Gaussian mixture regression modelling

Due: 8:00, Wednesday 26/9/2018, Turnitin submission on ClickUp

Objectives:

This assignment aims to achieve the following general learning objectives:

- To gain experience using mixtures of Gaussian regression;

- To critically evaluate the properties of mixture regression;

Plagiarism:

The Department of Statistics regards plagiarism as a serious offence. Your submission will be subject to plagiarism checks and appropriate action will be taken against offending parties. You may also refer to the the Library's website at *www.library.up.ac.za/plagiarism/index.htm* for more information.

Report,hand in and mark allocation:

Marks are awarded as indicated in each question.

Submit only the report. No additional files of any sort should be submitted. Do not submit files in any other format other than PDF. Failure to follow any of these instructions will result in a zero mark for the assignment.

Upload the report, PDF file, (named s99999999.pdf, where 99999999 is your student number). **Ensure that you name and student number is clearly indicated on the front page of your assignment.**

Suggested sources

1. Videos as discussed in class

2. Class discussion

3. There are many sources on the internet that can also assist.

Software:

You can use any appropriate software. Use the flexmix package (flexmix) when using *R*.

STK 802 (Clustering & Mixtures) Assignment:3

Data

Consider the file, *tdata.csv* , on ClickUp.The file contains data relating to the claims history of branches within an insurance company. Standardised claims data are reported for a group of 375 branches in a specific year. Data are aggregated to the branch level and the following variables are available:

Variable	Description
$Y = Claims$	Total claims cost
$X_1 = Beneficiaries$	Number of beneficiaries in the branch
$X_2 = Age$	Average age of members
$X_3 = Pens\_Ratio$	Ratio of pensioners to the total number of members
$X_4 = Female\_Ratio$	Ratio of female beneficiaries to the total number of members
$X_5 = Depen\_Ratio$	Average dependency ratio per policy

1: Question 1

Do an exploratory analysis of the observed data. Also make appropriate graphical representations of the observed data.

(**15**)

2: Question 2

Estimate a mixture of Gaussians regression model with three components.

1. Give the estimated mixture rergession model with all estimated parameters.

2. Overlay the graph of the mixture regression model on the observed data.

3. Give a FULL interpretation of the estimated regression model. Specifically address the behavioural aspects in the interpretation. (**30**)

Total [**45**]