COMP 4106 Assignment Two

Connor Stewart (101041125)
Gabrielle Latreille (101073284)
Abdaljabbar Hersi (100919600)

Monday, March 15, 2021

Matthew Holden
CARLETON UNIVERSITY

1 – Abstract

This report provides information on a self-directed project on classification by Bayes theorem and rule-based systems. The project is to design and implement the well-known naïve Bayes classifier and fuzzy classifiers to determine a dog breed based on a set of parameters such as girth, height, and weight. Input into the program is formatted as a function-call called a fuzzy classifier, which takes an array of three numbers as input. The three input parameters represent the girth, height, and weight of the respective dog, and the program outputs an array representing the class membership values along with the name of the most likely dog breed for the given input parameters. The program's output is returned through the terminal for review, and input is sent either through the program's code or via external function calls.

2 – Introduction

The goal of the project is to determine the breed of a dog based on input parameters representing a dog's physical dimension. The breed classification is using a naïve Bayes classifier with fuzzy classification. The system uses a set of fuzzy classifiers to help divide the dogs by breed and uses Goguen t-norms and s-norms for fuzzy rules. The Goguen-norms parameters are generated as the product of a function given as per the assignment specifications. All three parameters representing the girth, height, and weight have three sets of given parameters for the Goguen norms. The parameter sets are defined as small, medium, and large - respectively - and automatically have assigned t- and s-norm coefficients for variables a, b, c, and d. The defined fuzzy rule systems are in the given assignment's specifications, and we follow the requirements for the program input and output, as the specifications do not read or write to files.

3 – Datasets

The application uses input via the python-code to obtain the girth, height, and weight values for the provided dog. The program itself makes use of multiple given assumptions from the assignment. The fuzzy rules for the associated system include:

IF height is medium AND (girth is small OR weight is light) THEN beagle.
IF girth is medium AND height is short AND weight is medium THEN corgi.
IF girth is large AND height is tall AND weight is medium THEN husky.
IF (girth is medium OR height is medium) AND weight is heavy THEN poodle.

Given below are the values for each fuzzy set and characteristic. Note that the program generates the values for a, b, c, and d for the fuzzy-membership function. Also, note that the girth and height are units of centimeters, and the weight is in units of kilograms:

Girth:

Small: a = 0, b = 0, c = 40, d = 50
Medium: a = 40, b = 50, c = 60, d = 70

Large: a = 60, b = 70, c = 100, d = 100

Height:

Short: a = 0, b = 0, c = 25, d = 40

Medium: a = 25, b = 40, c = 50, d = 60

Tall: a = 50, b = 60, c = 100, d = 100

Weight:

Light: a = 0, b = 0, c = 5, d = 15

Medium: a = 5, b = 15, c = 20, d = 40

Heavy: a = 20, b = 40, c = 100, d = 100

As per the assignment specifications, we are provided the fuzzy membership functions and fuzzy rules systems. The fuzzy rules use the Goguen t-norm and the Goguen s-norm. The fuzzy membership function is assumed to be trapezoidal given by the following formula:

$$f(x) = \begin{cases} 0 \; if \; x \leq a \\ \dfrac{x-a}{b-a} \; if \; a < x < b \\ 1 \; if \; b \leq x \leq c \\ \dfrac{d-x}{d-c} \; if \; c < x < d \\ 0 \; if \; d \leq x \end{cases}$$

4 – Results and answers:

The implemented algorithm correctly outputs the breed of dog based on the inputted parameters. The class memberships for the classification made is also given, alongside the breed of dog.

Solutions to the provided questions in the assignment are given below:

1 What type of agent have you implemented?

The type of agent implemented is a simple reflex agent; since the agent only acts based on the current precept, namely the user's input. Using the user input, a specific output - namely, the breed of dog - is declared. Thus, this follows the condition-action rule, where if a condition set (in this case, measurements) is satisfied, a scripted result occurs (the breed of dog is declared).

2 Suggest a particular instance of this problem for the naïve Bayes classifier where using equal prior probabilities for each class changes the most likely class (compared to using the above-specified prior probabilities).

If we input values 68, 68, and 28 for Girth, Height, Weight, the mostly likely class is Poodle if we keep the prior probabilities as is (i.e., $P(beagle)$=0.30 $P(corgi)$=0.21 $P(husky)$=0.14 $P(poodle)$=0.35). But If we use equal prior probabilities for each class (i.e., $P(beagle)$=0.25 $P(corgi)$=0.25 $P(husky)$=0.25 $P(poodle)$=0.25) the most likely class is Husky.

3 Suggest a particular instance of this problem where the most likely class identified by the naïve Bayes classifier and the highest membership class identified by the fuzzy classifier are different.

An instance where the naïve Bayes classifier and the highest membership class identified by the fuzzy classifier are different is the case where the input list is [68, 68, 28]. This represents a dog with a girth of 68cm, a height of 68cm, and a weight of 28kg. We get the following input output stream from python when using the above parameters:

>>> input = [68, 68, 28]

>>> print(naive_bayes_classifier(input))

('poodle', [9.985770647425284e-34, 5.794314898340868e-54, 0.3772876775099364, 0.6227123224900636])

>>> print(fuzzy_classifier(input))

('huski', [0, 0.0, 0.48, 0.08000000000000002])

Thus, the naïve bayes classier returns a poodle as the breed, and the fuzzy classifier returns a husky as the breed.

4 Suppose we try to use our fuzzy classifier to classify a dog of a different breed (i.e., not beagle nor corgi nor husky nor poodle). How might we identify such a case?

If we try to identify a dog of a different breed completely, this would mean that the membership function for each breed beagle, corgi, husky, and poodle would be [0, 0, 0, 0]. Such a case can be identified by taking the characteristics of a Chihuahua and defining its inputs as [10,10,3]. Calculating the membership functions for each characteristic given the features we get:

Girth [1, 0, 0]

height [1, 0, 0]

weight [1, 0, 0]

Applying the fuzzy rules associated with the system we get the following output:

[beagle, corgi, husky, poodle]

[0, 0, 0, 0]

5 Suggest a particular instance of this problem for the fuzzy classifier where the class with the highest membership function would be different if we used the Gödel T-norm and Gödel S-norm (instead of using the Goguen T-norm and Goguen S-norm):

Let us look at example 2 given in the assignment, the girth, height, and weight are given as 65, 55, and 30, respectively. Calculating the membership functions for each characteristic given the features we get:

Girth [0, 0.5, 0.5]

height [0, 0.5, 0.5]

weight [0, 0.5, 0.5]

Where the order is [small, medium, large], after this you simply apply the fuzzy rules associated with the system (Note: this is where using Gödel and Goguen take affect)

If we apply Gödel the membership for each breed becomes the following:

[beagle, corgi, husky, poodle]

[0, 0, 0.5, 0.5]

Looking at the rules since husky if-then rule is satisfied first, Husky becomes the highest membership function.

But Goguen T-norm and S-norm gives us a differing answer which is:

[0.0, 0.0, 0.125, 0.375]

In this case poodle is clearly the highest membership function.

5 – Discussion and Conclusion:

In conclusion, the algorithm can accurately predict dog breeds based on input measurements. The program runs quickly and within a good timeframe and satisfies the assignment requirements.

6 – Statement and Contributions:

Overall, the project was completed in group meetings through discord. All group members made significant contributions to the overall project, with approximately equal contribution among each member. Since the group work was evenly divided amongst members and completed in a shared environment, all members of the group completed all aspects of the assignment together.