ARTIFICIAL INTELLIGENCE

UNIT 5: LEARNING

PROF. JESLEENA GONSALVES

TOPICS

- Inductive learning
- Types of learning
- Supervised - decision trees classification
- Unsupervised learning – K-means clustering.

LEARNING

- A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience.
- Learning agent can be thought of as containing a performance element that decides what actions to take and a learning element that modifies the performance element so that it makes better decisions.
- The design of a learning element is affected by three major issues:

1. Which components of the performance element are to be learned.
2. What feedback is available to learn these components.
3. What representation is used for the components.

- The type of feedback available for learning is usually the most important factor in determining the nature of the learning problem that the agent faces.
- The field of machine learning usually distinguishes three cases: supervised, unsupervised, and reinforcement learning.

INDUCTIVE LEARNING

- Reasoning from a set of examples to produce a general rules. The rules should be applicable to new examples, but there is no guarantee that the result will be correct.
- **Deductive Learning:** Reasoning from a set of known facts and rules to produce additional rules that are guaranteed to be true.

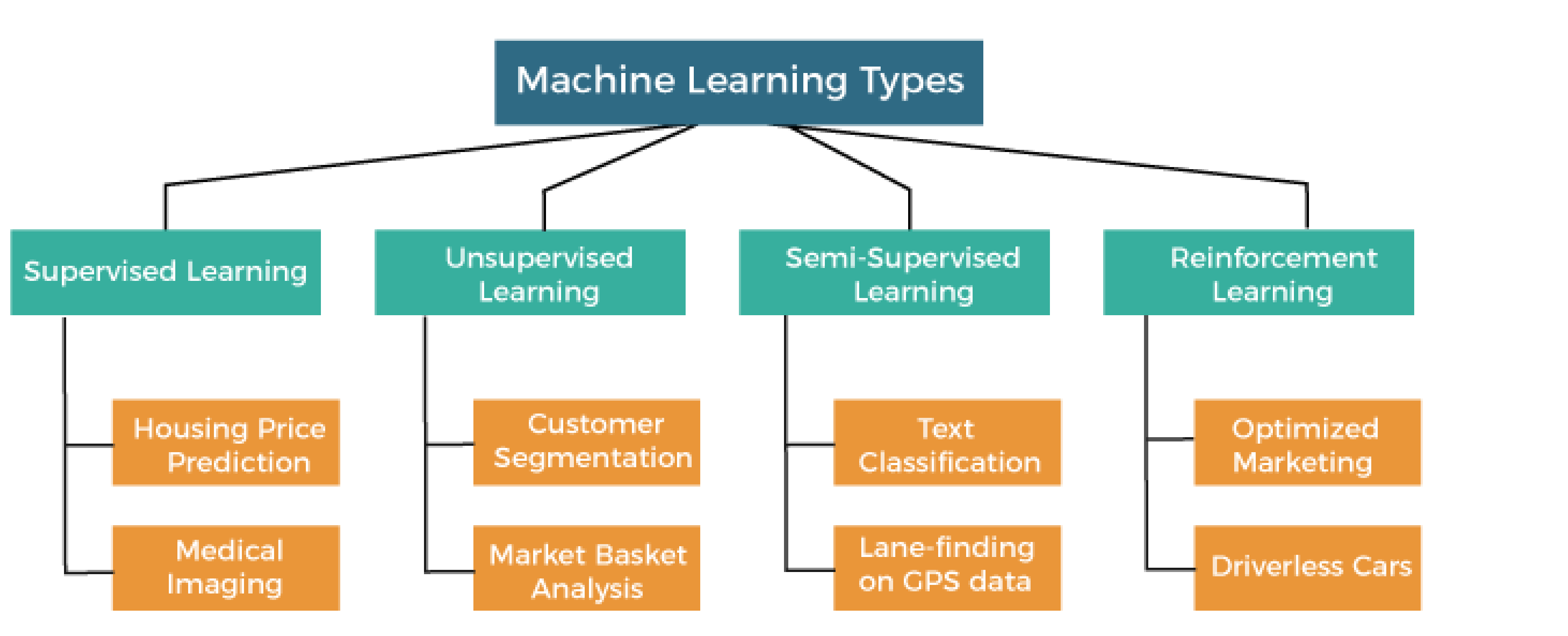
INDUCTIVE LEARNING

- An example is a pair (x, f (z)), where x is the input and f(x) is the output of the function applied to x. The task of pure inductive inference (or induction) is this:

Given a collection of examples of f , return a function h that approximates f .

- The function h is called a hypothesis. The reason that learning is difficult, from a conceptual point of view, is that it is not easy to tell whether any particular h is a good approximation of f .
- A good hypothesis will generalize well-that is, will predict unseen examples correctly.
- This is the fundamental problem of induction.

TYPES OF LEARNING



SUPERVISED LEARNING

- Supervised learning is commonly used in real world applications, such as face and speech recognition, products or movie recommendations, and sales forecasting.
- Supervised learning can be further classified into two types - Regression and Classification.
- Regression trains on and predicts a continuous-valued response, for example predicting real estate prices.
- Classification attempts to find the appropriate class label, such as analyzing positive/negative sentiment, male and female persons, benign and malignant tumors, secure and unsecure loans etc.
- In supervised learning, learning data comes with description, labels, targets or desired outputs and the objective is to find a general rule that maps inputs to outputs.
- This kind of learning data is called labeled data. The learned rule is then used to label new data with unknown outputs.

SUPERVISED LEARNING

- Supervised learning involves building a machine learning model that is based on labeled samples.
- For example, if we build a system to estimate the price of a plot of land or a house based on various features, such as size, location, and so on, we first need to create a database and label it. We need to teach the algorithm what features correspond to what prices. Based on this data, the algorithm will learn how to calculate the price of real estate using the values of the input features.
- Supervised learning deals with learning a function from available training data. Here, a learning algorithm analyzes the training data and produces a derived function that can be used for mapping new examples.
- There are many supervised learning algorithms such as Logistic Regression, Neural networks, Support Vector Machines (SVMs), and Naive Bayes classifiers.
- Common examples of supervised learning include classifying e-mails into spam and not-spam categories, labeling webpages based on their content, and voice recognition.

SUPERVISED LEARNING

Advantages:

- Since supervised learning work with the labelled dataset so we can have an exact idea about the classes of objects.
- These algorithms are helpful in predicting the output on the basis of prior experience.

Disadvantages:

- These algorithms are not able to solve complex tasks.
- It may predict the wrong output if the test data is different from the training data.
- It requires lots of computational time to train the algorithm.

UNSUPERVISED LEARNING

- Unsupervised learning is used to detect anomalies, outliers, such as fraud or defective equipment, or to group customers with similar behaviors for a sales campaign. It is the opposite of supervised learning. There is no labeled data here.
- When learning data contains only some indications without any description or labels, it is up to the coder or to the algorithm to find the structure of the underlying data, to discover hidden patterns, or to determine how to describe the data. This kind of learning data is called unlabeled data.
- Suppose that we have a number of data points, and we want to classify them into several groups. We may not exactly know what the criteria of classification would be. So, an unsupervised learning algorithm tries to classify the given dataset into a certain number of groups in an optimum way.
- Unsupervised learning algorithms are extremely powerful tools for analyzing data and for identifying patterns and trends. They are most commonly used for clustering similar input into logical groups. Unsupervised learning algorithms include K-means, Random Forests, Hierarchical clustering and so on.

UNSUPERVISED LEARNING

Advantages:

- These algorithms can be used for complicated tasks compared to the supervised ones because these algorithms work on the unlabeled dataset.
- Unsupervised algorithms are preferable for various tasks as getting the unlabeled dataset is easier as compared to the labelled dataset.

Disadvantages:

- The output of an unsupervised algorithm can be less accurate as the dataset is not labelled, and algorithms are not trained with the exact output in prior.
- Working with Unsupervised learning is more difficult as it works with the unlabeled dataset that does not map with the output.

SEMI-SUPERVISED LEARNING

- If some learning samples are labeled, but some other are not labeled, then it is semi-supervised learning.
- It makes use of a large amount of unlabeled data for training and a small amount of labeled data for testing. Semi-supervised learning is applied in cases where it is expensive to acquire a fully labeled dataset while more practical to label a small subset.
- For example, it often requires skilled experts to label certain remote sensing images, and lots of field experiments to locate oil at a particular location, while acquiring unlabeled data is relatively easy.

SEMI-SUPERVISED LEARNING

Advantages:

- It is simple and easy to understand the algorithm.
- It is highly efficient.
- It is used to solve drawbacks of Supervised and Unsupervised Learning algorithms.

Disadvantages:

- Iterations results may not be stable.
- We cannot apply these algorithms to network-level data.
- Accuracy is low.

REINFORCEMENT LEARNING

- Here learning data gives feedback so that the system adjusts to dynamic conditions in order to achieve a certain objective.
- The system evaluates its performance based on the feedback responses and reacts accordingly.
- Reinforcement learning is categorized mainly into two types of methods/algorithms:

1. **Positive Reinforcement Learning:** Positive reinforcement learning specifies increasing the tendency that the required behaviour would occur again by adding something. It enhances the strength of the behaviour of the agent and positively impacts it.
2. **Negative Reinforcement Learning:** Negative reinforcement learning works exactly opposite to the positive RL. It increases the tendency that the specific behaviour would occur again by avoiding the negative condition.

REINFORCEMENT LEARNING

Advantages

- It helps in solving complex real-world problems which are difficult to be solved by general techniques.
- The learning model of RL is similar to the learning of human beings; hence most accurate results can be found.
- Helps in achieving long term results.

Disadvantage

- RL algorithms are not preferred for simple problems.
- RL algorithms require huge data and computations.
- Too much reinforcement learning can lead to an overload of states which can weaken the results.

SUPERVISED - DECISION TREES CLASSIFICATION

- Predicts categorical class labels (discrete or nominal)
- Classifies data (constructs a model) based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data.

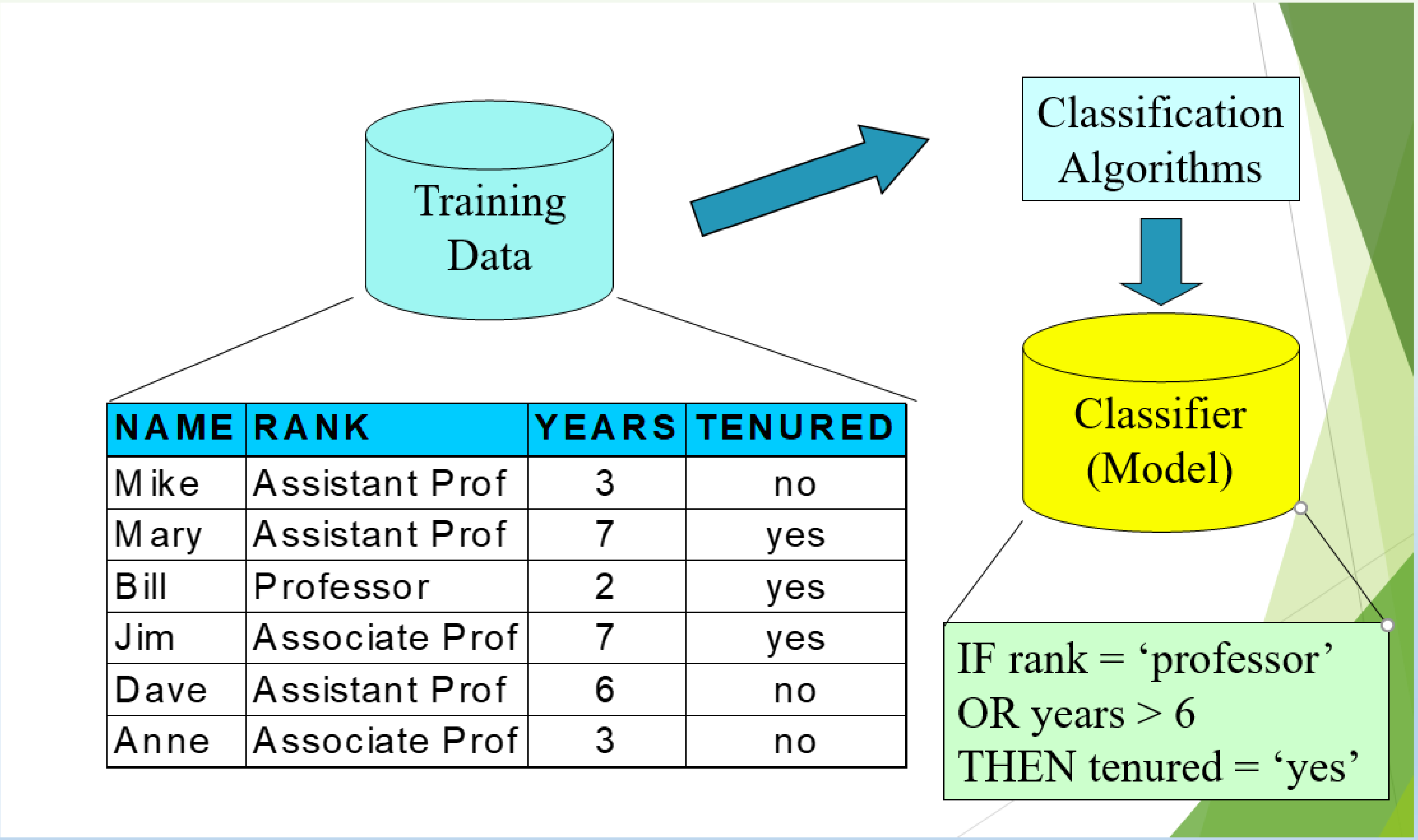
CLASSIFICATION- A TWO STEP PROCESS

- Model construction: describing a set of predetermined classes
  - Each tuple/sample is assumed to belong to a predefined class, as determined by the class label attribute
  - The set of tuples used for model construction is training set
  - The model is represented as classification rules, decision trees, or mathematical formulae

- Model usage: for classifying future or unknown objects
  - Estimate accuracy of the model
    - The known label of test sample is compared with the classified result from the model.
    - Accuracy rate is the percentage of test set samples that are correctly classified by the model.
    - Test set is independent of training set (otherwise overfitting)
  - If the accuracy is acceptable, use the model to classify new data

SUPERVISED - DECISION TREES CLASSIFICATION

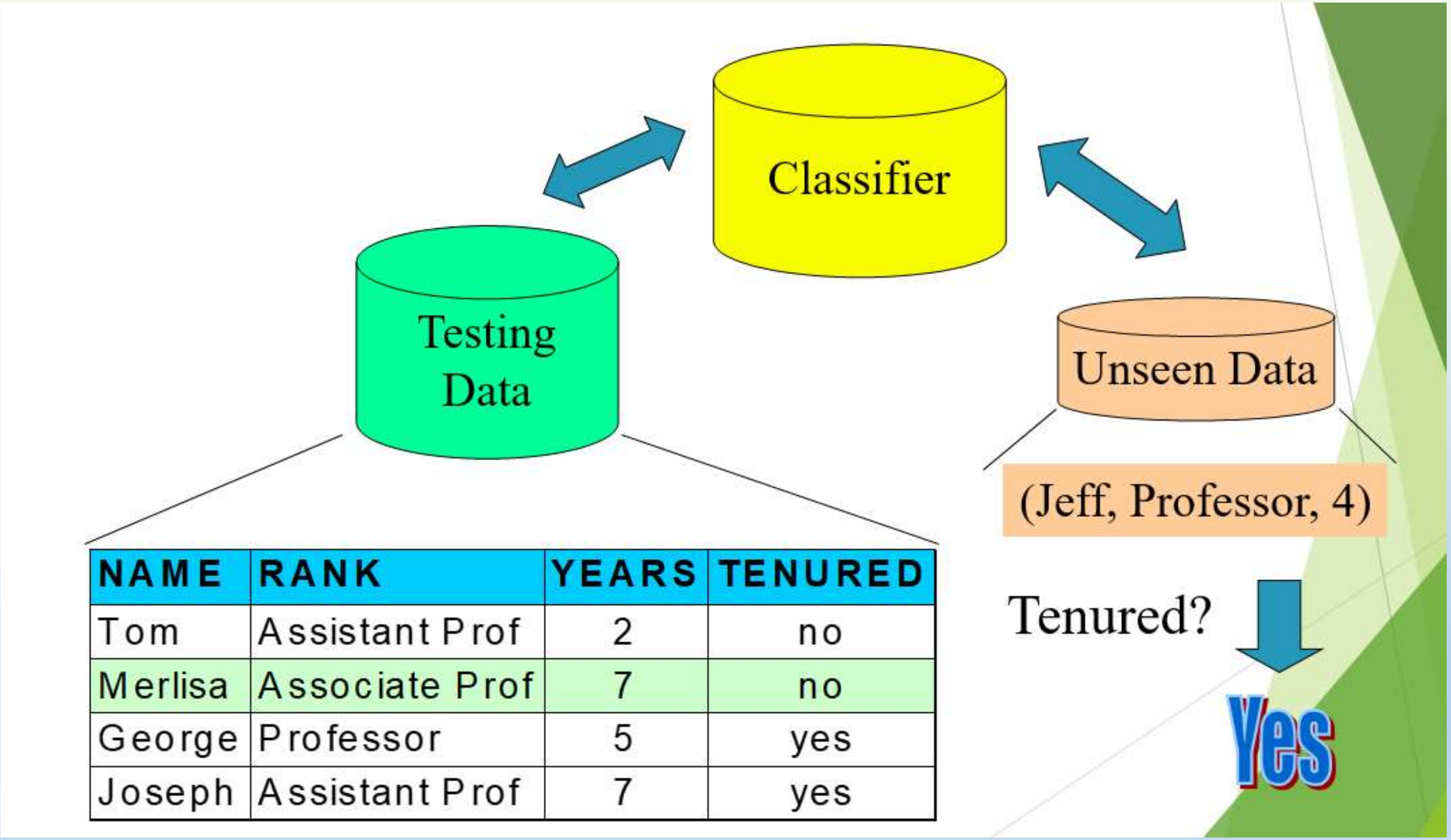
Step 1: Model Construction



NAME	RANK	YEARS	TENURED
Mike	Assistant Prof	3	no
Mary	Assistant Prof	7	yes
Bill	Professor	2	yes
Jim	Associate Prof	7	yes
Dave	Assistant Prof	6	no
Anne	Associate Prof	3	no

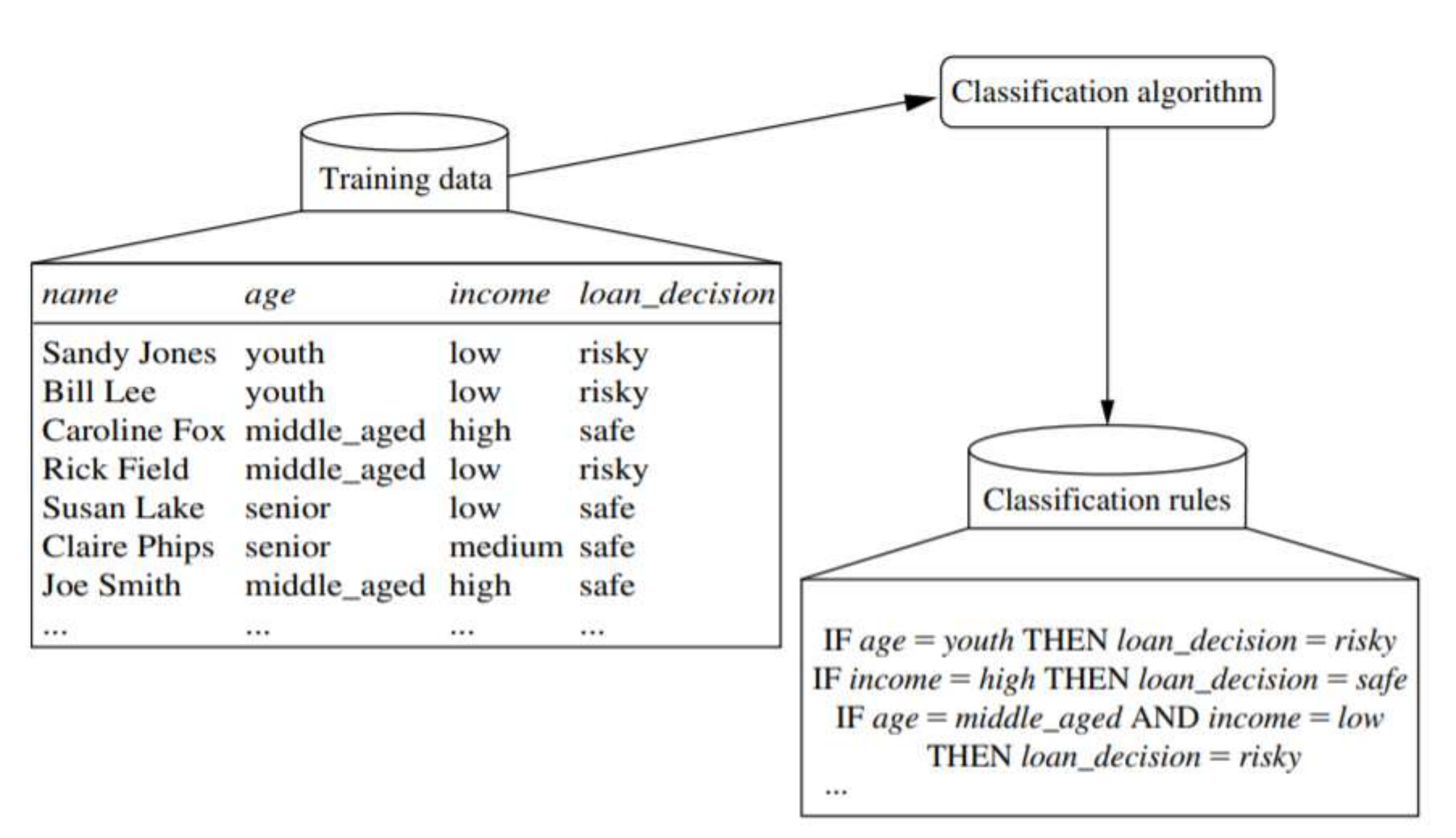
SUPERVISED - DECISION TREES CLASSIFICATION

Step 2: Using the model in Prediction

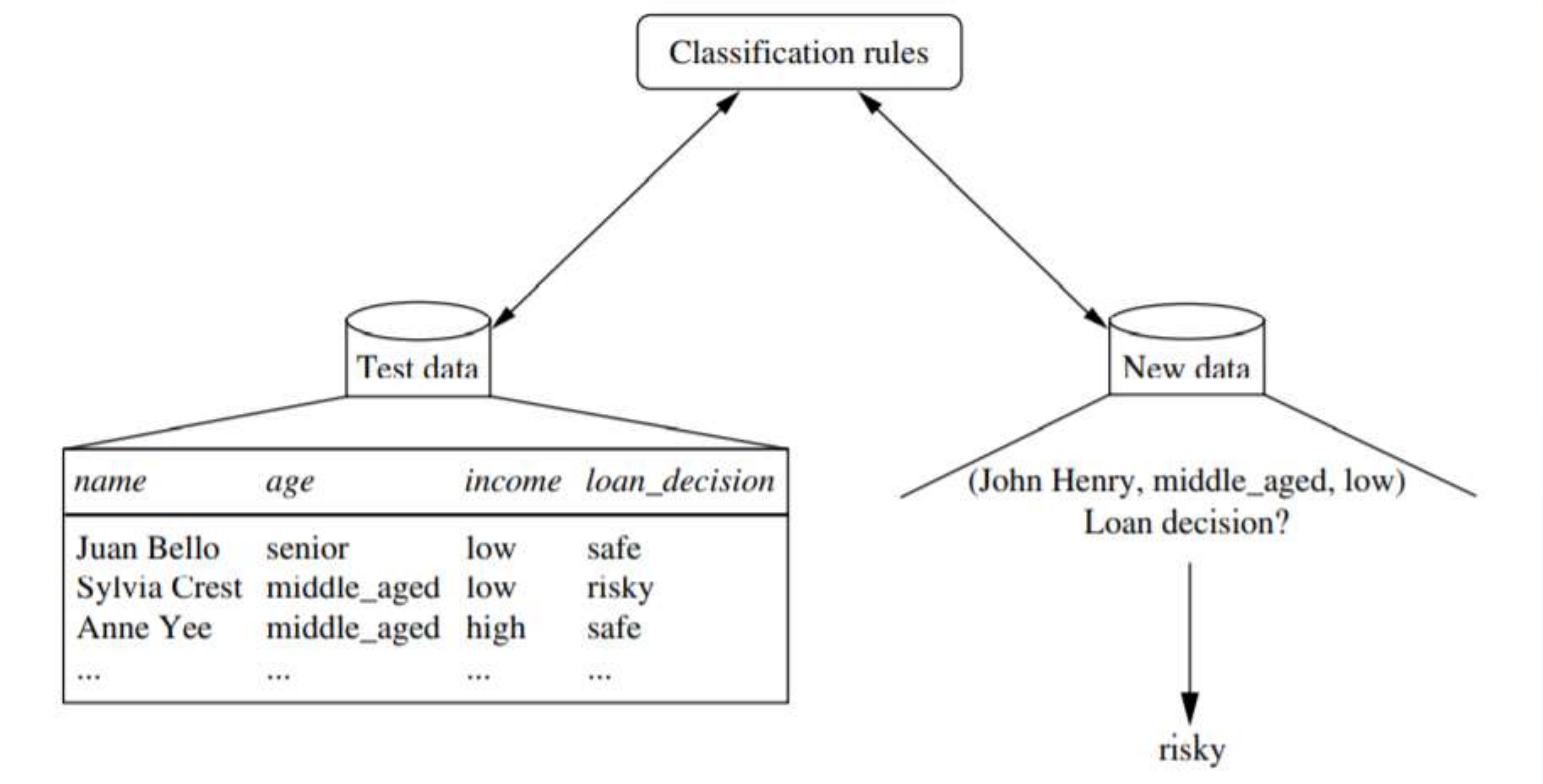


NAME	RANK	YEARS	TENURED
Tom	Assistant Prof	2	no
Merlisa	Associate Prof	7	no
George	Professor	5	yes
Joseph	Assistant Prof	7	yes

SUPERVISED - DECISION TREES CLASSIFICATION



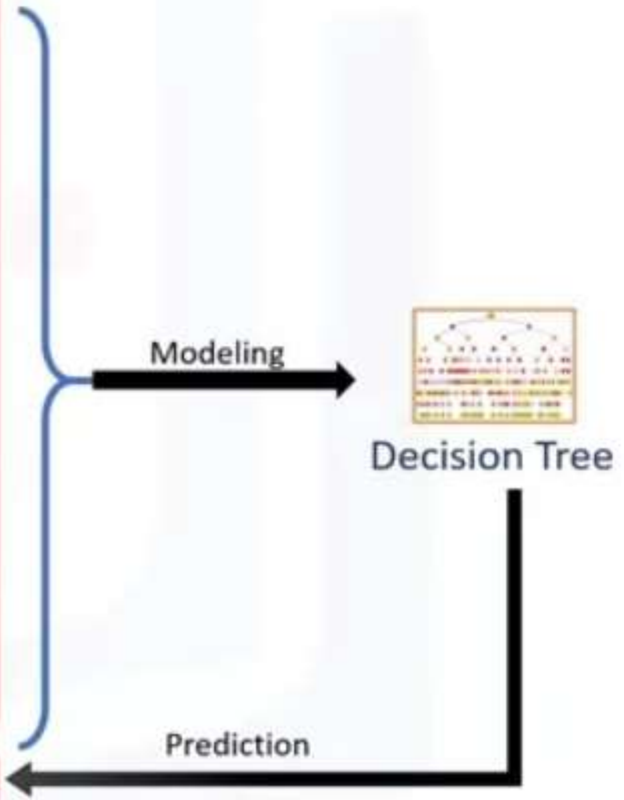
SUPERVISED - DECISION TREES CLASSIFICATION



SUPERVISED - DECISION TREES CLASSIFICATION

How to build a decision tree?

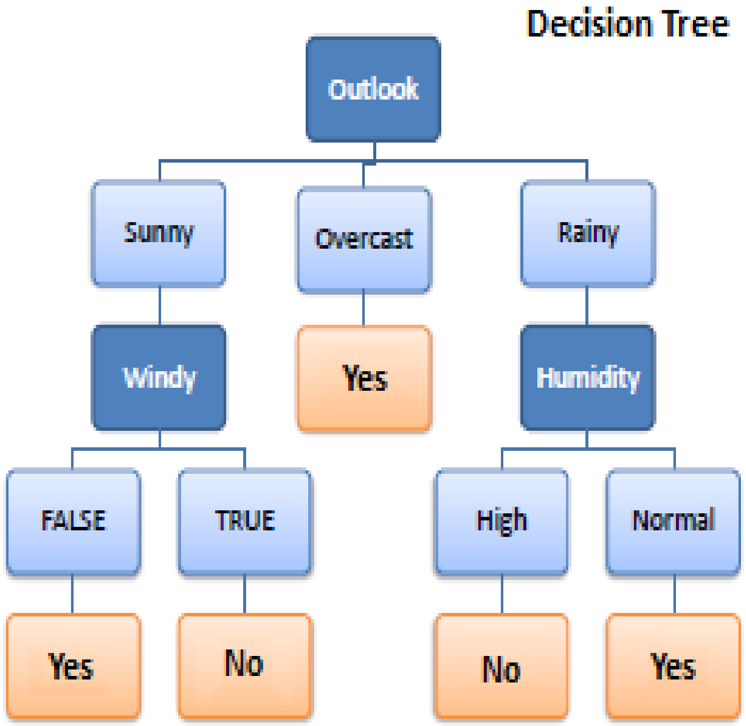
Patient ID	Age	Sex	BP	Cholesterol	Drug
p1	Young	F	High	Normal	Drug A
p2	Young	F	High	High	Drug A
p3	Middle-age	F	Hiigh	Normal	Drug B
p4	Senior	F	Normal	Normal	Drug B
p5	Senior	M	Low	Normal	Drug B
p6	Senior	M	Low	High	Drug A
p7	Middle-age	M	Low	High	Drug B
p8	Young	F	Normal	Normal	Drug A
p9	Young	M	Low	Normal	Drug B
p10	Senior	M	Normal	Normal	Drug B
p11	Young	M	Normal	High	Drug B
p12	Middle-age	F	Normal	High	Drug B
p13	Middle-age	M	High	Normal	Drug B
p14	Senior	F	Normal	High	Drug A
p15	Middle-age	F	Low	Normal	?



SUPERVISED - DECISION TREES CLASSIFICATION

Predictors | Target

Outlook	Temp.	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



Refer this link to understand the entire sum:

https://kindsonthegenius.com/blog/how-to-build-a-decision-tree-for-classification-step-by-step-procedure-using-entropy-and-gain/

SUPERVISED - DECISION TREES CLASSIFICATION

Construction of Decision Tree:

- A tree can be “learned” by splitting the source set into subsets based on Attribute Selection Measures.
- Attribute selection measure (ASM) is a criterion used in decision tree algorithms to evaluate the usefulness of different attributes for splitting a dataset.
- The goal of ASM is to identify the attribute that will create the most homogeneous subsets of data after the split, thereby maximizing the information gain.
- This process is repeated on each derived subset in a recursive manner called recursive partitioning.
- The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions.
- The construction of a decision tree classifier does not require any domain knowledge or parameter setting and therefore is appropriate for exploratory knowledge discovery.
- Decision trees can handle high-dimensional data.

SUPERVISED - DECISION TREES CLASSIFICATION

Entropy

- Entropy is a measure of the randomness in the information being processed.
- The higher the entropy, the harder it is to draw any conclusions from that information.
- Flipping a coin is an example of an action that provides information that is random.

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

Play Golf	
Yes	No
9	5

Entropy(PlayGolf) = Entropy (5,9)
= Entropy (0.36, 0.64)
= $-(0.36 \log_2 0.36) - (0.64 \log_2 0.64)$
= 0.94

Where **S → Current state, and Pi → Probability of an event *i* of state S or Percentage of class *i* in a node of state S.**

SUPERVISED - DECISION TREES CLASSIFICATION

Information Gain

- Information gain or IG is a statistical property that measures how well a given attribute separates the training examples according to their target classification.
- Constructing a decision tree is all about finding an attribute that returns the highest information gain and the smallest entropy.
- Information gain is a decrease in entropy. It computes the difference between entropy before split and average entropy after split of the dataset based on given attribute values.

$$\text{Information Gain}(T,X) = \text{Entropy}(T) - \text{Entropy}(T, X)$$

$$\begin{aligned} IG(\text{PlayGolf, Outlook}) &= E(\text{PlayGolf}) - E(\text{PlayGolf, Outlook}) \\ &= 0.940 - 0.693 \\ &= 0.247 \end{aligned}$$

SUPERVISED - DECISION TREES CLASSIFICATION

Gini Index

- You can understand the Gini index as a cost function used to evaluate splits in the dataset.
- It is calculated by subtracting the sum of the squared probabilities of each class from one.
- It favors larger partitions and easy to implement whereas information gain favors smaller partitions with distinct values.

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2$$

Gini Index

- Gini Index works with the categorical target variable “Success” or “Failure”. It performs only Binary splits.

UNSUPERVISED : K- MEANS CLUSTERING

- **Cluster:** A collection of data objects
  - ***similar*** (or related) to one another within the same group
  - ***dissimilar*** (or unrelated) to the objects in other groups
- Cluster analysis (or *clustering*, *data segmentation*, …)
  - Finding similarities between data according to the characteristics found in the data and grouping similar data objects into clusters
- Unsupervised learning: no predefined classes (i.e., *learning by observations* vs. learning by examples: supervised)
- Typical applications
  - As a stand-alone tool to get insight into data distribution
  - As a preprocessing step for other algorithms

UNSUPERVISED : K- MEANS CLUSTERING

Applications of Clustering:

- **Biology:** taxonomy of living things: kingdom, phylum, class, order, family, genus and species
- **Information retrieval:** document clustering
- **Land use:** Identification of areas of similar land use in an earth observation database
- **Marketing:** Help marketers discover distinct groups in their customer bases, and then use this knowledge to develop targeted marketing programs
- **City-planning:** Identifying groups of houses according to their house type, value, and geographical location
- **Earth-quake studies:** Observed earth quake epicenters should be clustered along continent faults
- **Climate:** understanding earth climate, find patterns of atmospheric and ocean
- **Economic Science:** market research

UNSUPERVISED : K- MEANS CLUSTERING

- A good clustering method will produce high quality clusters
  - high intra-class similarity: cohesive within clusters
  - low inter-class similarity: distinctive between clusters

- The quality of a clustering method depends on
  - the similarity measure used by the method
  - its implementation, and
  - Its ability to discover some or all of the hidden patterns

UNSUPERVISED : K- MEANS CLUSTERING

- Dissimilarity/Similarity metric
  - Similarity is expressed in terms of a distance function, typically metric:

$$d(i, j)$$

  - The definitions of distance functions are usually rather different for interval-scaled, boolean, categorical, ordinal ratio, and vector variables
  - Weights should be associated with different variables based on applications and data semantics
- Quality of clustering:
  - There is usually a separate "quality" function that measures the "goodness" of a cluster.
  - It is hard to define "similar enough" or "good enough"

UNSUPERVISED : K- MEANS CLUSTERING

- Partitioning criteria
  - Single level vs. hierarchical partitioning (often, multi-level hierarchical partitioning is desirable)
- Separation of clusters
  - Exclusive (e.g., one customer belongs to only one region) vs. non-exclusive (e.g., one document may belong to more than one class)
- Similarity measure
  - Distance-based (e.g., Euclidian, road network, vector) vs. connectivity-based (e.g., density or contiguity)
- Clustering space
  - Full space (often when low dimensional) vs. subspaces (often in high-dimensional clustering)

UNSUPERVISED : K- MEANS CLUSTERING

- Given *k*, the *k-means* algorithm is implemented in four steps:
  - Partition objects into *k* nonempty subsets
  - Compute seed points as the centroids of the clusters of the current partitioning (the centroid is the center, i.e., *mean point*, of the cluster)
  - Assign each object to the cluster with the nearest seed point
  - Go back to Step 2, stop when the assignment does not change

THANK-YOU