CSE 446: Machine Learning
Winter 2018

Assignment 1

from
Lukas Nies
University of Washington

01/18/18

Contents

0 Policies

0.1 List of Collaborators

0.2 List of Acknowledgments

0.3 RTFM

I have read and understood these policies.

1 Problem: Criteria for Choosing a Feature to Split

We build a tree with the dataset D consisting n negative examples (label 0) and p positive examples (label 1).

1.1 Not Splitting

If we are at the bottom of our decision tree and don't have any features left to split, consider the subset D' of data D with n' negative and p' positive examples. The smallest number of mistakes we can make in this subset is given by

$$\mathrm{err}(D') = \min(n', p') = \begin{cases} p' & \text{if } p' < n' \\ n' & \text{if } n' < p' \end{cases} \tag{1}$$

The node itself is labeled accordingly, with 0 if $n' < p'$ or with 1 if $n' > p'$.

1.2 Splitting

Now we have a new feature Φ which splits the subsection D' according to the contingency table: By splitting we generate two new sub-nodes: (n_0, p_0) and (n_1, p_1).

Table 1: Add caption

y	phi	
	0	1
0	n_0	n_1
1	p_0	p_1

The error for splitting is given by the sum of the errors of both nodes

$$\mathrm{err}(D') = \min(n_0, p_0) + \min(n_1, p_1), \tag{2}$$

where $n' = n_0 + n_1$ and $p' = n_1 + p_1$. The error reduction rate (err_red) is given by the reduction of error if comparing the error of "not splitting" with the error of "splitting", divided by the total number of examples in node D':

$$\text{err\_red}(D') : \frac{\min(n', p') - (\min(n_0, p_0) + \min(n_1, p_1))}{|D'|}. \tag{3}$$

Consider the maximal possible error (in this case for binary data) when $n' = p' = 0.5|D'|$

$$\text{err}_{\max}(D') = \min(n', p') = 0.5|D'|, \tag{4}$$

then the maximal possible error reduction is given by

$$\text{err\_red}(D') : \frac{\min(0.5|D', 0.5|D') - (\min(n_0, p_0) + \min(n_1, p_1))}{|D'|} = 0.5, \tag{5}$$

where either $p_0 = 0$ or $n_0 = 0$ and $p_1 = 0$ or $n_1 = 0$ (maximal information gain).

1.3 Mutual Information

The mutual information is a measure of information gain when splitting a dataset. In figure 1 the mutual information and the error reduction rate are plotted for following example:

References

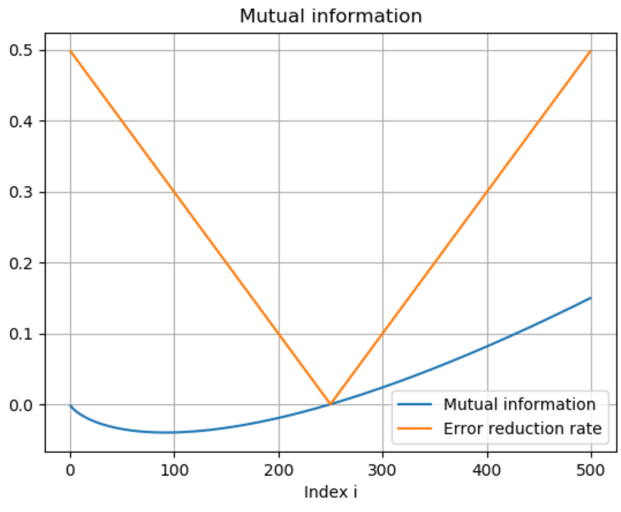

Figure 1: Comparison of mutual information between feature and label and error reduction rate for splitting the dataset.