AIE
ACADEMY OF INNOVATIVE EDUCATION

DAY 03

INTRODUCTION TO ML CLASSIFIERS - IV

DECISION TREE CLASSIFIER

Thakshila Dasun
BSc. Eng in Mechatronics Eng
CIMA (UK)

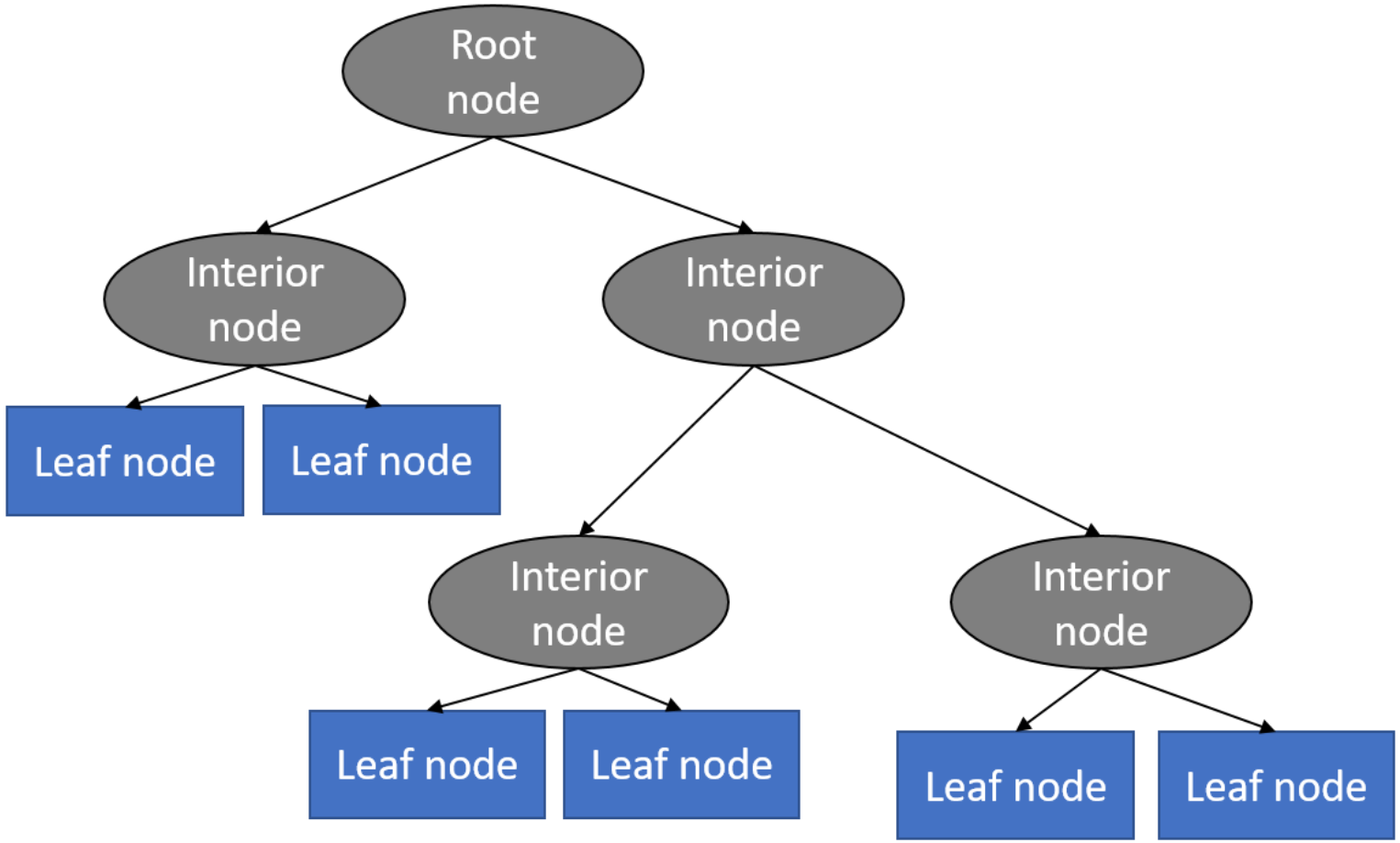
Decision Tree Classifier(1)

- Decision trees are supervised learning algorithms used for both, classification and regression tasks.
- Decision trees are assigned to the information based learning algorithms which use different measures of information gain for learning.
- We can use decision trees for issues where we have continuous but also categorical input and target features
- The main idea of decision trees is to find those descriptive features which contain the most "information" regarding the target feature and then split the dataset along the values of these features such that the target feature values for the resulting sub_datasets are as pure as possible

Decision Tree Classifier(2)

- The descriptive feature which leaves the target feature most purely is said to be the most informative one.
- This process of finding the "most informative" feature is done until we accomplish a stopping criteria where we then finally end up in so called **leaf nodes**. The leaf nodes contain the predictions we will make for new query instances presented to our trained model
- This is possible since the model has kind of learned the underlying structure of the training data and hence can, given some assumptions, make predictions about the target feature value (class) of unseen query instances.
- A decision tree mainly contains of a **root node**, **interior nodes**, and **leaf nodes** which are then connected by **branches**.

Decision Tree Classifier(3)



Entropy

- the information gain is the measure of how good a descriptive feature is suited to split a dataset on.
- To be able to calculate the information gain, we have to first introduce the term *entropy* of a dataset.
- The entropy of a dataset is used to measure the impurity of a dataset and we will use this kind of informativeness measure in our calculations

$$H(x) = -\sum_{for\ k \in target} (P(x = k) * log_2(P(x = k)))$$

Info Gain

$$InfoGain(feature_d) = Entropy(D) - Entropy(feature_d)$$

- How can we check which of the descriptive features most accurately splits the dataset, that is, remains the dataset with the lowest impurity(entropy)
- we use each descriptive feature and split the dataset along the values of these descriptive feature and then calculate the entropy of the dataset once we have split the data along the feature values.
- This gives us the remaining entropy after we have split the dataset along the feature values.
- Next, we subtract this value from the originally calculated entropy of the dataset to see how much this feature splitting reduces the original entropy. The information gain of a feature is calculated with: