Assignmentn4 Memo

Dr. Niladri Chakraborty

2025-04-25

Q1.

Dataset: Obtain the Asn4_data.csv file from Blackboard. This dataset consists of 50 rows and 500 columns, where each row represents a data string of 500 variables (as indicated by the columns). Additionally, a dummy variable classifies each data string into one of five types. The id variable serves as the index for each row.

Objective: Create a single plot featuring multiple line charts, one for each of the 50 rows. Use the types of the data strings, as specified by the dummy variable, to color the line charts.

Ans.

```r
rm(list = ls())
# Load necessary libraries
library(tidyr)
library(ggplot2)

data_df = read.csv("C:\\Users\\ChakrabortyN\\OneDrive - University of the
Free State\\Desktop\\Asn4_data.csv", header = T)

# Melt the dataset to a long format for plotting
data_long = gather(data_df, key = "time", value = "value", -dummy_variable, -
id)
data_long$time <- as.numeric(as.character(gsub("V", "", data_long$time)))

# Plot using ggplot with facet_wrap
# Plot using ggplot
p = ggplot(data_long, aes(x = time, y = value, group = id, color =
dummy_variable)) +
  geom_line(alpha = 0.7) +  # Set alpha for better visualization
  labs(title = "Multiple Time Series Plot", x = "Time", y = "Value") +
  scale_color_manual(values = c("Type1" = "darkblue", "Type2" = "#f67e1f",
"Type3"="#cb181d","Type4"="#238b45", "Type5"="#6a51a3")) +  # Color by dummy
variable
  theme_classic() +
  theme(legend.title = element_blank())  # Modify legend

print(p)
```

Multiple Time Series Plot