Machine Learning Lab Class III

Ankit Satpute, 120825
Hsueh Wei, 120820
Sagar Nagaraj Simha, 120797
(M.Sc. CS4DM)

09/12/2019

Exercise 2.a

D :- a set of examples over a feature space X
Set of classes C = $\{c_1, c_2, c_3, c_4\}$, with D= 24 and two possible decision trees

Considering only first split each of the two trees makes, we have to compute $\Delta\iota(\{D_{1,1}, D_{1,2}\})$ and $\Delta\iota(\{D_{2,1}, D_{2,2}\})$ with the misclassification rate $\iota_{misclass}$ and the entropy criterion $\iota_{entropy}$ as splitting criterion.

We will use formulas given on the lecture slides to calculate $\Delta\iota_{misclass}$ and $\Delta\iota_{entropy}$ for both the trees respectively.

- **For tree t_1**

  $\Delta\iota_{misclass} = \iota_{misclass}(D) - \left\{\frac{|D_{1,1}|}{|D|} * \iota_{misclass}(D_{1,1}) + \frac{|D_{1,2}|}{|D|} * \iota_{misclass}(D_{1,2})\right\}$

  we calculate values needed to put in the above formula

  $\iota_{misclass}(D) = 1 - \max\left\{\frac{6}{24}, \frac{6}{24}, \frac{6}{24}, \frac{6}{24}\right\} = \frac{3}{4}$

  $\iota_{misclass}(D_{1,1}) = 1 - \max\left\{\frac{6}{12}, \frac{6}{12}\right\} = \frac{1}{2}$

  $\iota_{misclass}(D_{1,2}) = 1 - \max\left\{\frac{6}{12}, \frac{6}{12}\right\} = \frac{1}{2}$

  Now substituting all values we get

  - $\Delta\iota_{misclass} = \frac{3}{4} - \left\{\frac{12}{24} * \frac{1}{2} + \frac{12}{24} * \frac{1}{2}\right\} = \frac{1}{4}$

  $\Delta\iota_{entropy} = \iota_{entropy}(D) - \left\{\frac{|D_{1,1}|}{|D|} * \iota_{entropy}(D_{1,1}) + \frac{|D_{1,2}|}{|D|} * \iota_{entropy}(D_{1,2})\right\}$

  Again we calculate values needed to put in the above formula

  $\iota_{entropy}(D) = -\frac{6}{24} * \log_2 \frac{6}{24} * 4 = 2$

  $\iota_{entropy}(D_{1,1}) = -\left(\frac{6}{12} * \log_2 \frac{6}{12} + \frac{6}{12} * \log_2 \frac{6}{12}\right) = 1$

  $\iota_{entropy}(D_{1,2}) = -\left(\frac{6}{12} * \log_2 \frac{6}{12} + \frac{6}{12} * \log_2 \frac{6}{12}\right) = 1$

  Now substituting all values we get

- $\Delta\iota_{entropy} = 2 - \left(\frac{12}{24} * 1 + \frac{12}{24} * 1\right) = 1$

- **For tree** t_2

  $\Delta\iota_{misclass} = \iota_{misclass}(D) - \left\{\frac{|D_{2,1}|}{|D|} * \iota_{misclass}(D_{2,1}) + \frac{|D_{2,2}|}{|D|} * \iota_{misclass}(D_{2,2})\right\}$

  we calculate values needed to put in the above formula

  $\iota_{misclass}(D) = 1 - \max\left\{\frac{6}{24}, \frac{6}{24}, \frac{6}{24}, \frac{6}{24}\right\} = \frac{3}{4}$

  $\iota_{misclass}(D_{2,1}) = 1$ - max $\left\{\frac{4}{12}, \frac{2}{12}, \frac{6}{12}\right\} = \frac{1}{2}$

  $\iota_{misclass}(D_{2,2}) = 1$ - max $\left\{\frac{4}{12}, \frac{2}{12}, \frac{6}{12}\right\} = \frac{1}{2}$

  Now substituting all values we get

  - $\Delta\iota_{misclass} = \frac{3}{4} - \left\{\frac{1}{2} * \frac{1}{2} + \frac{1}{2} * \frac{1}{2}\right\} = \frac{1}{4}$

  $\Delta\iota_{entropy} = \iota_{entropy}(D) - \left\{\frac{|D_{2,1}|}{|D|} * \iota_{entropy}(D_{2,1}) + \frac{|D_{2,2}|}{|D|} * \iota_{entropy}(D_{2,2})\right\}$

  Again we calculate values needed to put in the above formula

  $\iota_{entropy}(D) = -\frac{6}{24} * \log_2 \frac{6}{24} * 4 = 2$

  $\iota_{entropy}(D_{2,1}) = -\left(\frac{1}{3} * \log_2 \frac{1}{3} + \frac{1}{6} * \log_2 \frac{1}{6} + \frac{1}{2} * \log_2 \frac{1}{2}\right) = 1.457$

  $\iota_{entropy}(D_{2,2}) = -\left(\frac{1}{3} * \log_2 \frac{1}{3} + \frac{1}{6} * \log_2 \frac{1}{6} + \frac{1}{2} * \log_2 \frac{1}{2}\right) = 1.457$

  Now substituting all values we get

  - $\Delta\iota_{entropy} = 2 - \left(\frac{12}{24} * 1.457 + \frac{12}{24} * 1.457\right) = 0.543$

  From the above values, we can see that miss-classification rate for both the trees t_1 and t_2 is same. Hence, we cannot differentiate on basis of this factor but in case of impurity function based on entropy we know that $\Delta\iota_{entropy}$ corresponds to the information gain. On basis of our calculations this information gain is clearly greater in tree t_1 than in t_2. Therefore, we can say that split $\{D_{1,1}, D_{1,2}\}$ is better first split than $\{D_{2,1}, D_{2,2}\}$

Exercise 2.b

By looking at trees t_1 and t_2 we can see that splitting in tree t_1 constantly minimizes the impurity of subsets of D than in t_2. Although in case of tree t_2 the impurity at each level is also reduced in some extent but looking at the leaf node of t_1 we can clearly conclude that two classes has occupied separate leaf node. hence further partitioning will not be done in those two (c_3, c_4) classes. Based on this observations, we can say that t_1 is better decision tree than t_2

Exercise 2.c

In case of ID3 algorithm, each splitting is based on one nominal feature and considers its complete domain. Also, splitting criterion is information gain i.e. in our case we calculated in terms of $\Delta\iota_{entropy}$. Based on this information and given two trees as only possible combination, the ID3 algorithm should construct tree t_1. Although our observation is based on only first splitting which we calculated in 2.a, with provided splitting criteria if we proceed further then we should get same leaf nodes as in tree t_1.

References

1 Tom Mitchell. Machine Learning. 1st edition, McGraw-Hill, 1997.