COMP 3106 – Assignment 2

Date: October 20th, 2025

Group 133 Members:

Rahul Rodrigues (101145082),
Emily Amos (101311817),
Abaan Noman (101305538)

Statement of Contributions:

All group members made approximately equal contributions.

Team:
- Brainstorming ideas to implement Naive Bayes classifier
- PDF report and questions

Rahul:
- Parse CSV function
- Helped check Fuzzy model norms and proposition logic values

Emily:
- Fuzzy model equations and calculations
- Helper functions for mean, standard deviation

Abaan:
- AI performance measures and validation strategies
- Helper function for getting the log of the Gaussian probability density function

1) What type of agent have you implemented (simple reflex agent, model-based reflex agent, goal-based agent, or utility-based agent)?

Model-based reflex agent.
It maps the current percept (length, weight, speed) to an action (class) in one step. It does keep an internal model of the world learned from data (class priors and class-conditional Gaussians), then applies a reflex rule "choose argmax posterior." It's not goal-based (no explicit goal tests) and not a utility-based planner (no trade-offs across states).

2) Is the task environment:
   a) Fully or partially observable?
      Fully observable. We observe all features needed to classify (length, weight, speed).

   b) Single or multiple agent(s)?
      Single agent. Only our classifier is acting.

   c) Deterministic or stochastic?
      Stochastic. The real mapping features class has noise. We model it probabilistically.

   d) Episodic or sequential?
      Episodic. Each classification is independent. There's no dependence on previous cases.

   e) Static or dynamic?
      Static. The world (aka the snake's measurements) doesn't change while we decide.

   f) Discrete or continuous?
      Continuous. Inputs are continuous, outputs are discrete classes therefore this task environment is continuous.

   g) Known or unknown?
      Unknown (learned). Before training the data-generating distribution is unknown. We estimate it from labelled data. After training, parameters are known but still approximate.

3) Suppose we wish to measure how well our methods work. Suggest what measure(s) of performance and/or what validation scheme should be used. Assume that we have a labelled set of instances available for this task.

   Gather the performance measures:
   - Accuracy: the % of snakes classified correctly ($\frac{\#\ of\ correct\ predictions}{\#\ of\ predictions}$)
   - AUC (Area Under the Curve). For multi-class, the slides suggest reporting mean one-vs-one or mean one-vs-all AUC.

Report for our 3-class task:

- Compute the above metrics per class using a one-vs-rest view, and then average them:
  - Macro-average (equal weight to each class).
  - Micro-average (equal weight to each instance).

Validation scheme: To check how well our Naive Bayes classifier works, we can use cross-validation and some performance measures (accuracy, precision, etc).

- K-fold cross-validation:
  - Split our dataset into parts, e.g. 10 parts.
  - Train the model on 9 parts and test it on the remaining 1 part.
  - Repeat this 10 times, so each part is used once for testing.
  - Finally average the results to get an overall score. This helps make sure our model works well on new, unseen data.
- Leave-one-out cross-validation:
  - If the dataset is very small, leave-one-out cross validation is acceptable.
  - Assume we have N samples, called x_i
  - for $i = 1...N$:

    train model M on all samples x_k for $k \neq i$

    Test M on x_i
  - Performance is mean over all x_i

4) Suggest a particular instance of this problem where using equal prior probabilities in the naïve Bayes classifier for each class changes the most likely class (compared to using the prior probabilities computed from the dataset).

If priors are empirical and imbalanced, the MAP decision can flip.
Example (hypothetical numbers for one snake x):

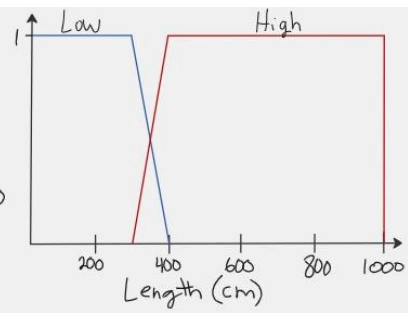
- Likelihoods: $p(x \mid anaconda) = 0.90,\ p(x \mid python) = 1.00$
- Uniform priors: $\pi_a = \pi_b = \frac{1}{3}$ which leads to posteriors $\alpha\ 0.90\ and\ 1.00$. Python wins.
- Empirical priors: suppose dataset is skewed, $\boldsymbol{\pi}_a = 0.99, \pi_b = 0.01$.

  Posteriors $\alpha\ 0.90 \times 0.99 = 0.891\ vs\ 1.00 \times 0.01 = 0.01$. Anaconda wins.
  So the same x flips label when we change from uniform to empirical priors. (Currently 3 per class, priors are equal, so we won't see this flip unless an imbalanced dataset.)

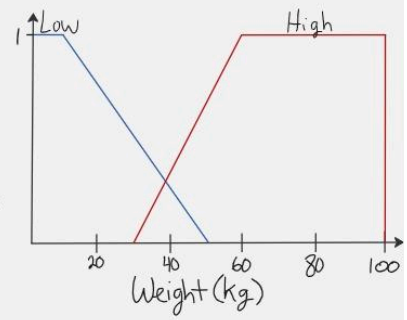
5) Consider, instead, a fuzzy rule-based system for classifying whether a snake is an anaconda, cobra, or python with the following rules.
IF length is high AND weight is high OR speed is high THEN snake is anaconda
IF length is high AND weight is low AND speed high THEN snake is cobra
IF length is low AND weight is high OR speed is low THEN snake is python

All fuzzy membership functions are trapezoidal with parameters given below. Units for length are in centimeters, units for weight are in kilograms, and units for speed are in kilometres per hour.
Length Low (LL): a = 0, b = 0, c = 300, d = 400
Length High (LH): a = 300, b = 400, c = 1000, d = 1000
Weight Low (WL): a = 0, b = 0, c = 10, d = 50
Weight High (WH): a = 30, b = 60, c = 100, d = 100
Speed Low (SL): a = 0, b = 0, c = 10, d = 20
Speed High (SH): a = 10, b = 20, c = 100, d = 100
Suppose a snake has the following measurements. Compute the highest membership class (i.e. anaconda, cobra, or python) and the membership value for each class..

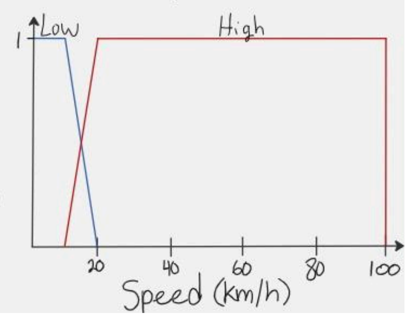
$$m_{LL}(x) = \begin{cases} 0 & \text{if } x \le 0 \\ \cancel{\frac{x-0}{0-0}} & \text{if } 0 \le x \le 0 \\ 1 & \text{if } 0 \le x \le 300 \\ \frac{400-x}{400-300} & \text{if } 300 \le x \le 400 \\ 0 & \text{if } x \ge 400 \end{cases} \qquad m_{LH}(x) = \begin{cases} 0 & \text{if } x \le 300 \\ \frac{x-300}{400-300} & \text{if } 300 \le x \le 400 \\ 1 & \text{if } 400 \le x \le 1000 \\ \cancel{\frac{1000-x}{1000-1000}} & \text{if } 1000 \le x \le 1000 \\ 0 & \text{if } x \ge 1000 \end{cases}$$



$$m_{WL}(x) = \begin{cases} 0 & \text{if } x \le 0 \\ \cancel{\frac{x-0}{0-0}} & \text{if } 0 \le x \le 0 \\ 1 & \text{if } 0 \le x \le 10 \\ \frac{50-x}{50-10} & \text{if } 10 \le x \le 50 \\ 0 & \text{if } x \ge 50 \end{cases} \qquad m_{WH}(x) = \begin{cases} 0 & \text{if } x \le 30 \\ \frac{x-30}{60-30} & \text{if } 30 \le x \le 60 \\ 1 & \text{if } 60 \le x \le 100 \\ \cancel{\frac{100-x}{100-100}} & \text{if } 100 \le x \le 100 \\ 0 & \text{if } x \ge 100 \end{cases}$$



$$m_{SL}(x) = \begin{cases} 0 & \text{if } x \le 0 \\ \cancel{\frac{x-0}{0-0}} & \text{if } 0 \le x \le 0 \\ 1 & \text{if } 0 \le x \le 10 \\ \frac{20-x}{20-10} & \text{if } 10 \le x \le 20 \\ 0 & \text{if } x \ge 20 \end{cases} \qquad m_{SH}(x) = \begin{cases} 0 & \text{if } x \le 10 \\ \frac{x-10}{20-10} & \text{if } 10 \le x \le 20 \\ 1 & \text{if } 20 \le x \le 100 \\ \cancel{\frac{100-x}{100-100}} & \text{if } 100 \le x \le 100 \\ 0 & \text{if } x \ge 100 \end{cases}$$



a) Length: 350 cm, Weight: 40 kg, Speed: 15 km/h

Compute fuzzy truth values:

$m_{LL}(350) = \frac{400-350}{400-300} = 0.5$ $m_{LH}(350) = \frac{350-300}{400-300} = 0.5$

$m_{WL}(40) = \frac{50-40}{50-10} = 0.25$ $m_{WH}(40) = \frac{40-30}{60-30} = 0.\overline{33}$

$m_{SL}(15) = \frac{20-15}{20-10} = 0.5$ $m_{SH}(15) = \frac{15-10}{20-10} = 0.5$

Use Godel norms to compute the rule strengths of each proposition:

i) IF length is high AND weight is high OR speed is high THEN snake is anaconda
$s(t(0.5, 0.\overline{33}), 0.5) = s(0.\overline{33},\ 0.5) = 0.5$

ii) IF length is high AND weight is low AND speed high THEN snake is cobra
$t(0.5, 0.25, 0.5) = 0.25$

iii) IF length is low AND weight is high OR speed is low THEN snake is python
$s(t(0.5, 0.\overline{33}), 0.5) = s(0.\overline{33}, 0.5) = 0.5$

∴ The highest membership classes are anaconda and python at 0.5

b) Length: 310 cm, Weight: 20 kg, Speed: 12 km/h

Compute fuzzy truth values:

$m_{LL}(310) = \frac{400-310}{400-300} = 0.9$ $m_{LH}(310) = \frac{310-300}{400-300} = 0.1$

$m_{WL}(20) = \frac{50-20}{50-10} = 0.75$ $m_{WH}(20) = 0$

$m_{SL}(12) = \frac{20-12}{20-10} = 0.8$ $m_{SH}(12) = \frac{12-10}{20-10} = 0.2$

Use Godel norms to compute the rule strengths of each proposition:

i) IF length is high AND weight is high OR speed is high THEN snake is anaconda
$s(t(0.1, 0), 0.2) = s(0,\ 0.2) = 0.2$

ii) IF length is high AND weight is low AND speed high THEN snake is cobra
$t(0.1, 0.75, 0.2) = 0.1$

iii) IF length is low AND weight is high OR speed is low THEN snake is python
$s(t(0.9, 0), 0.8) = s(0, 0.8) = 0.8$

∴ The highest membership class is python at 0.8

c) Length: 392 cm, Weight: 55 kg, Speed: 19 km/h

Compute fuzzy truth values:

$m_{LL}(392) = \frac{400-392}{400-300} = 0.08$ $m_{LH}(392) = \frac{392-300}{400-300} = 0.92$

$m_{WL}(55) = 0$ $m_{WH}(55) = \frac{55-30}{60-30} = 0.8\overline{33}$

$m_{SL}(19) = \frac{20-19}{20-10} = 0.1$ $m_{SH}(19) = \frac{19-10}{20-10} = 0.9$

Use Godel norms to compute the rule strengths of each proposition:

i) IF length is high AND weight is high OR speed is high THEN snake is anaconda
$s(t(0.92, 0.8\overline{33}), 0.9) = s(0.8\overline{33},\ 0.9) = 0.9$

ii) IF length is high AND weight is low AND speed high THEN snake is cobra
$t(0.92, 0, 0.9) = 0$

iii) IF length is low AND weight is high OR speed is low THEN snake is python
$s(t(0.08, 0.8\overline{33}), 0.1) = s(0.08, 0.1) = 0.1$

∴ The highest membership class is anaconda at 0.9

d) Length: 315 cm, Weight: 32 kg, Speed: 18 km/h

Compute fuzzy truth values:

$m_{LL}(315) = \frac{400-315}{400-300} = 0.85$ $m_{LH}(315) = \frac{315-300}{400-300} = 0.15$

$m_{WL}(32) = \frac{50-32}{50-10} = 0.45$ $m_{WH}(32) = \frac{32-30}{60-30} = 0.0\overline{66}$

$m_{SL}(18) = \frac{20-18}{20-10} = 0.2$ $m_{SH}(18) = \frac{18-10}{20-10} = 0.8$

Use Godel norms to compute the rule strengths of each proposition:

i) IF length is high AND weight is high OR speed is high THEN snake is anaconda
$s(t(0.15, 0.0\overline{66}), 0.8) = s(0.0\overline{66},\ 0.8) = 0.8$

ii) IF length is high AND weight is low AND speed high THEN snake is cobra
$t(0.15, 0.45, 0.8) = 0.15$

iii) IF length is low AND weight is high OR speed is low THEN snake is python
$s(t(0.5, 0.0\overline{66}), 0.2) = s(0.0\overline{66}, 0.2) = 0.2$

∴ The highest membership class is anaconda at 0.8

6) Compare the results from the above question to the results from your naïve Bayes classifier. Briefly comment on why the predicted classes are the same in some cases and different in other cases.

The fuzzy system and the Naïve Bayes classifier often predict the same class but differ in some borderline cases.

Example	Naive Bayes Most Likely Class (Code)	Fuzzy Highest Class	Comment
0	anaconda (0.515 vs 0.485)	Tie (anaconda = python = 0.5)	Both models see 350 cm, 42 kg, 13 km/h as near the decision boundary. Naive Bayes slightly favours anaconda because of its learned Gaussian means and variances, while fuzzy rules treat overlapping sets as equal.
1	cobra	cobra	Agreement. The short, light, and fairly fast features clearly match the “cobra” region in both systems.
2	python	python	Agreement. Low length, low weight, and low speed make python dominant. Both models agree.
3	anaconda	anaconda	Agreement. Very similar to example 0 but slightly different distributions make anaconda dominate strongly under Naive Bayes.

Both methods capture the same general pattern i.e. anacondas are long / heavy / fast, cobras are lighter and faster, pythons are shorter / slower but they differ because:
Naive Bayes is statistical: it computes smooth Gaussian likelihoods and priors from data. Even tiny numeric differences in mean / variance can shift the posterior, so one class may slightly outweigh another (as in example 0).
Fuzzy logic is rule-based and symbolic: it uses crisp linguistic boundaries with min-max (Gödel) operators, creating broad plateaus where two classes can tie if the inputs lie in overlapping trapezoids.
Thus, they agree when the measurements are clearly in one region and diverge when the data fall near overlapping fuzzy boundaries.