Sage Plagarism Report

Powered by schoolworkspro.com

Rohit Shrestha

160490

19

Results Found

20.32 %

Match Percentage

Submitted to: Softwarica	2.88%
Submitted to: Softwarica	2.64%
Submitted to: Softwarica	1.52%
Submitted to: Softwarica	1.4%
Submitted to: Softwarica	1.36%
Submitted to: Softwarica	1.36%
Submitted to: Softwarica	1.32%
Submitted to: Softwarica	1.2%
Submitted to: Softwarica	1.2%
Submitted to: Softwarica	1.16%
Submitted to: Softwarica	1.12%
Submitted to: Softwarica	0.96%
Submitted to: Softwarica	0.64%
Submitted to: Softwarica	0.48%
Submitted to: Softwarica	0.44%
Submitted to: Softwarica	0.44%
Submitted to: Softwarica	0.12%
Submitted to: Softwarica	0.04%

REFLECTIVE REPORT
STW205CDE- Developing the
Modern Web
Name: Rohit Shrestha
College ID: 160490
Github Link: https://github.com/RohitShrestha343/160490_rohit_shrestha_code.git

Contents

Introduction Electroencephalography (EEG) signals the measurement of brain electrical activity which can
be used to study brain function.
Polynomial regression is a type of statistical modelling that can
be used to analyze these signals.
In this approach a polynomial equation is fit to the EEG data
and the coefficients of the equations are used to describe the underlying patterns in the data.
This type of modeling can be used for identifying patterns in EEG signals that are related to specific brain functions such as attention or memory.
It is a non-parametric approach and often
used when underlying relationship between the EEG signals is not known.
Task 1
Preliminary data analysis
Preliminary data analysis is the initial step in analyzing EEG signals using polynomial
It involves cleaning and preprocessing the data, and exploring the basic
characteristics if the signals.
Time Series Plot
Time"series plot is a useful tool for analyzing EEG signals because it allows to visually explore"
patterns in the data over time and can be sued to identify patterns that are related to specific

brain functions.
The x-axis of a time series display indicates time, and the y-axis indicate amplitude of the EEG
R is used to create the time series plot of the input and output of the EEG data which is”
shown in the below mentioned diagram.
The major factor to be considered while interpreting a time series display is that there should not
be any unexpected shift or outliers in the plot caused by abnormal data.
By examining the plot
created through the provided data there seems to be no sudden spikes which signifies that the there are no outside factors or entry error that may have influenced the data.
The second figure displays the output of the four input values against the same date, whereas the four input values are plotted against a timestamp in the first.
Numerous trends in the data
may be seen by carefully examining the graphs.
The time series graph of the input signal that
noise and spikes have been much reduced.
Among the four signals X4 tends to be highly steady.
The plot is showing that after 120ms the
plot seems to be more steady and is in noticeable pattern.
There spikes tend to be decreasing
The output signal is progressive up until 60ms baseline then there we can discover
sudden surge which is very detectable.
This might have been caused due to the noise and
external factor.
After 120ms the spikes tends to be more balancing and follows a steady
Distribution Plot
It is the connection between the observed values and samples of data which is arranged from lower to the greatest values.
The diagram shows the input signal delivered to the brain which is”
shown below.
The above image provides us the bell-shaped curved which means when the histogram and density is observing the distribution its input signal appear around “0”.
When the input peak of
input (X) seems to be ”0” then the highest number of the input signal may be in between number 9”and 10 to the brain whereas the lowest signal is -10.”The distribution shape began to expand from -4 and it decrease form 5, thus show us that the signal mainly lies between -4 and 5. length between the both given tails seems nearly equal which make it easy for locating the distribution’s center.
Let's now explore each of the four input signals that the brain gets at once.
In the figure below,
every input signal from X1 through X4 displays the same pattern.
The X2 input signal doesn't
seem to differ significantly from the outliers.
The output signal, often known as the "Y signal, is sometimes referred to as a left -skewed distribution because, while examining figure we can see that the tail of left side’s is seriously longer than the right one.The majority of the output appears to come from 2 and -2 since the minimum output is 9 and the highest output is 5, therefore the signal Y might be less than 0.5. or
Correlation and Scatter Plots

In our case, the connection of output and input of dataset is shown by using the diagram of
The individual piece in the scatterplot diagram represents a single piece of data
whereas the different dots in the diagram represent different pattern which depends on how the data are closely grouped.
When the correlation of data is perfect positive the relationship of data
is trend to set a straight form of line in the beginning of y axis origin, whereas if the x-axis line's beginnings are not straight, there will be a negative correlation., otherwise it referred as a relationship of no correlation.
There may not be always the negative or a perfect positive
correlation, the relationship may sometimes be slightly positive or slightly negative.
study the data's input and output signals, which are shown below..
The image below shows us
the relationship between the signals of X1, X2 and X3 as well as Y signal, also we can see the pattern of data which appears to follow in a negative correlation the data were not distributed when the dot plot had a value larger than the Y-value on the X-axis.
Since"the data don't follow any pattern so the X2 and Y signals data don't have any"correlation.
Task 2
Task 2.1
An estimator is very much applicable for observing the different characterstics of plot as the exact value of distribution is not known.
Estimator which also can be termed as random variable
can be denoted as Least squares which are generally
used for getting the elegent feasible line of the regression equation also termed as date point that demonstrates the relationship between unknow dependent value with independent value.
Least squares will be used for calculation of each possible module of EEG data to determine the
estimator model parameters.
"C" is used for indicating the least squares.
As input and output
data of EEG signals can be represented as X and Y, we can calculate the least square using
This formula in R can be written as
Prior to computing the least ^ square, the value or column of the X (input data) must first be bound based on the provided dataset.
The ideal method for binding the dataset is by using
Once the data is binded the above mentioned formula can be used toe perform the Least squares form.
The outcomes of all the candidate model is mentioned below:
Task 2.2
The estimator for Sum Squared is referred to as model residual error (RSS) which is used to find out the value of squared average that's means the difference between the value of actual average square and estimated value.
Usually, it is used to control the estimator quality because
the RSS closer value is never negative but it is better to zero.First, we must get each model's error from the of task 2.1 in order to calculate the RSS.
When implementing the R programming which is shown above, it can be calculate as (RSS=sum((Y-Yhat) 2) and Yhat is known as the product of XiC and n is known as the total length of Y and ∑ is the sum of all the output Y.
The ^ table below shows the RSS values of the
given model whereas Xi is also all the outcomes of X.

C ^ is the value of task that we have
calculated form 2.1 task where all models have different value of 0.
At last, for calculating the ^
formula RSS is RSS= sum((Y-Y_hat) ^2).
The output signal is Y and sum is every possible
Task 2.3
Log-Likelihood Function
This can be defined as a function which is generally used for identifying the measuring value that fits in sample data which in provided in the model as an unknown parameter.
identify several approaches for fitting a data distribution to a particular observed value.
logarithmic modification of the likelihood function which also maximizes likelihood and is sometimes called maximum log-likelihood.
By maximizing estimation, it may be used for
practical work objectives.
In the Equation mentioned above
The results of likelihood obtained are listed below
Task 2.4 Akaike Information criterion (AIC)
The Akaike Information criterion (AIC) is an estimate that aids in the prediction of errors for identifying the fact that without overfitting the provided dataset how correctly the data is fitted.
The AIC assists in evaluating the model's quality by comparing many candidate models.
k indicates the length of the parameters to be evaluated, as calculated by task 2.1.
L denotes for the likelihood function based on 2.3.
Bayesian Information criterion (BIC)
BIC can be defined as a standard that helps for choosing the best modules from many candidate models and scoring them.
During the selection of BIC, the candidate with the lowest score is most preferable BIC and AIC are interconnected, and it also depends on some extent on the likelihood function.
The following is a list of the BIC calculation formula.
k indicates the length of the parameters to be evaluated, as calculated by task 2.1.
L denotes for the likelihood function based on 2.3
Task 2.5
For each candidate model in which the output EEG signal has a normal/Gaussian distribution, Task 2.2 calculates the RSS, also known as the modal residual error with additive Gaussian noise is to be plotted.
Expected Y(Output) or Y hat value was deducted from output value (Y)
and is used in ggnorm and ggline plotting functions of ggplot2 which are pre-defiend in R programming, for calculation RSS.
For determining the most efficient regression model and dat
trend a normal/gaussian distribution was plotted for each model.
Task 2.6
For choosing the top model of all the given task i.e from 2.1 to 2.5 were finished.
For eg, plotting
normal graphs distribution, RSS, likelihood.
All the steps were completed carefully.
From 2.5
plot and calculation of BIC and AIC I will choose best candidate model for this regression.
are all aware that BIC and AIC were mostly used throughout the selecting process.
models help to minimize the score of models while choosing the model.
AIC choose the model

which is based on relative distance from likelihood of data and unknown likelihood of the model.
So, that the model can be chosen if there is lower AIC because it is close to the truth value.
Also, BIC compares the model of function which follow the certain of Bayesian system being a true model.
That why the model can be chosen when the BIC is lower cause it seems to be close to the truth.
So, for this assignment, I'll choose a model using both the BIC and the AIC.
By examining the output produced by each BIC and AIC of task 2.4 as well as the normal distribution plot of task 2.5, I have found that model 2 is the best model from the other listed
The value generated from the model 2 seems to be lower and consistent for both BIC and AIC also the model 2 plot seems closed for normal distribution.
Task 2.7
The input of (X) and output of (Y) of given data has been split into two parts one is a testing dataset, and the other is the training dataset.. Each of the dataset was divided by the information provided which is for training 70% and rest is for the testing purpose With the assistance of choosing the best model, which is model 2, the estimate of the model parameters was done during the training of the data.
The value of q is obtained there.
RSS the model output/prediction were performed while testing of the data after the estimation of
model parameter.
Following the calculation of the RSS model, 95% accuracy was obtained for model prediction.
Task 3
Approximate Bayesian computation is one of the methods which can help to calculate and also evaluate a posterior distribution of model parameter which is perform without calculating parameter of likelihood.
In this task we were using the ABC method for the posterior distribution
and furthermore to perform the rejection of ABC was done using the model that was selected from task 2.
When completing task 2, I selected model 2 as a potential candidate for a regression module later on which that I will use to complete this task.
I have selected the two values from the model
2 for estimation of the model parameter that was carried out in 2.1 task for computing posterior
The chosen parameter should be the largest possible value.
After choosing different values, by the assistance of runif function which is predefined in R programming generates two new values the range of the chosen parameters/value is to be
By the combination of new value with the constant value stored in the estimated parameter calculation of Y hat and error of output signal can be done.
As the Y hat and error is
generated, we can calculate the RSS as we did in task 2.2.
Once the RSS has been calculated, Rejection ABC can be performed.
rejection ABC, if the RSS value is higher than the threshold value, those values are to be
Otherwise, the values are approved and stored for plotting if the threshold values are higher than the RSS.
The plots of both the parameters are shown below.
From model 2 two parameters with the highest value were chosen for the calculation of the range for performing the rejection sampling and identification of the range of estimated

For finding the RSS as we did in task 2.2 the calculation of the range, Y hat and error were calculated.
As the RSS was calculated ABC was performed in which the values which has higher RSS than that of estimated threshold were rejected and values that has lower RSS that that of threshold we acceped and plotted.
Overall, in this task, 2 parameters with highest value from mode 3 were selected for calculating the range in order to perform rejection sampling and identifying the range of estimated
To find RSS after computing the range, Y hat and error were obtained just like in task 2.2.
After calculating the RSS rejection ABC was performed where values having higher RSS with a lower threshold value were accepted and plotted, whereas RSS with the estimated threshold were rejected.
The plots for them are shown above.
In this assignment, we had to select the top regression model out of all the candidate models.
By performing different calculations and considering the results obtained I have chosen model 2
among the 5 models because of the low AIC and BIC than that of other models.
R programming language was used for this process and the codes are provided in the below mentioned appendix

https://study.com/learn/lesson/time-series-graph-plot-examples.html
https://onlinelibrary.wiley.com/doi/10.1002/9781119386957.ch4
https://www.shmoop.com/study-guides/math/basic-statistics-probability/scatter-plots-correlation
https://byjus.com/maths/scatter-plot/
https://www.researchgate.net/publication/302509467_Preliminary_Data_Analysis