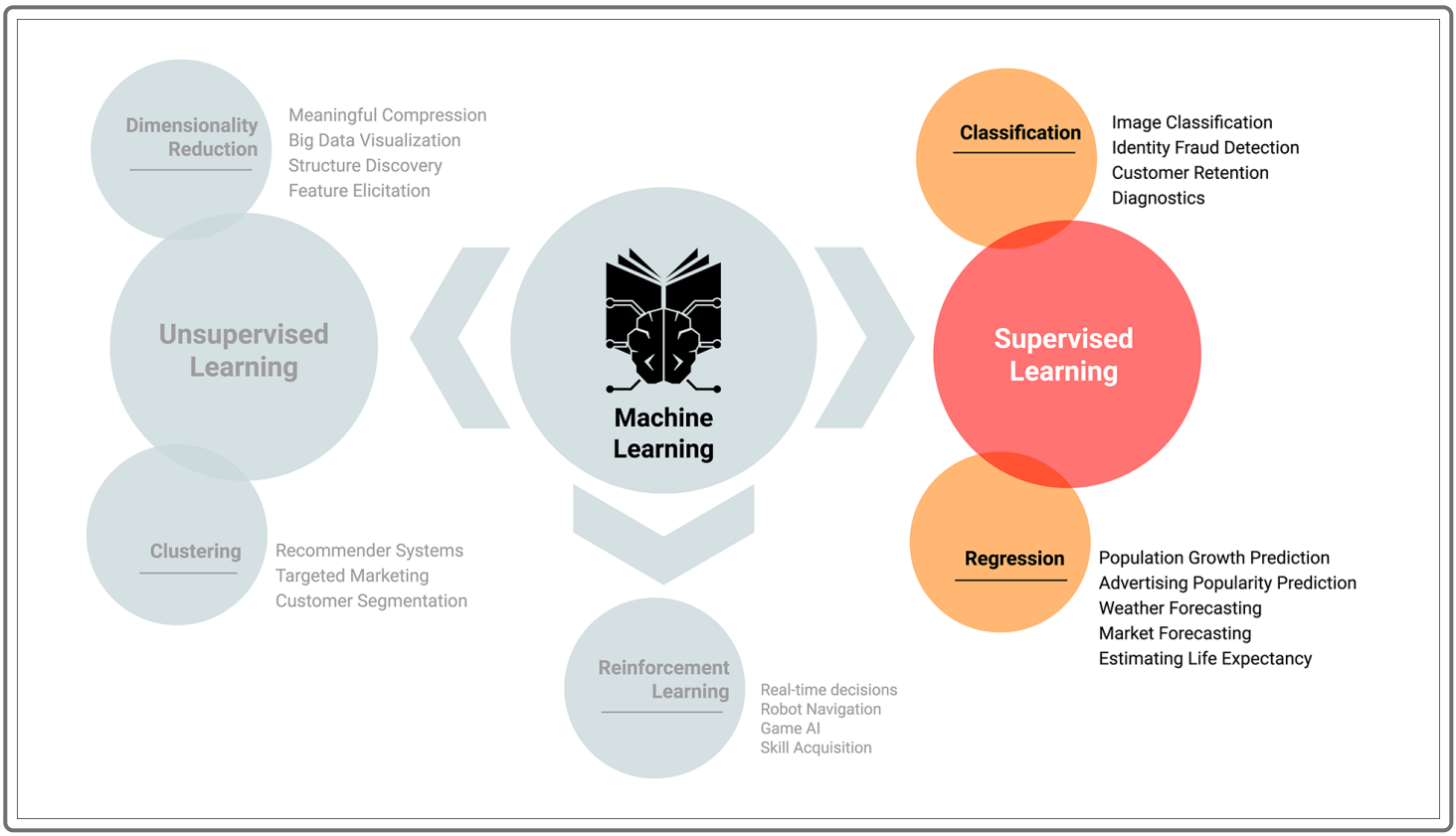
17.2.2

Supervised Learning: Regression and Classification

After getting an overview of supervised learning and unsupervised learning, you would like to find out more about supervised learning. Jill, your data scientist boss, asks you to research the difference between regression and classification models, then report back to her. Here is what you find out.

Supervised learning can be broadly divided into regression and classification. In this module, we will focus mainly on using supervised machine learning for classification:



Regression

Recall that **regression** is used to predict continuous variables. For example, let's say that we're interested in predicting a person's weight based on factors like height, dietary preferences, and exercise patterns. To accomplish this task, we would collect data on a number of people. The regression model's algorithms would attempt to learn patterns that exist among these factors. If presented with the data of a new person, the model would make a prediction of that person's weight, based on previously learned patterns from the dataset.

REWIND

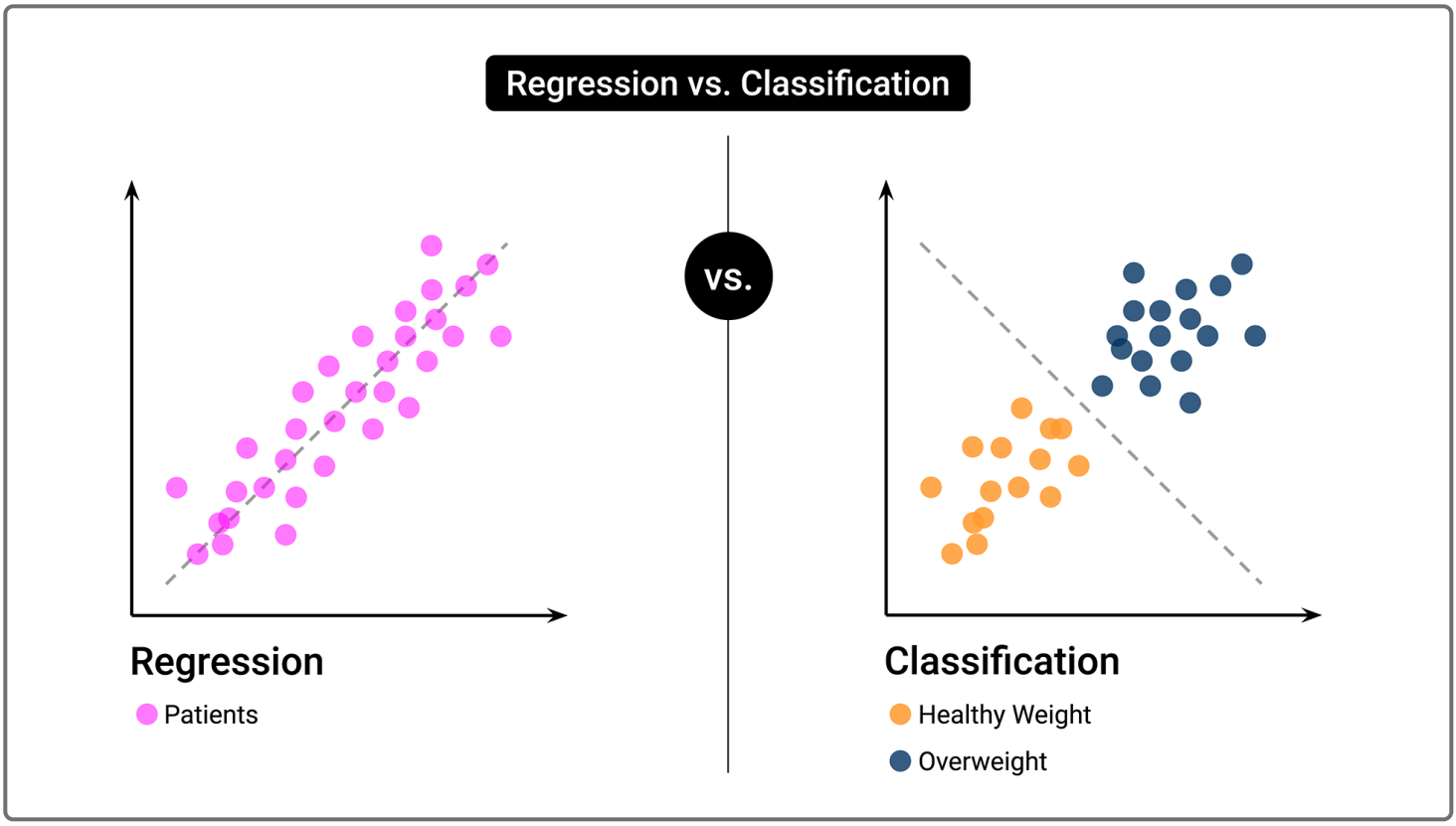
Regression is used to predict continuous variables.

Classification

Classification, on the other hand, is used to predict discrete outcomes. For example. Let's say that we are interested in using a person's traits, such as age, sex, income, and geographic location, to predict how she or he will vote on a particular issue. The outcome, in this case, is whether the person will vote "Yes" or "No." The classification model's algorithms would attempt to learn patterns from the data, and if the model is successful, gain the ability to make accurate predictions for new voters.

Regression vs. Classification

There is a major difference between regression and classification models. In our regression example, the target variable, or what we're trying to predict, is weight. Weight is a continuous variable—a person's weight can be any numerical value within a certain range. In our classification example, on the other hand, the target variable only has two possible values: whether a person votes "Yes" or "No." When the classification model encounters new data, it would attempt to predict whether the votes cast by people will be "Yes" or "No."



NOTE

In both classification and regression problems, a dataset is divided into features and target. **Features** are the variables used to make a prediction. **Target** is the predicted outcome.

Retake

In the examples above, we saw a basic pattern:

- A machine learning model is presented with a dataset.
- The model algorithms analyze the data and attempt to identify patterns.
- Based on these patterns, the model makes predictions on new data.

This pattern applies whether we're using regression or classification. Let's look at an example of creating and using a machine learning model.