

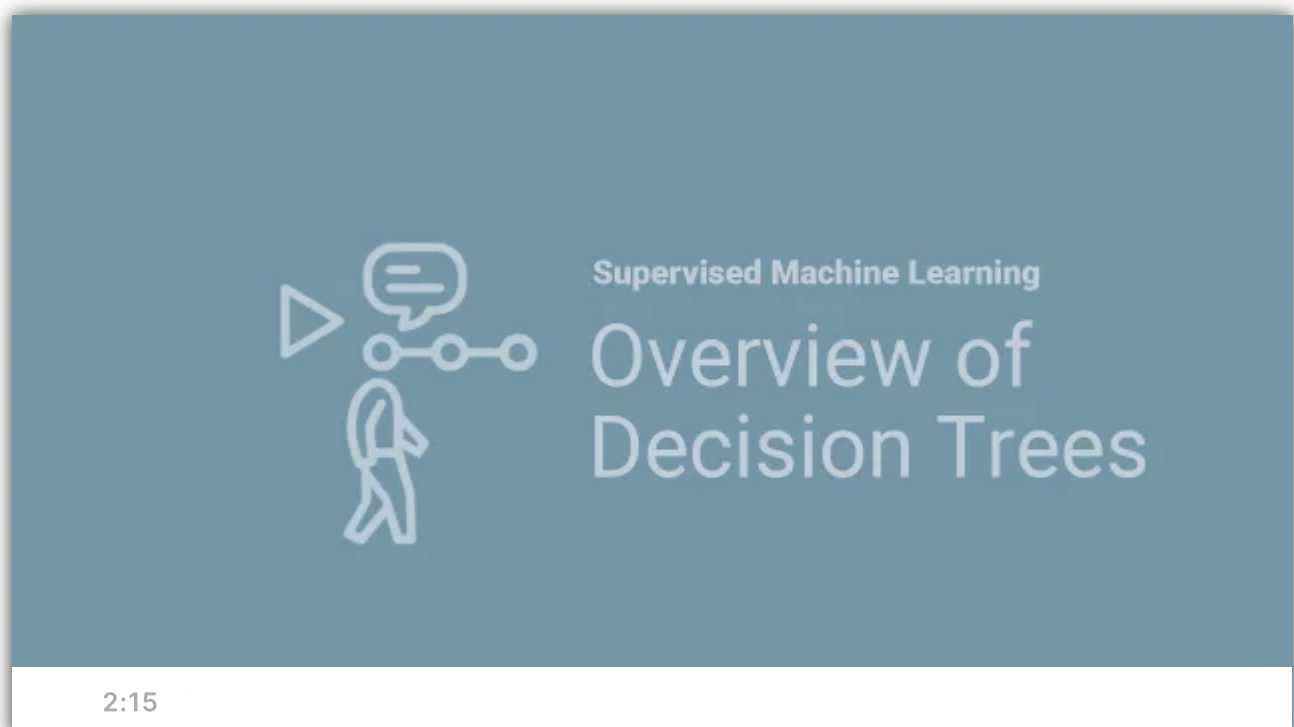
## 17.7.1

## Overview of Decision Trees

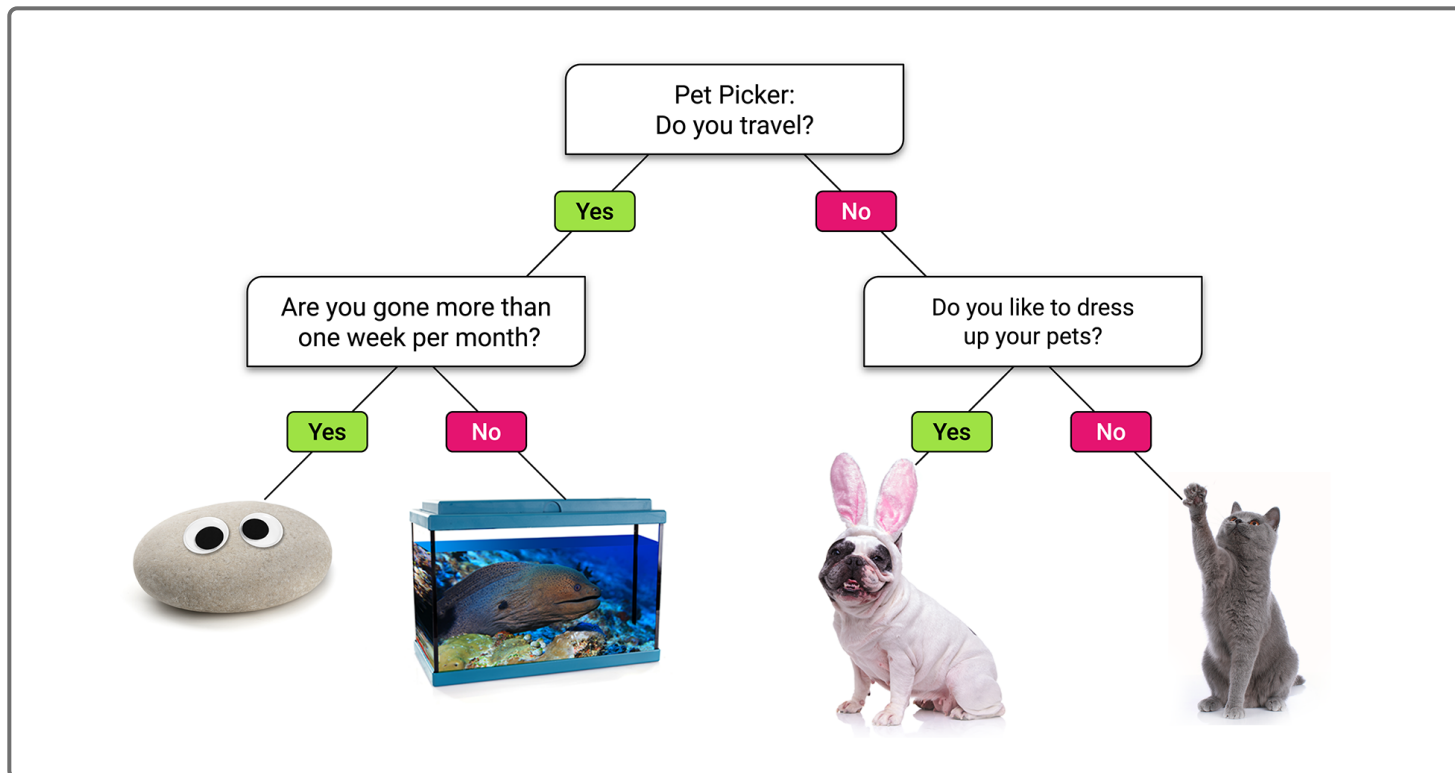
**You're** becoming more and more comfortable with supervised learning, and Jill suggests that you look into decision trees. The basic idea behind decision trees is simple. Furthermore, decision trees can be combined into even more powerful classifiers. In fact, decision trees are the basis of many models that win machine learning competitions.

As you may have guessed, decision trees are used in decision analysis, like determining if a coin flip will be heads or tails, or whether a loan application is approved. In short, decision trees encode a series of true/false questions that are represented by a series of if/else statements. Decision trees are one of the most interpretable models, as they provide a clear representation of how the model works.

Decision trees are natural ways in which you can classify or label objects by asking a series of questions designed to zero in on the true answer. However, decision trees can become very complex and very deep, depending on how many questions have to be answered. Deep and complex trees tend to overfit to the data and do not generalize well.



The following example, a pet picker decision tree, covers some key concepts:

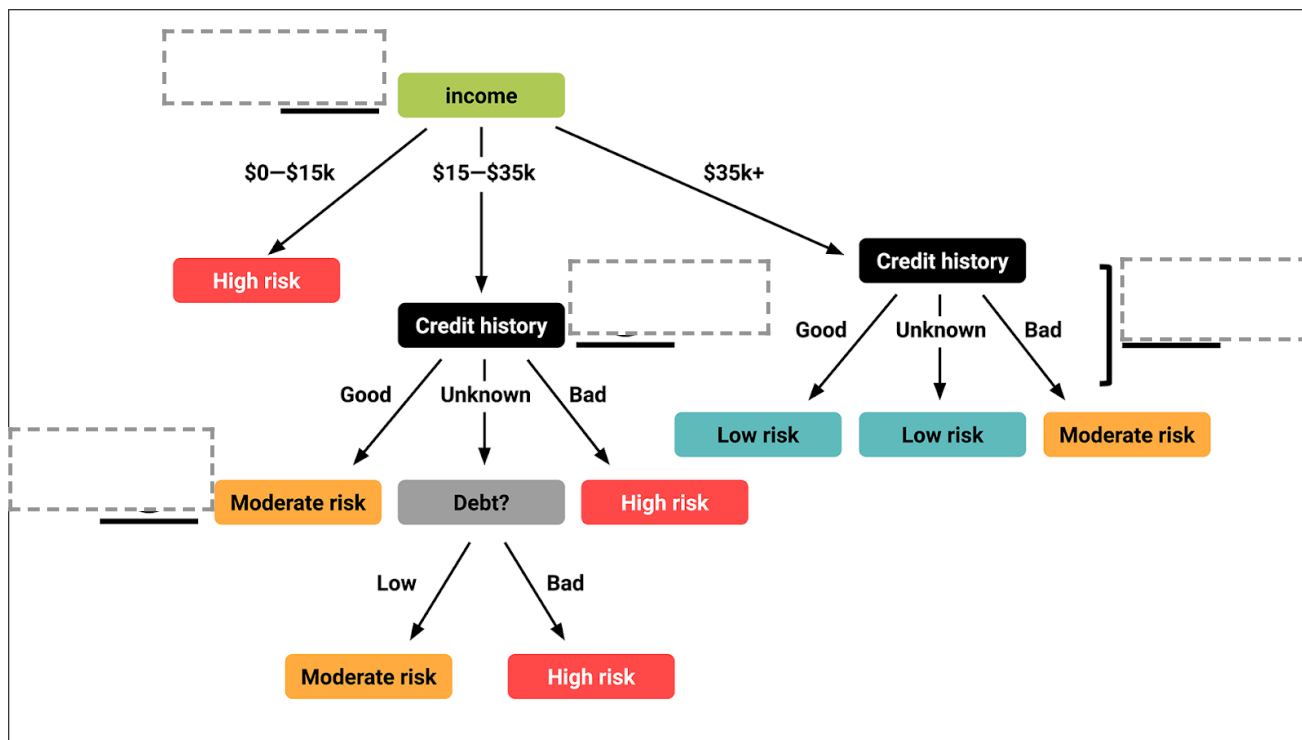


At the top of this decision tree is the **root node** or the **parent node**: "Pet Picker: Do you travel?" The root node represents the entire population. This node gets divided into two or more homogeneous sets in our decision tree to answer the question, "Do you travel?" Each answer, "Yes" or "No," is a **branch** or subsection of the tree. When we divide a node into two or more subnodes, it's called **splitting**.

When we split the root node into two subnodes, they are called **child nodes**. Our child nodes include two questions: "Are you gone more than one week per month?" and "Do you like to dress up your pets?" These child nodes are split again into subnodes—the images in the decision tree—making each child node a **decision node**. The four images do not split further and are referred to as a **leaf** or **terminal node**.

Our pet picker decision **tree depth** is not that deep because, at two, it has a low number of decision nodes one encounters before making a decision: "Are you gone more than one week per month?" and "Do you like to dress up your pets?"

Label the parts of the following credit approval decision tree using the following terms: child node, leaf node, root node, and splitting.



child node

leaf node

splitting

root node

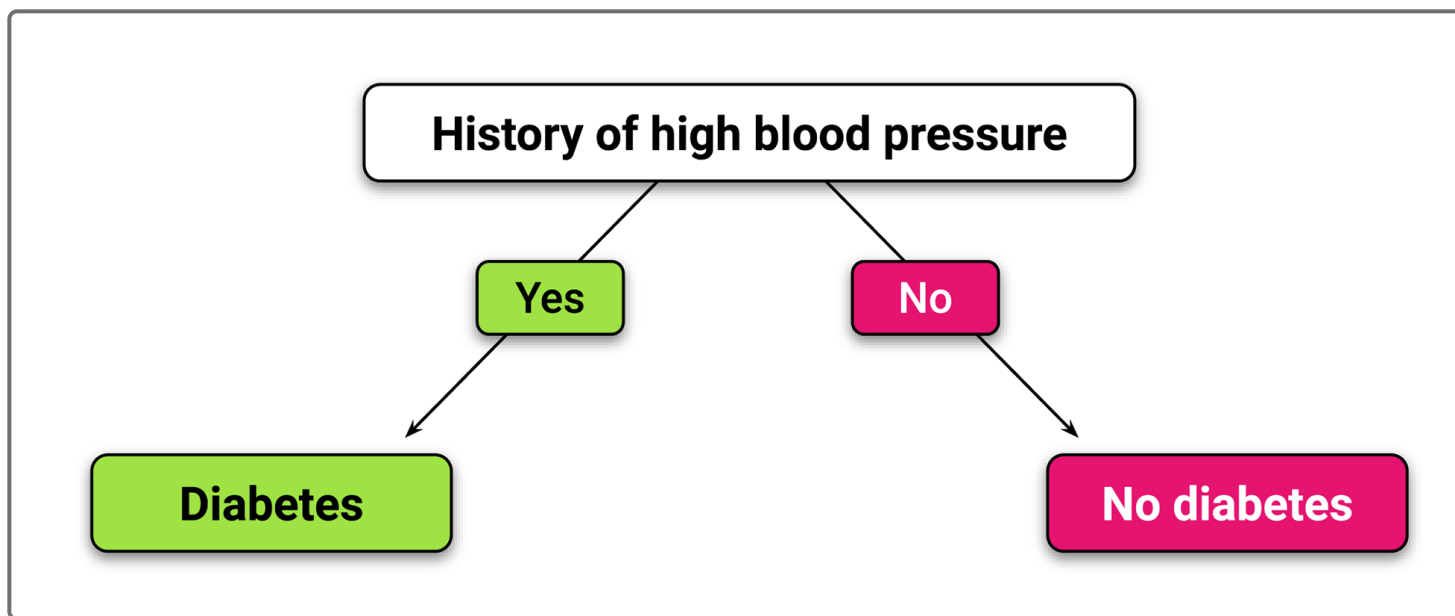
Check Answer

Finish ►

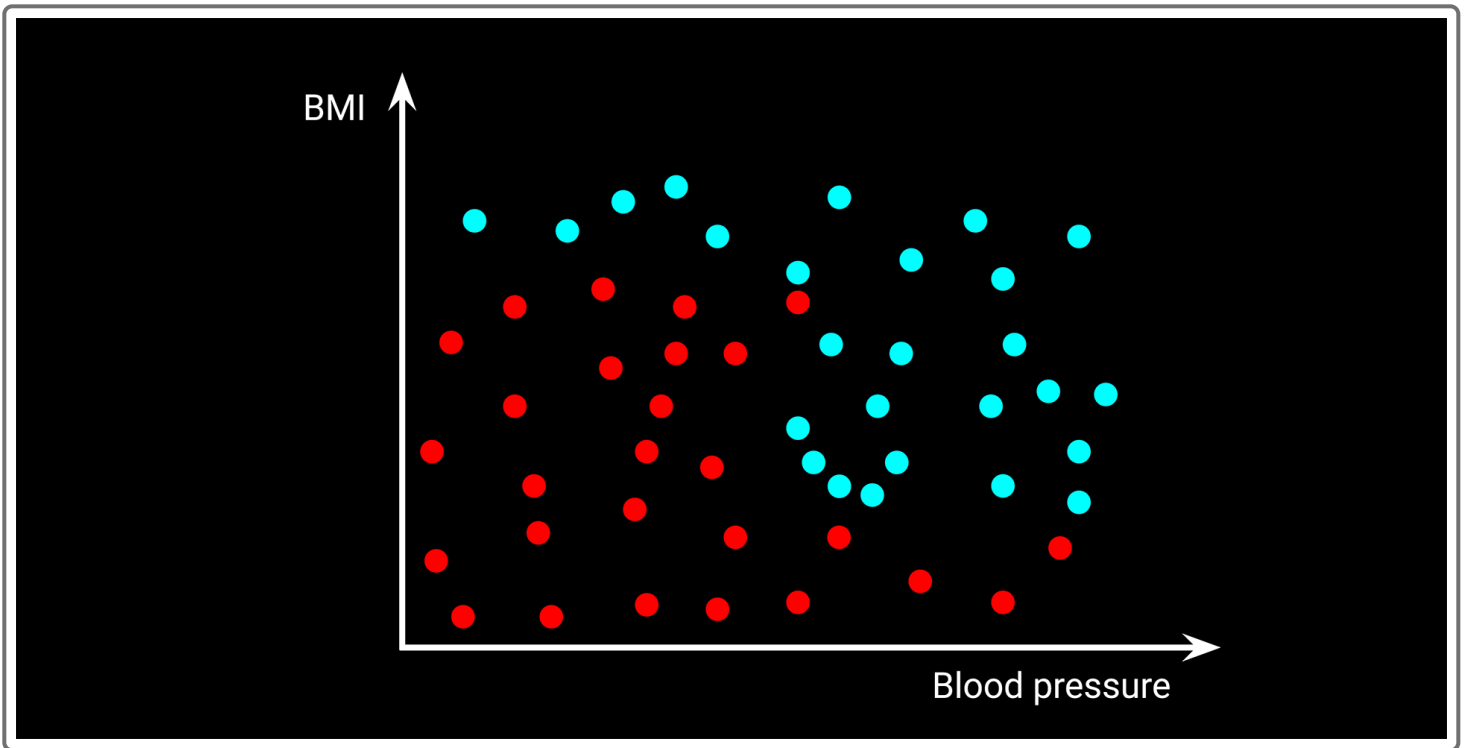
Decision trees, as you have seen, are like a series of if-else statements. But how does a decision tree determine which if-else statement to use first? In other words, how does it decide on the root node?

To understand this process, let's imagine a scenario in which we use multiple factors, such as body mass index (BMI) over 30, and a history of high blood pressure, to predict whether a patient has diabetes. There are two candidates for the root node in this example: BMI and blood pressure. The classification that creates the best split becomes the root node.

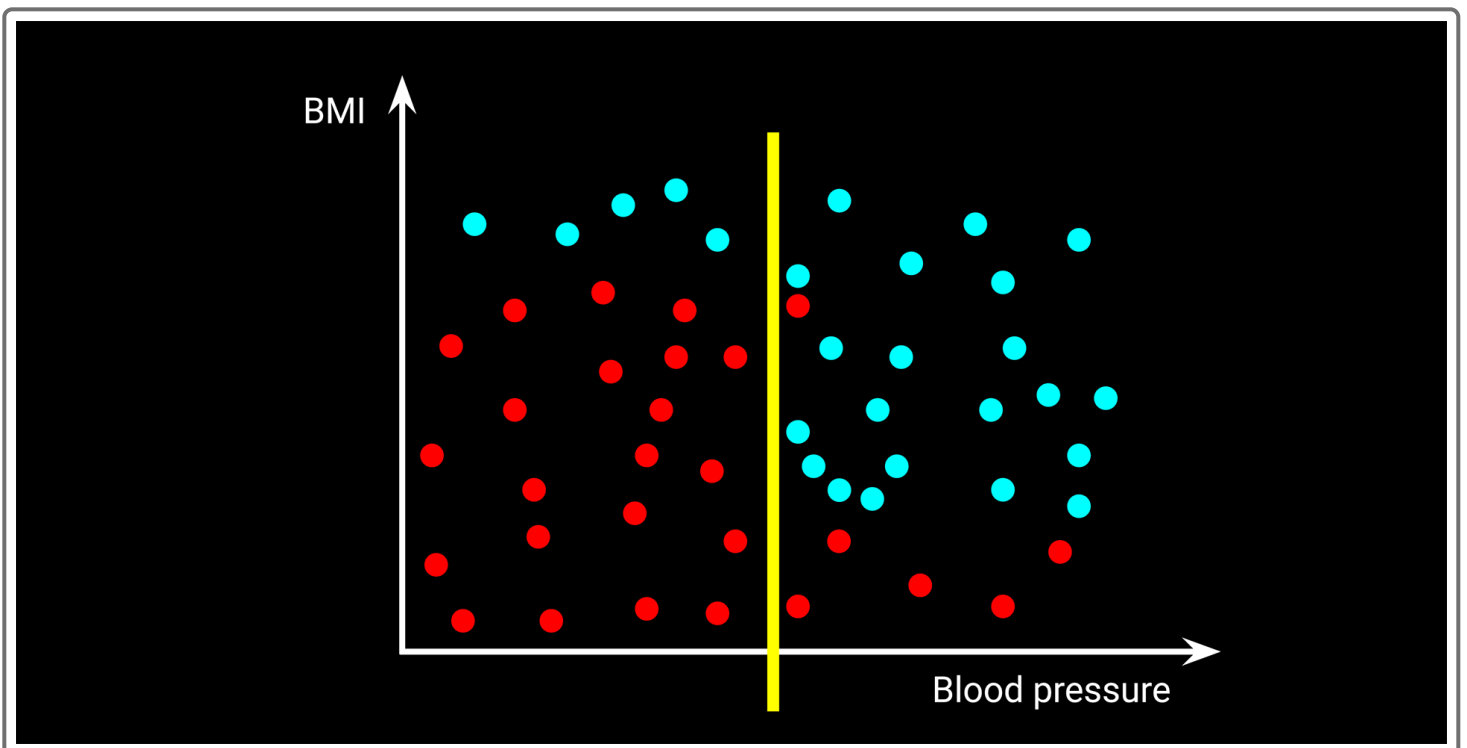
For example, let's say that a history of hypertension best predicts whether a person will be diabetic. In that case, it becomes the root node:



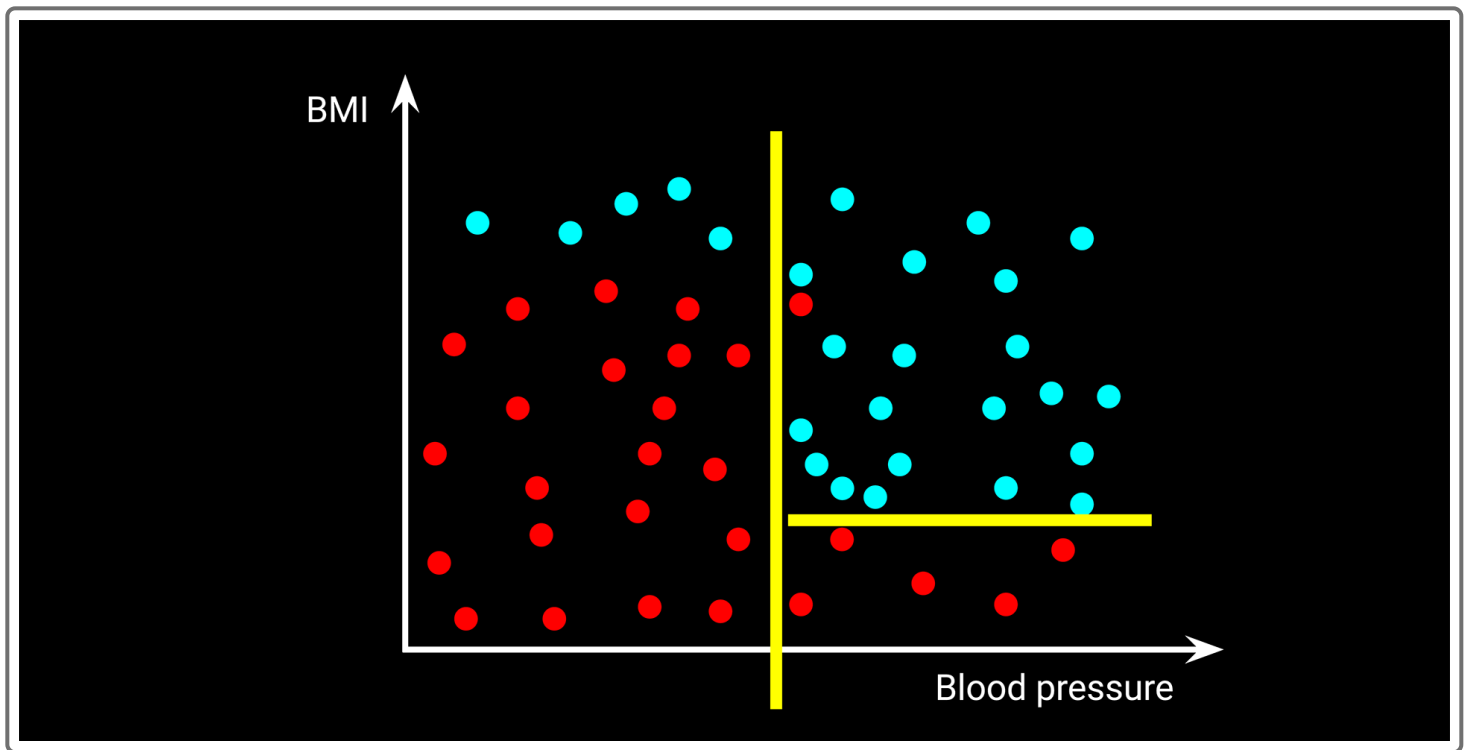
The same idea holds when the variables are continuous rather than discrete. Let's say that we're dealing with the same two variables, BMI and blood pressure, to predict diabetes. The decision tree looks for the best value along each axis to split the diabetics from non-diabetics:



Looking at this image, we can see that a vertical line along the x-axis (blood pressure) would best split the dataset into diabetics and non-diabetics, resulting in the fewest errors. In other words, a specific blood pressure value can be used to split the dataset into diabetic and non diabetic groups:



The same process is repeated to create subsequent child nodes. In this example, a horizontal line that represents a particular BMI will become the next node:



© 2020 - 2022 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.