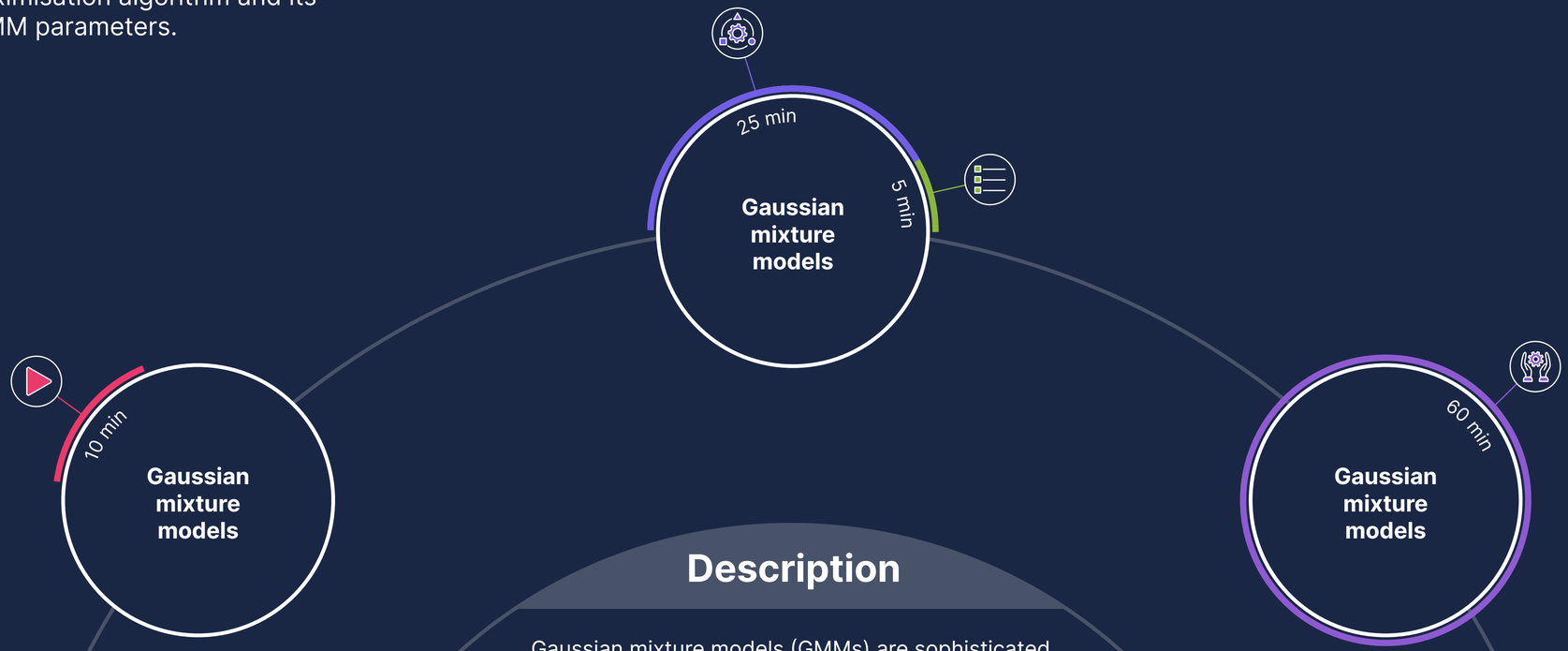
Gaussian mixture models – Lesson overview

Learning objectives:

- Describe the concept of mixture models and how GMMs represent a probabilistic model for clustering.
- Differentiate GMMs from k-means clustering in terms of cluster assignment and flexibility.
- Explain the Expectation-Maximisation algorithm and its importance in optimising GMM parameters.
- Implement GMMs using scikit-learn to perform clustering on real-world datasets.

Video | **Knowledge questions** | **Examples** | **Exercise**



Description

Gaussian mixture models (GMMs) are sophisticated unsupervised machine learning tools, used to uncover **latent structures** within data. Unlike more simplistic clustering techniques, GMMs accommodate **mixed membership** of data points to **multiple clusters**, making them highly effective for **complex data distributions** involving **overlapping clusters** and **non-spherical shapes**.

In this lesson, we'll delve into the **theory and application of Gaussian mixture models**. We'll start by **comparing GMMs with simpler clustering methods** like k-means to highlight their advantages in soft clustering scenarios. Following that, we will explore the **mathematical foundation of GMMs**, including the Expectation-Maximisation (EM) algorithm, which plays a crucial role in estimating the parameters of GMMs.