

Normalised Accuracy

Frans Coenen

30 May 2014

When comparing the operation of a prediction (classification) algorithm, in terms of accuracy, where one of the parameters is the size of the class set C we might get a set of results as shown in Table 1 and Figure 1. Note that the probabilities expressed in the table and figure assume a normal distribution which may not necessarily be the case. From the table we might argue the $|C|=1$ is the best because it produced 100% accuracy. Alternatively we might argue that a classifier designed to allocate all records to one class is pointless and therefore $|C|=2$ is the best because it gives a 95% accuracy. However, intuitively $|C|=3$ might be argued to be the best because it has an accuracy of 94% but using three classes. A prediction accuracy of 80% with respect to five alternatives, where we have a 20% chance of getting it right, seems better than a prediction accuracy of 95% where we have a 50% chance of getting it right.

Num Classes	Acc %	Prob. %
1.00	100.00	100.00
2.00	95.00	50.00
3.00	94.00	33.33
4.00	93.00	25.00
5.00	92.00	20.00
6.00	91.00	16.67
7.00	90.00	14.29
8.00	89.00	12.50
9.00	88.00	11.11
10.00	87.00	10.00

Table 1: Example data

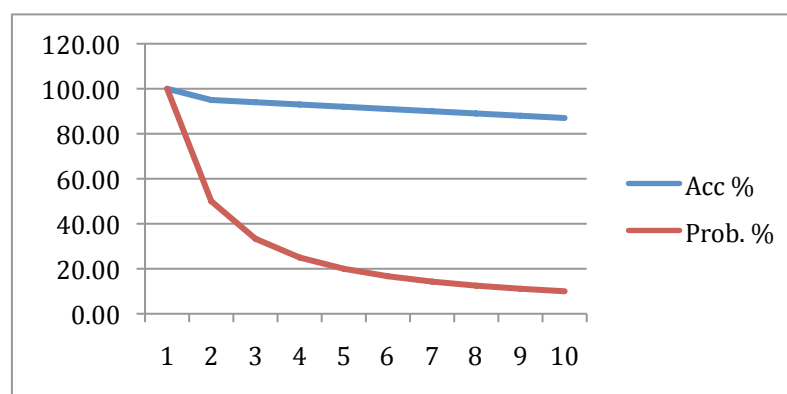


Figure 1: Plot of Data in Table 1

To express this the concept of normalised accuracy is suggested (normalised with respect to the probability of guessing the correct classification). Thus the percentage of records correctly classified beyond that which can be achieved simply by guessing. Thus:

$$\text{Norm. Acc.} = ((\text{Acc.} - \text{Prob.}) / (100 - \text{Prob.})) * 100$$

Applying this to the data in Table 1 we get results as shown in Table 2 and Figure 3. From the table and figure it can be seen that when the probability of correctly guessing a class label is taken into account, $|C|=3$ produces the best result with a normalised accuracy (that takes into account the probability of a correct guess) of 91%

Num Classes	Acc %	Prob. %	Acc.-Prob.	100 - Prob	(Acc.-Prob) / (100-Prob.)	Norm. Acc. %
1	100.00	100.00	0.00	0.00	0.00	0.00
2	95.00	50.00	45.00	50.00	0.90	90.00
3	94.00	33.33	60.67	66.67	0.91	91.00
4	93.00	25.00	68.00	75.00	0.91	90.67
5	92.00	20.00	72.00	80.00	0.90	90.00
6	91.00	16.67	74.33	83.33	0.89	89.20
7	90.00	14.29	75.71	85.71	0.88	88.33
8	89.00	12.50	76.50	87.50	0.87	87.43
9	88.00	11.11	76.89	88.89	0.87	86.50
10	87.00	10.00	77.00	90.00	0.86	85.56

Table 2: *Normalised accuracy calculation*

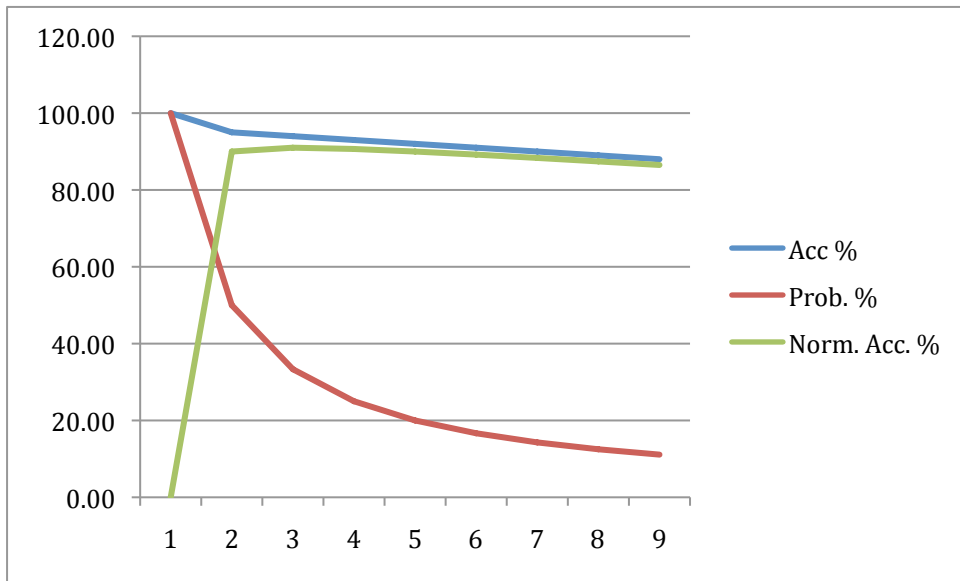


Figure 2: *Plot of Data in Table 2*