Data Mining & Machine Learning
Project 4 - Classification Based on Probability

Steinarr Hrafn Höskuldsson

September 2022

Section 2.2

The accuracy of both method is the same, 98%. The confusion matrices are also identical:

$$\mathrm{CM} = \begin{bmatrix} 19 & 0 & 0 \\ 0 & 19 & 1 \\ 0 & 0 & 20 \end{bmatrix}$$

And it makes sense that they are identical because our dataset is equally split between classes. When estimating the maximum likelihood we assume that the prior probability of each class is equal. In this particular dataset this is actually true. So incorporating the class prior probability of into our maximum likelihood estimation does not change any of the probability estimates.

Independent

A maximum likelihood classification can be seen a special case of a posteriori classification when prior probability of each class is equal. So when the dataset is balanced there is no difference. If the features do a good job of describing and differentiating the data then incorporating the prior probability, assuming an unbalanced dataset, will only do harm to the accuracy. So for a posteriori classification to outperform a maximum likelihood classification the dataset needs to be unbalanced, and the features should not do a good job of describing the targets. As an example of such a dataset I created a dataset with 1000 simulations of flipping 10 coins to predict the throw of a unfair 4 sided die. Using the same methods as were used in sections 1 and 2 the following confusion matrices and accuracies were found:

Using maximum likelihood estimation:

$$\mathrm{CM}_{MLE} = \begin{bmatrix} 4 & 6 & 5 & 10 \\ 4 & 1 & 9 & 11 \\ 20 & 15 & 24 & 44 \\ 13 & 22 & 31 & 34 \end{bmatrix}$$

$Accuracy = 25\%$

Using a posteriori classification:

$$\mathrm{CM}_{MAP} = \begin{bmatrix} 1 & 1 & 15 & 8 \\ 0 & 0 & 16 & 9 \\ 6 & 0 & 52 & 45 \\ 3 & 4 & 51 & 42 \end{bmatrix}$$

$Accuracy = 37.5\%$

As can be seen the a posteriori classification doesn't waste many guesses on the unlikely sides of the unfair die, making it more likely to guess correctly.