Centre for infrastructure, Sustainable Transportation and Urban Planning

Indian Institute of Science (IISc), Bengaluru
Summer Internship Program 2024

Test for "Choice modelling intern" Position

The aim of this exam is to test your problem-solving skills and basic understanding of Python. You are encouraged to submit the best possible answer to the question.

Instructions:

- Plagiarism will result in instant disqualification. You must write your own code.
- To make your submission, use the following Google form: https://forms.gle/DY7itiFGxJYR8JHC9
- You are allowed to make only one submission for this test. While submitting, you will be asked to upload three documents:
  - A Python file (format: .py) containing all the codes related to the question
  - An output file (format: .txt) from the above code
  - A report (format: pdf) summarising your findings from the question. The report should include essential components such as clearly stated assumptions, informative visualisations, and your findings. Do not copy and paste code in the report.
- If possible, the Python code should be written according to PEP 8 – style guide (reference).
- Your submissions will be evaluated based on the quality of the report and codes.
- If selected, it will be mandatory for you to join in an in-person capacity. Please refrain from attempting the test if you cannot attend in-person.
- The test commences on 22nd March 2024 (10:00 AM). The last date for submission is 24th March 2024 (10:00 AM). Late submissions will not be accepted.
- The test will be used to shortlist for "Choice modelling intern", supervised by Dr. Abdul Rawoof Pinjari
- For any clarifications, contact Anil Koushik (anil.koushik@fsid-iisc.in)

All the best.
CiSTUP
IISc, Bengaluru

Question

Write a Python function to calculate the probability of each alternative in a multinomial choice setting using the logistic function, given a set of parameters and independent variables. The function should be generic enough to handle any number of alternatives and independent variables.

In a multinomial logit model, the probability of each alternative is calculated using a logistic function. For each alternative, a deterministic utility (V) is computed based on a linear combination of independent variables and their respective coefficients (β). The probability of each alternative is the exponential of its utility divided by the sum of exponentials for all utilities.

Given Sample Data:

```
data = {
   'X1': [2,1,3,4,2,1,8,7,3,2],
   'X2': [8,7,4,1,4,7,2,2,3,1],
   'Sero': [0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
   'S1': [3,8,4,7,1,6,5,9,2,3],
   'AV1' : [1,1,1,1,1,0,0,1,1,0],
   'AV2': [1,1,1,0,0,1,1,1,0,1],
   'AV3': [1,1,0,0,1,1,1,1,1,1]
}
```

Deterministic Utilities:

$$V_1 = \beta_{01} + \beta_1 X1 + \beta_{S1,13} S1$$

$$V_2 = \beta_{02} + \beta_2 X2 + \beta_{S1,23} S1$$

$$V_3 = \beta_{03} + \beta_1 Sero + \beta_2 Sero$$

Probabilities to Compute:

$$P_1 = \frac{AV1 \times exp(V_1)}{AV1 \times exp(V_1) + AV2 \times exp(V_2) + AV3 \times exp(V_3)}$$

$$P_2 = \frac{AV2 \times exp(V_2)}{AV1 \times exp(V_1) + AV2 \times exp(V_2) + AV3 \times exp(V_3)}$$

$$P_1 = \frac{AV3 \times exp(V_3)}{AV1 \times exp(V_1) + AV2 \times exp(V_2) + AV3 \times exp(V_3)}$$

Parameters:

$$\beta_{01} = 0.1, \beta_1 = -0.5, \beta_2 = -0.4, \beta_{02} = 1, \beta_{03} = 0, \beta_{S1,13} = 0.33, \beta_{S1,23} = 0.58$$

Tasks

Write a Python function called 'calculate_probabilities' that takes the following inputs:

- Parameters: A dictionary containing the β coefficients.

- Data: A dictionary containing the independent variables (X1, X2, Sero, etc.).

- Utilities: A list of functions that define the deterministic utilities for each alternative based on the given parameters and data.

Your function should output a new dictionary with keys representing each alternative and values as lists containing the calculated probabilities for each data point. Save this output in .txt file format.

Ensure your code is well-commented to explain the logic used at each step.

Bonus: Include error handling for possible input errors such as mismatched dimensions between parameters and data points.

Evaluation Criteria:

- Correctness of the logistic function implementation.
- Ability to handle a dynamic number of alternatives and independent variables.
- Code readability and use of comments.
- Proper error handling.