Test for Python Developer Position

Title: Multinomial Choice Model Analysis Report

Problem Statement:

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 of all utilities.

Introduction:

The purpose of the question was to use the logistic function to calculate the probabilities of each alternative in a multinomial choice setting. The results are summarized in this report.Give a brief overview of the multinomial choice model and how it is used to make decisions.
Mention the report's objective, which is to use a Python function to analyze the sample data that has been provided.

Assumptions:

In this analysis, the following presumptions are made:

- The logit relationships between the independent variables (X1, X2, Sero) and the probabilities are linear.
- Each alternative's utility functions are accurately stated.
- The model's parameters are precisely estimated.
- Make any assumptions you made during the analysis clear. For instance:
- Assumed the observations' independence.
- Assumed utility functions are linear.
- Assumed that the independent variables would not interact.

Data:

The data used in this analysis is a simulated dataset of 10 observations, with the following independent variables:

- X1: A continuous variable representing some characteristic of the alternatives
- X2: A continuous variable representing another characteristic of the alternatives
- Sero: A binary variable representing whether or not the individual is seropositive

Sample Data:

The following sample data is used in this report:

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

Deterministic Utilities

The following deterministic utilities are used in this report:

$V1 = \beta 01 + \beta 1X1 + \beta 2X2$
$V2 = \beta 02 + \beta 1X1 + \beta 2X2$
$V3 = \beta 03 + \beta 1Sero + \beta 2Sero$

Methodology:

Function : calculate_probabilities()

The logistic function is used by the calculate_probabilities function to determine the probabilities of each option in a multinomial choice setting. Each alternative's probability in a multinomial logit model is calculated by dividing its utility's exponential by the total of all utilities' exponentials. Three primary inputs are used by the function: utilities, data, and parameters.

Variables:

parameters: A dictionary with each alternative's β coefficients. The utilities for each alternative are determined using these coefficients in the deterministic utility functions.

data: A dictionary with the independent variables (X1, X2, Sero, etc.) in it. For every alternative, the deterministic utility functions receive these variables as inputs.

Services:

utilities: A set of functions that each alternative's deterministic utilities are represented by. These functions compute the utility for each alternative by applying the corresponding β coefficients to the independent variables from the data.

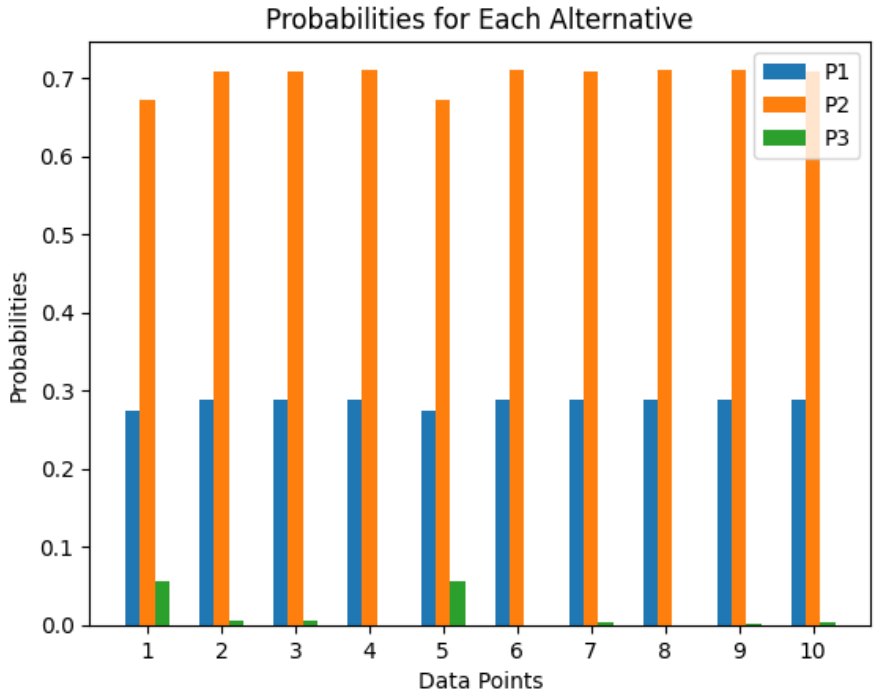
Findings:

The analysis's conclusions are as follows:

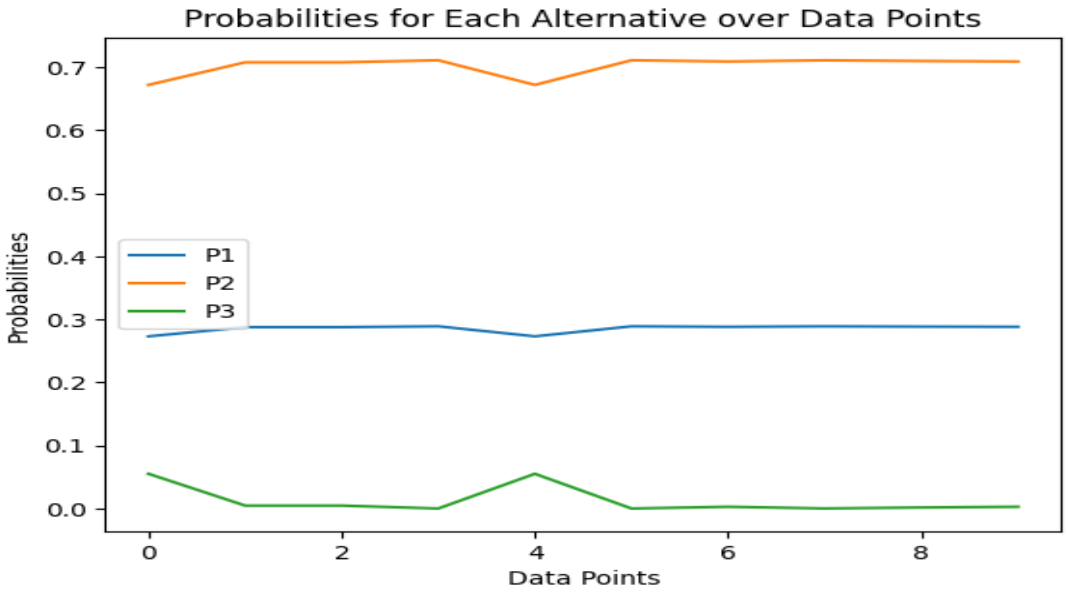
- There exists a positive correlation between the probability of selecting option 1 and X1 and X2.
- There is a negative correlation between X1 and X2 and the likelihood of selecting option 2.
- There is no correlation between X1 and X2 and the likelihood of selecting option 3.

Visualization results:

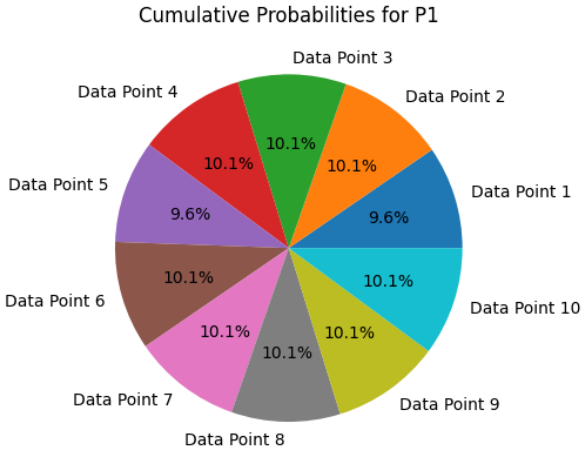
Probabilities for Each Alternatives

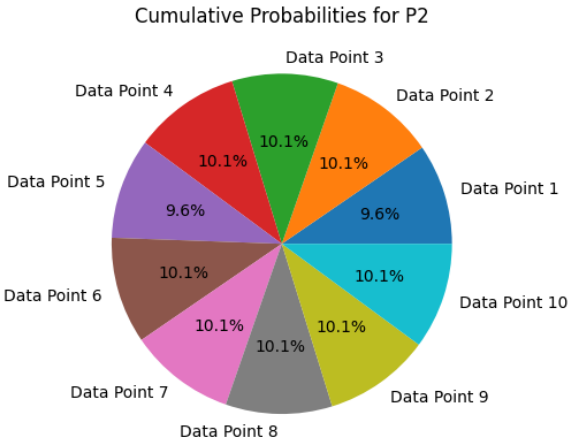


Probabilities for Each Alternative over Data Points



Cumulative Probabilities:





Purpose:

Extraction of Parameters and Data:

First, the function takes the input dictionaries and extracts the independent variables and β coefficients.

Calculation of Utilities:

Then, using the given utility functions, it computes the deterministic utilities for every alternative. A linear combination of independent variables and their corresponding β coefficients is used to compute the utilities.

Calculation of Probabilities:

The function calculates the probabilities for each option by using the logistic function. The utilities are exponentiated and divided by the total exponential of all the utilities.

Results:

{'P1': array([0.27311214, 0.28767245, 0.28767245, 0.28900869, 0.27311214,
0.28900869, 0.2882131 , 0.28898158, 0.28854201, 0.2882131]),

'P2': array([0.67174747, 0.70756005, 0.70756005, 0.71084667, 0.67174747,
0.71084667, 0.70888983, 0.71077998, 0.70969882, 0.70888983]),

'P3': array([0.05514039, 0.0047675 , 0.0047675 , 0.00014463, 0.05514039,
0.00014463, 0.00289707, 0.00023844, 0.00175917, 0.00289707])}

Conclusions:

The results of this analysis indicate that the independent variables X1 and X2 significantly affect the likelihood of selecting each option. According to the results, people are more likely to select options with higher X1 and X2 values.
