Applied Data Mining Homework 2

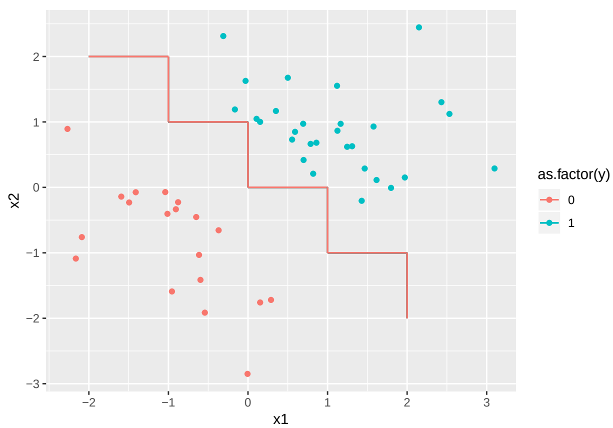
Xun Zhao, xz2827

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 dicision in tree classifier.



2.

No, the Bayes-Optimal is defined as:

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

And the risk is defined as:

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

Thus, the risk $R(f)$ depends on the possible data distribution instead of only existing data points. Moreover, the Bayes-optimal classifier is defined by the possible distribution, so it can always minimize the risk under certain data model. However, the tree classifier cannot fit the linear boundary, which causes the misclassification compared with Bayes-optimal.

As the Bayes-optimal has the lowest risk, the risk of tree classifier will always differ from Bayes-optimal, unless the Bayes-optimal's boundary is axis-parallel.

3.

$$g(\vec{x}) = \begin{cases} f_1(\vec{x}) > 0 & \begin{cases} f_2(\vec{x}) > 0 & class\ A \\ f_2(\vec{x}) < 0 & class\ B \end{cases} \\ f_1(\vec{x}) < 0 & \begin{cases} f_3(\vec{x}) > 0 & class\ B \\ f_3(\vec{x}) < 0 & class\ C \end{cases} \end{cases}$$