

Applied Data Mining Homework 2

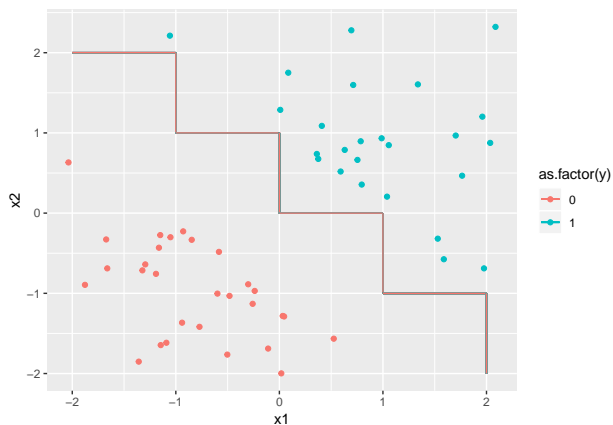
Xun Zhao, xz2827

Problem 1: Trees

1.

Yes. But it needs the tree to be very high than some simple tree classifiers.

As the tree classifier boundary is made of some line segments that are parallel with x-axis or y-axis, we can draw a zigzag curve along the linear boundary. Thus, the sloping linear boundary becomes many segments, and every segment is a decision in tree classifier.



2.

No, the Bayes-Optimal is defined as:

$$f(\vec{x}) = \underset{y}{\operatorname{arg\,max}} P(\mathbf{Y} = y | \mathbf{X} = \vec{x})$$

And the risk is defined as:

$$R(f) = \sum_{+1, -1} \int L(y, f(\vec{x})) P(\vec{x}, y) d\vec{x}$$