Machine Learning Exercise Sheet 1

Math Refresher

Group_369

Fan Xue – fan98.xue@tum.de

Xing Zhou – xing.zhou@tum.de

Jianzhe Liu – jianzhe.liu@tum.de

November 2, 2021

Problem 6

Yes. We can define two new features $y_1 = x_1 - x_2$ and $y_2 = x_2$. By doing this, we can split the dataset by judging the condition $y_1 \leqslant 0$. With only one split the dataset is split into two parts and each part only has one sort of class. It means, there exists a decision tree of depth 1 that classifies this dataset with 100% accuracy.

Problem 7

a)

$$\begin{aligned} i_H(y) &= -p(y=W)\log p(y=W) - p(y=L)\log p(y=L) \\ &= -\frac{4}{10}\log\frac{4}{10} - \frac{6}{10}\log\frac{6}{10} \\ &= 0.971 \end{aligned}$$

b) Spliting by $x_1 = T$

$$\begin{aligned} \Delta i_H &= i_H(y) - p(x_1=T)i_H(x_1=T) - p(x_1=I)i_H(x_1=I) \\ &= 0.971 - \frac{1}{2}(-\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\log\frac{3}{5}) - \frac{1}{2}(-\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\log\frac{3}{5}) \\ &= 0 \end{aligned}$$

Spliting by $x_2 = M$

$$\begin{aligned} \Delta i_H &= i_H(y) - p(x_2=M)i_H(x_1=M) - p(x_2=P)i_H(x_2=P) \\ &= 0.971 - \frac{4}{10}(-\frac{2}{4}\log\frac{2}{4} - \frac{2}{4}\log\frac{2}{4}) - \frac{6}{10}(-\frac{2}{6}\log\frac{2}{6} - \frac{4}{6}\log\frac{4}{6}) \\ &= 0.020 \end{aligned}$$

Spliting by $x_3 = S$

$$\begin{aligned}\Delta i_H &= i_H(y) - p(x_3 = S) i_H(x_3 = S) - p(x_3 = C) i_H(x_3 = C) \\ &= 0.971 - \frac{1}{2}(-\frac{3}{5}\log\frac{3}{5} - \frac{2}{5}\log\frac{2}{5}) - \frac{1}{2}(-\frac{1}{5}\log\frac{1}{5} - \frac{4}{5}\log\frac{4}{5}) \\ &= 0.125\end{aligned}$$

According to the calculation the split judgement will be $x_3 = S$, since in this case, the Δi_H is the biggest. If $x_3 = S$, the instance will be classified as W. Otherwise it will be classified as L.

Problem 8

Let $i^2 = \frac{125}{i}$, we get $i = 5$. Figure 1 shows the 2-d space of the dataset.

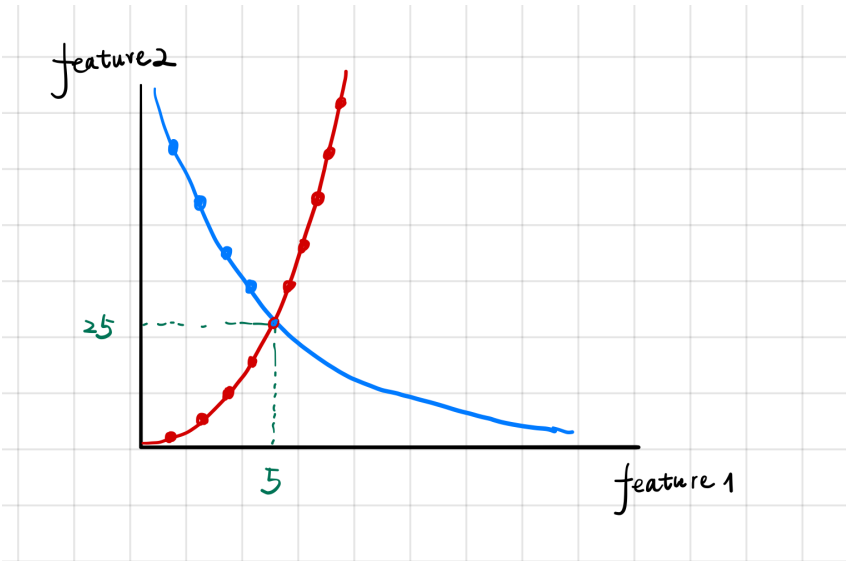

Figure 1: the 2-d space of the dataset

We can easily split the dateset into 4 parts. The first split uses the threshold feaure2 $\leqslant 25$. The second split uses the threshold feature1 $\leqslant 5$ for both child nodes.
In this way, the depth of the decision tree is 2. Only one datapoint (5, 25) is missclassified, which is unavoidable.