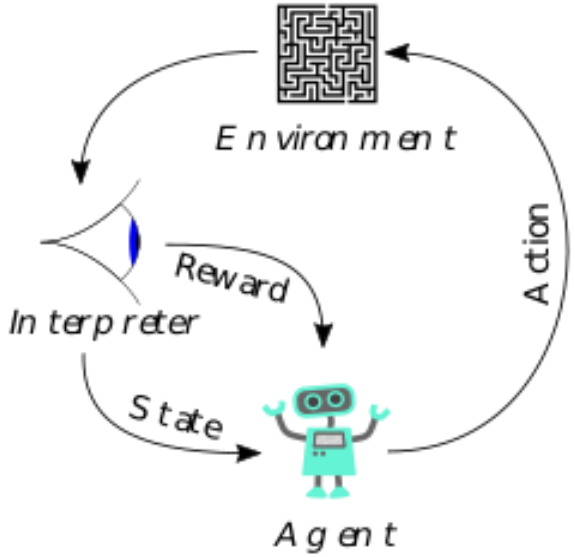
- **Filter Strategy**
- **Wrapper Strategy**
- **Embedded Strategy**

Feature Projection

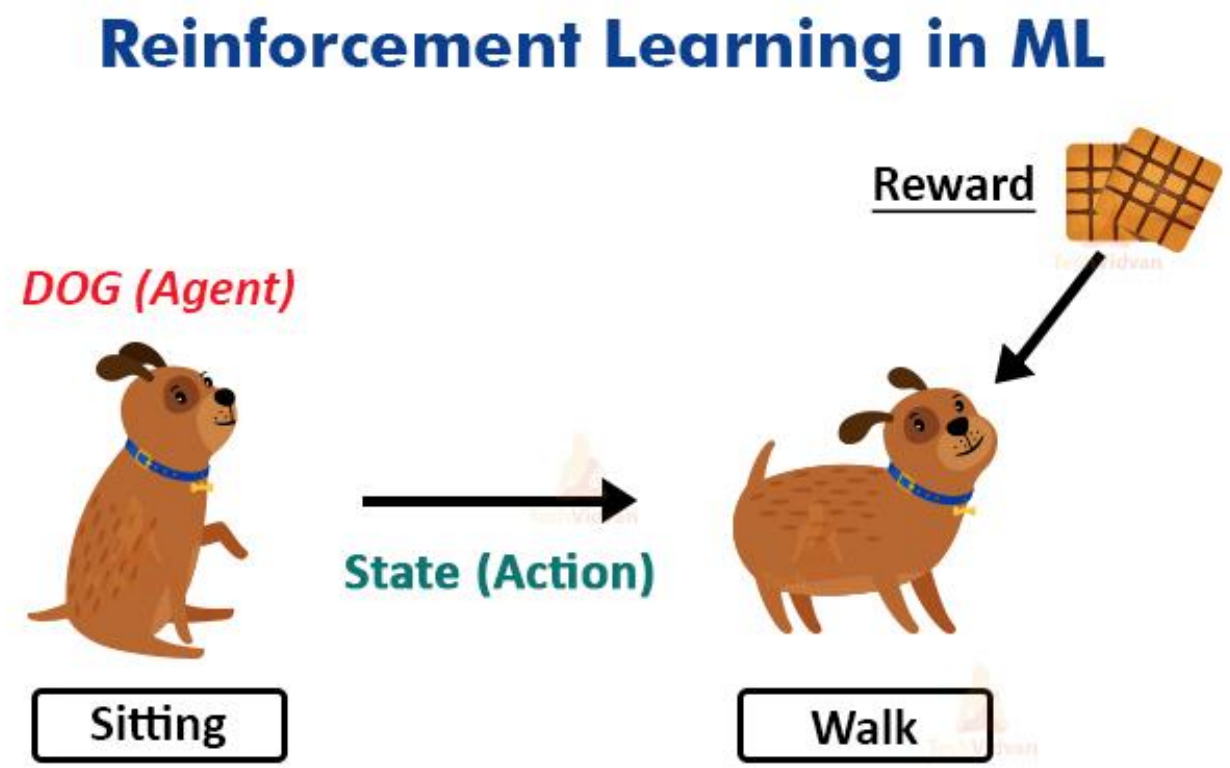
- **Single Valued Decomposition (SVD)**
- **Principle Component Analysis (PCA)**
- **Linear Discriminant Analysis (LDA)**
- Auto Encoders
- t-SNE
- UMAP

Reinforcement Learning

- **Reinforcement learning** (**RL**) is an area of machine learning concerned with how intelligent agents ought to take actions in an environment in order to maximize the notion of cumulative reward.



Reinforcement Learning



Let's say we have a dog and we are trying to train your dog to sit.

We would give certain instructions to the dog to try to make it learn.

If the dog executes the instruction perfectly, it would get a biscuit as a reward. If not, it would not get anything.

The dog learns from this after some tries that it would get a biscuit if it sits.

Reinforcement Learning Algorithms

- Reinforcement learning is studied in many disciplines, such as game theory, control theory, operations research, information theory, simulation-based optimization, multi-agent systems, swarm intelligence, and statistics.
- The major algorithms for reinforcement learning includes Q-learning, Model-Based learning (Value Iteration, Policy Iteration), and Temporal Difference Learning

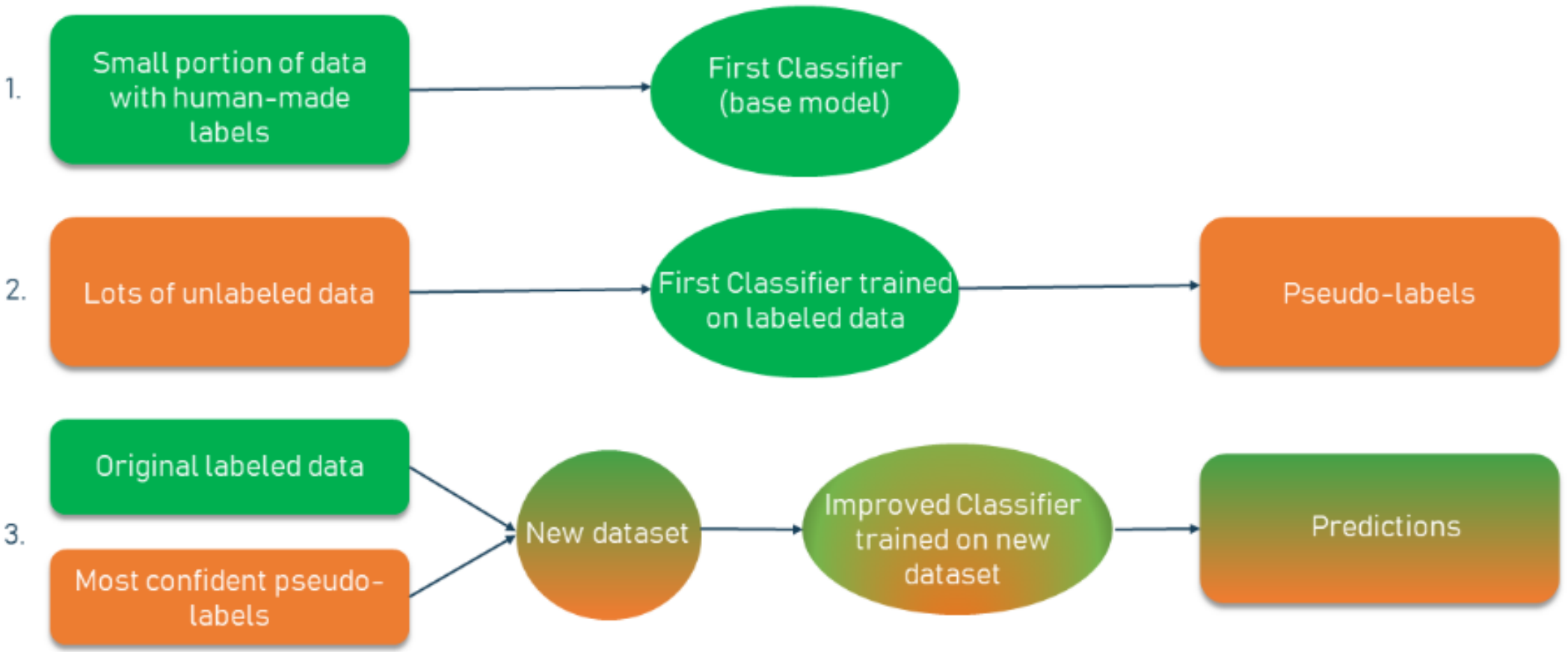
Semi-Supervised Learning

- This is a combination of supervised and unsupervised learning.
- This method helps to reduce the shortcomings of both the above learning methods.
- In supervised learning, labelling of data is manual work and is very costly as data is huge. In unsupervised learning, the areas of application are very limited.
- To reduce these problems, **semi-supervised** learning is used.

Semi-Supervised Learning Contd…

- In this, the model first trains under unsupervised learning. This ensures that most of the unlabeled data divide into clusters.
- For the remaining unlabeled data, the generation of labels takes place and classification carries with ease.
- This technique is very useful in areas like speech recognition and analysis, protein classification, text classification, etc.

SEMI-SUPERVISED SELF-TRAINING METHOD



Self-Supervised Learning

- It is a more advanced version of unsupervised learning which requires supervisory data along with it.
- Only in this case, the labelling of the data is not done by humans.
- The model itself extracts and labels the data. It does so with the help of the embedded metadata as supervisory data.
- A common example of self-supervised learning is computer vision where a corpus of unlabeled images is available and can be used to train a supervised model, such as identify the same image with different rotations.
- A general example of self-supervised learning algorithms are autoencoders and Generative Adversial Networks (GANs)

Multi-Instance Learning

- Multiple Instance Learning or MIL is another variation of supervised learning. Here, the training data isn't labelled individually, it is nicely arranged in bags.
- An arranged set of training data is called bags and the entire bag is labelled. This is a weaker but an interesting form of supervised learning method.
- If one entity is fitted with the result, it's entire bag is given positive. But if it does not fit, the entire bag equates to negative.
- The goal of this method is to classify **unseen** bags based on labelled bags.

Recent Trends in ML Learning

Multi-Task Learning

- A type of supervised learning that involves fitting a model on one dataset that addresses multiple related problems.
- Applications: Word Embeddings, Spam Filtering, Web Searching

Ensemble Learning

- Ensemble learning is an approach where two or more modes are fit on the same data and the predictions from each model are combined.

Transfer Learning

- Is a type of learning where a model is first trained on one task, then some or all of the model is used as the starting point for a related task.
- Applications in NLP and Image Classification