17.9.2

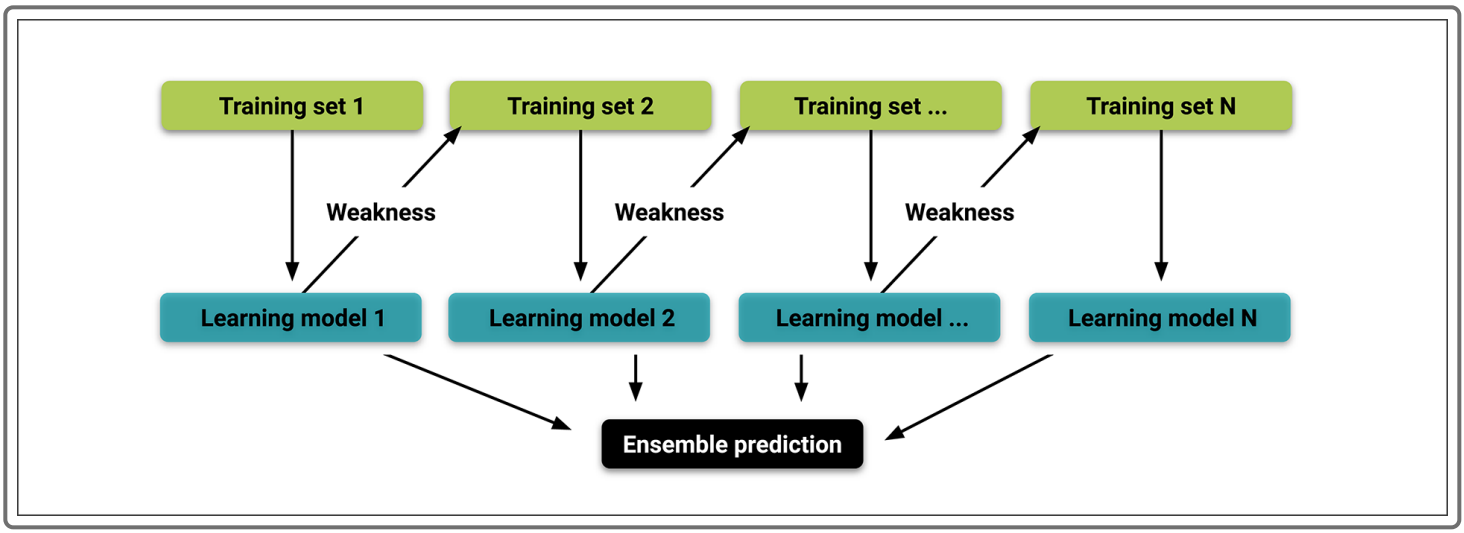
Boosting

Boosting is another technique to combine weak learners into a strong learner. However, there is a major difference between bagging and boosting. In bagging, as you have seen, multiple weak learners are combined at the same time to arrive at a combined result.

In boosting, however, the weak learners are not combined at the same time. Instead, they are used sequentially, as one model learns from the mistakes of the previous model.

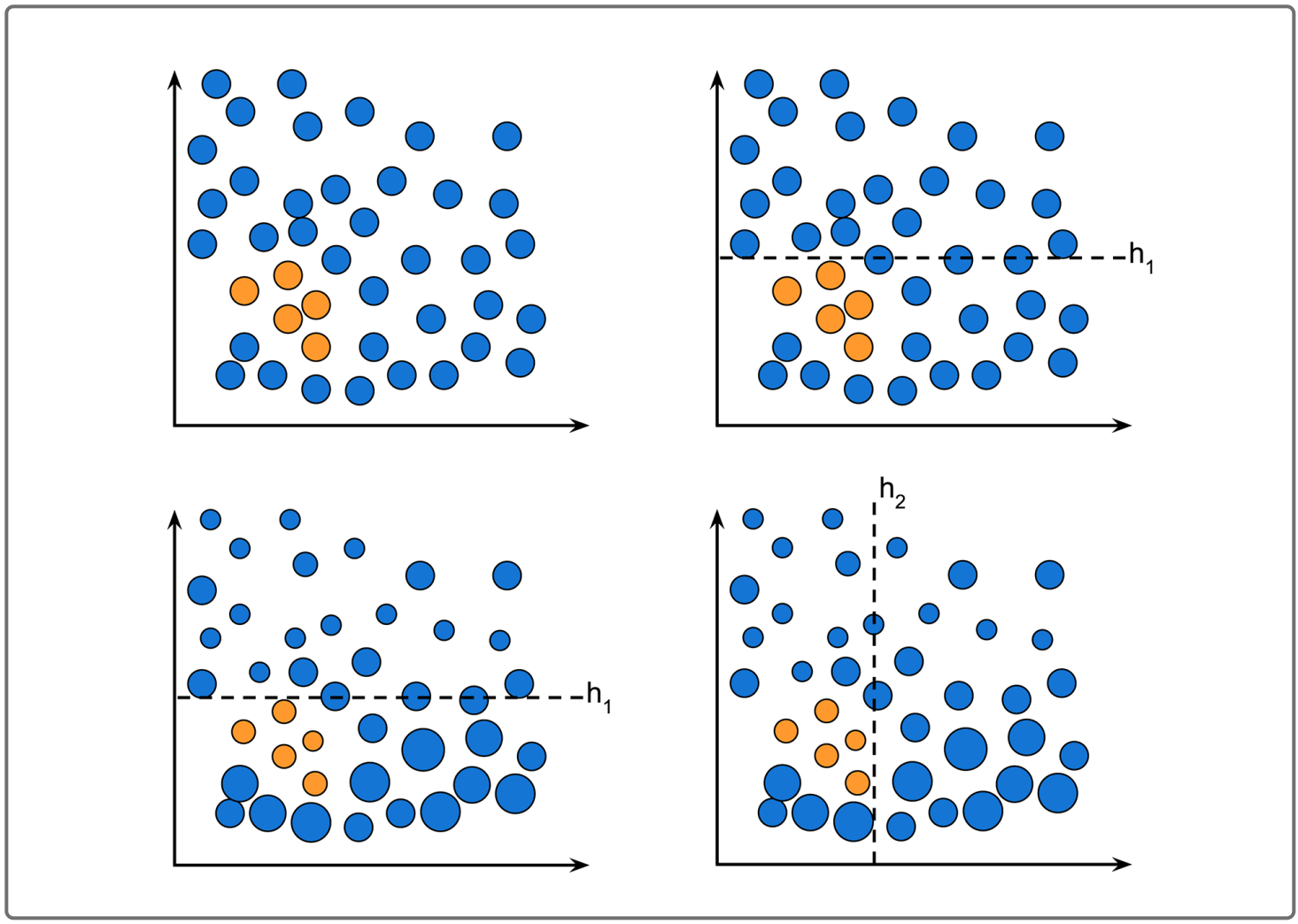
Jill assures you that learning this ensemble learning technique will be worth your time. After all, many machine learning competitions have been won with this powerful technique.

Like bagging, boosting is also a technique to combine a set of weak learners into a strong learner. We saw in bagging that the different models work independently of one another. In contrast, boosting trains a sequence of weak models. As shown below, each model learns from the errors of the previous model, and the models form an ensemble:

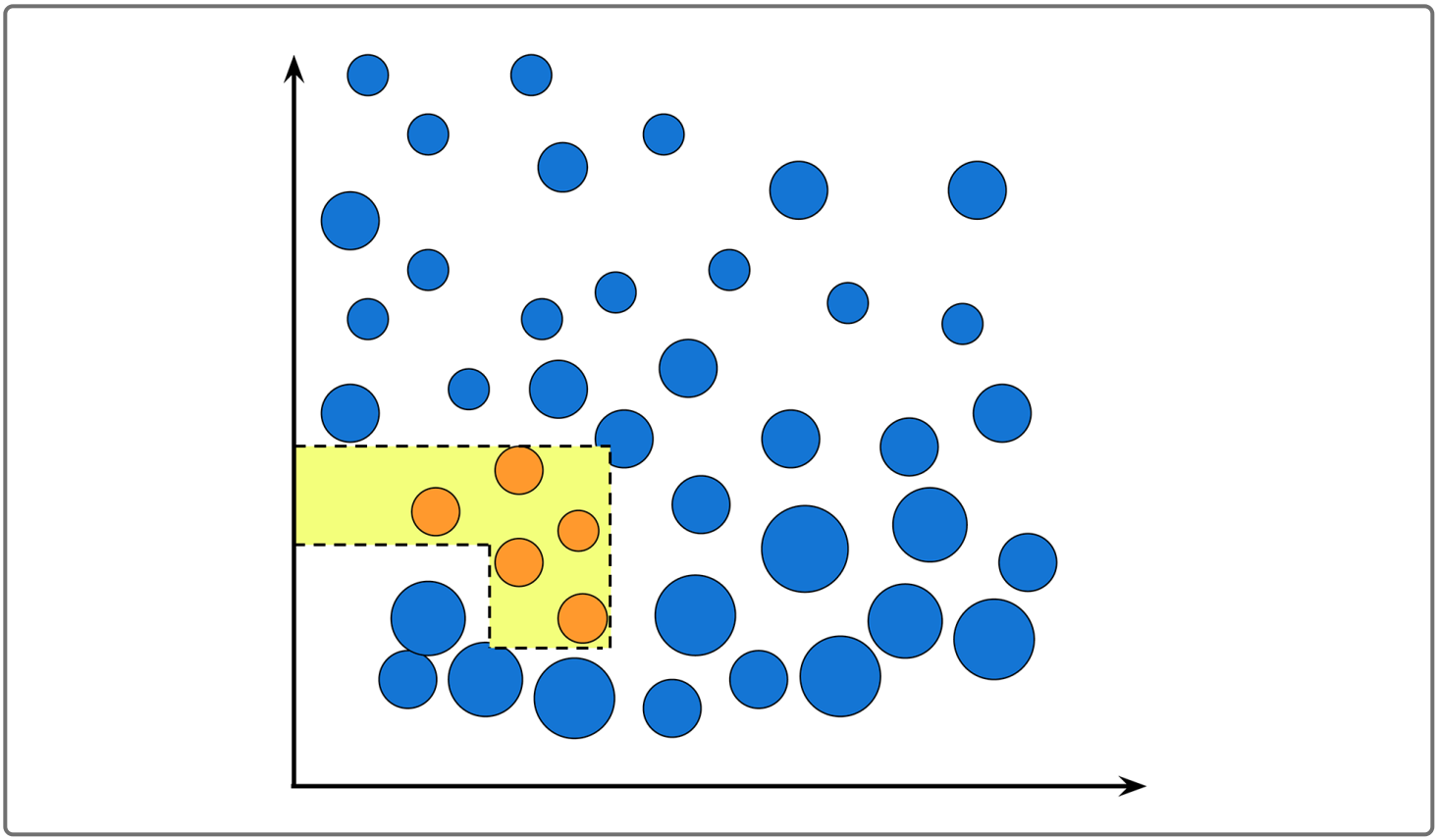


Adaptive Boosting

The idea behind Adaptive Boosting, called AdaBoost, is easy to understand. In AdaBoost, a model is trained then evaluated. After evaluating the errors of the first model, another model is trained. This time, however, the model gives extra weight to the errors from the previous model. The purpose of this weighting is to minimize similar errors in subsequent models. Then, the errors from the second model are given extra weight for the third model. This process is repeated until the error rate is minimized:



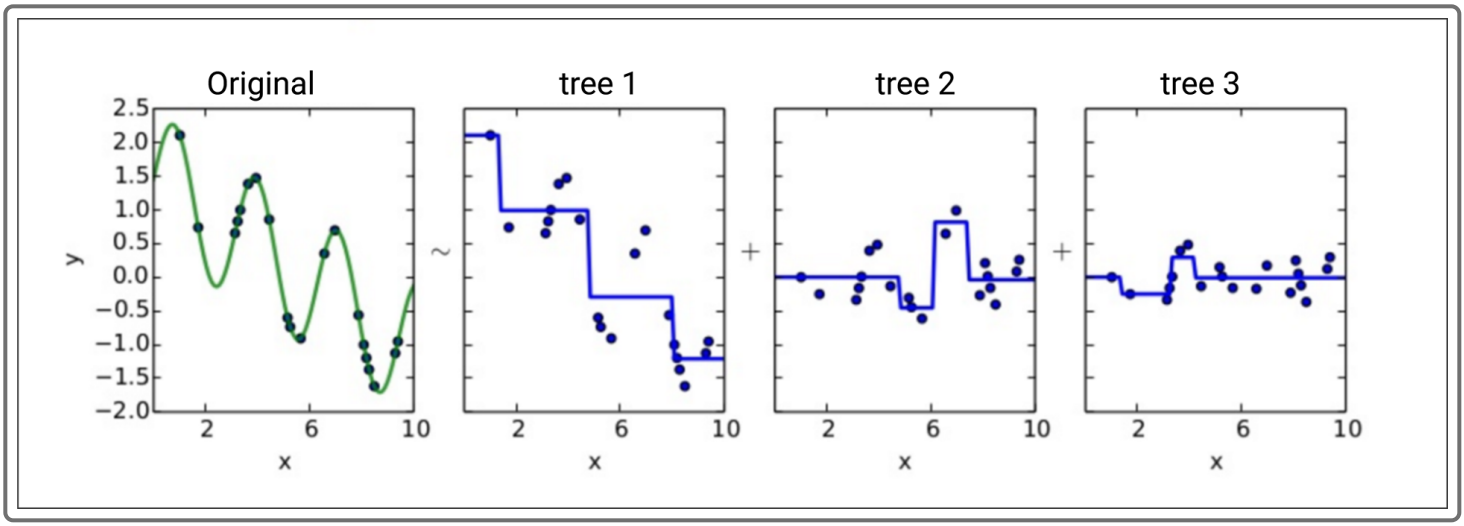
The final classifier might appear like the following:



Gradient Boosting

Gradient boosting, like AdaBoost, is an ensemble method that works sequentially. In contrast to AdaBoost, gradient boosting does not seek to minimize errors by adjusting the weight of the errors. Instead, it follows this process:

1. A small tree (called a stump) is added to the model, and the errors are evaluated.
2. A second stump is added to the first and attempts to minimize the errors from the first stump. These errors are called pseudo-residuals.
3. A third stump is added to the first two and attempts to minimize the pseudo-residuals from the previous two.
4. The process is repeated until the errors are minimized as much as possible, or until a specified number of repetitions has been reached:



In gradient boosting, the learning rate refers to how aggressively pseudo-residuals are corrected during each iteration. In general, it is preferable to begin with a lower learning rate and, if necessary, adjust the rate upward.

Gradient boosting is a powerful technique that is often used in machine learning competitions.