BRACE - Simulation Study

This R Markdown can be used to generate the simulation results of the paper "Flexible aggregation of compositional predictors with shared effects"

Load required libraries

```r
library(ggplot2)
library(gridExtra)
```

Load processed simulation outputs for BRACE for the simulation scenarios Dep1 and Dep2 where $n = 300$, $\text{SNR} = 1$, $p = 100, 300, 1000$.

```r
df<-readRDS("./Reproduce_Results/Simulated_Data/Processed_Simulation_Outputs_Metrics_Plot.rds")
```

Create the metrics plot. This plot reproduces the BRACE simulation results shown in Table 1 and 2 in the main paper

```r
# Function to create plot for each metric
create_metric_plot <- function(mean_col, sd_col, title) {
  # Ensure that the column names are properly accessed
  if (!(mean_col %in% colnames(df)) || !(sd_col %in% colnames(df))) {
    stop(paste("Column names not found:", mean_col, sd_col))
  }

  # Prepare data for ggplot
  depth_labels <- c("100", "300", "1000")
  dep_labels <- c("Dep1", "Dep1", "Dep1", "Dep2", "Dep2", "Dep2")

  plot_data <- data.frame(
    Depth = factor(rep(depth_labels, 2), levels = depth_labels),
    Dep = factor(dep_labels, levels = c("Dep1", "Dep2")),
    Mean = df[[mean_col]],
    SD = df[[sd_col]]
  )

  # Create the plot
  ggplot(plot_data, aes(x = Dep, y = Mean, fill = Depth)) +
    geom_bar(stat = "identity", position = position_dodge(width = 0.7), color = "black") +
    geom_errorbar(aes(ymin = Mean - SD, ymax = Mean + SD),
                  position = position_dodge(width = 0.7), width = 0.2) +
    scale_fill_manual(values = c("skyblue", "lightgreen", "salmon")) +
    labs(title = title, x = "Simulation", y = "Mean Value") +
    theme_minimal() +
    theme(legend.title = element_blank())
}
```

```
# Create each plot
plot_pe <- create_metric_plot("PE.mean", "PE.sd", "PE Mean and SD")
plot_accuracy <- create_metric_plot("accuracy.mean", "accuracy.sd", "L2 loss Mean and SD")
plot_fp <- create_metric_plot("fp.mean", "fp.sd", "False Positive Mean and SD")
plot_fn <- create_metric_plot("fn.mean", "fn.sd", "False Negative Mean and SD")
```

Generate the prediction error and L2-loss plots

```
# Arrange all plots in a 2x2 grid
#grid.arrange(plot_pe, plot_accuracy,plot_fp,plot_fn,ncol = 2)
grid.arrange(plot_pe, plot_accuracy,ncol = 2)
```
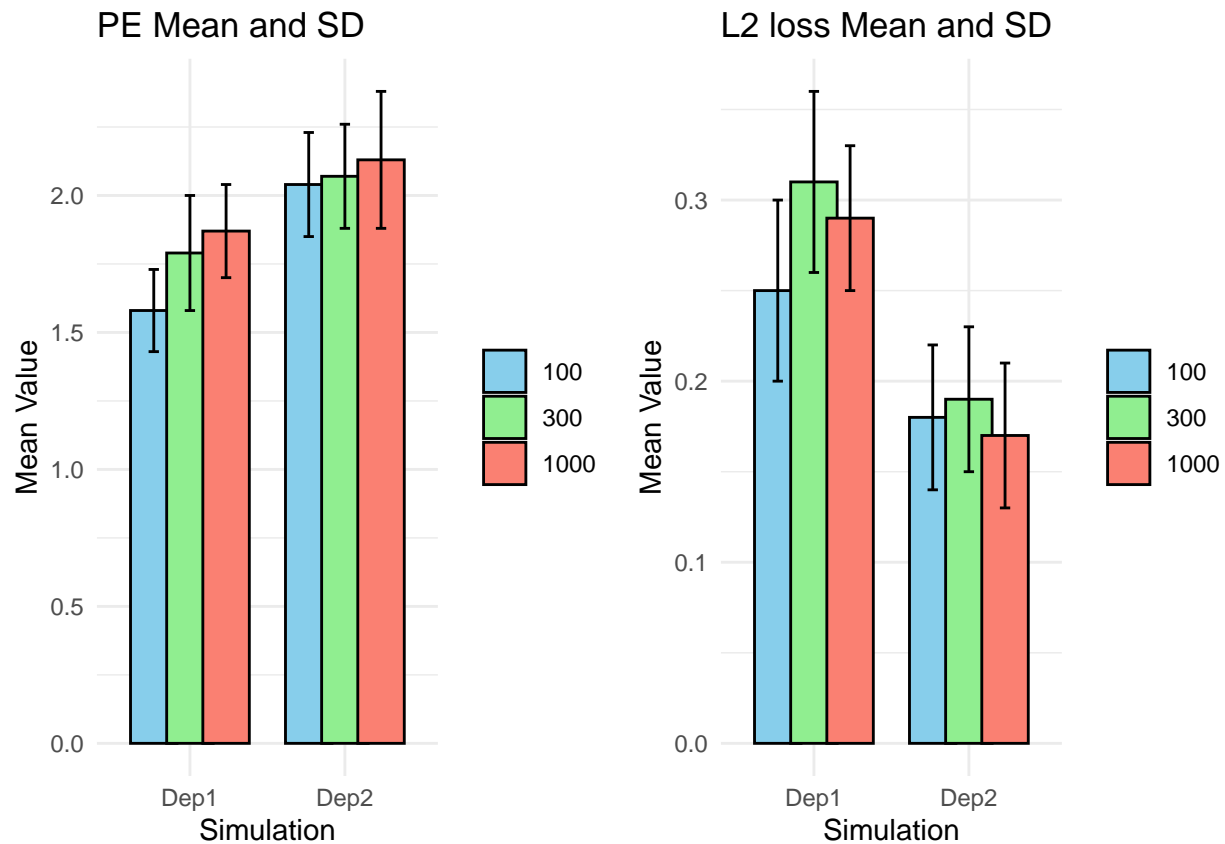

We have omitted the plots for the false positive and false negative metrics, as their mean and standard deviation values are zero.

Main Paper Figure 2: Boxplots showing the adjusted Rand indices calculated by comparing the true and predicted cluster labels in the two dependent setup with the number of predictors set to $p = 100, 300$, and 1000.

```
df<-readRDS("./Reproduce_Results/Simulated_Data/Processed_Outputs_ARI_Plot.rds") # Create the plot with
custom labels and legend at the bottom
pplot<-ggplot(df, aes(x = Deplabels, y = values)) +
  geom_boxplot(aes(fill = plabels)) +
  labs(x = "",           # Rename x-axis
       y = "Adjusted Rand Index",            # Rename y-axis
       fill = "Number of predictors") +      # Change legend title
```

```
  theme_classic() +
  theme(legend.position = "bottom") +# Move legend to the bottom
  ylim(0.8, 1)
pplot
```
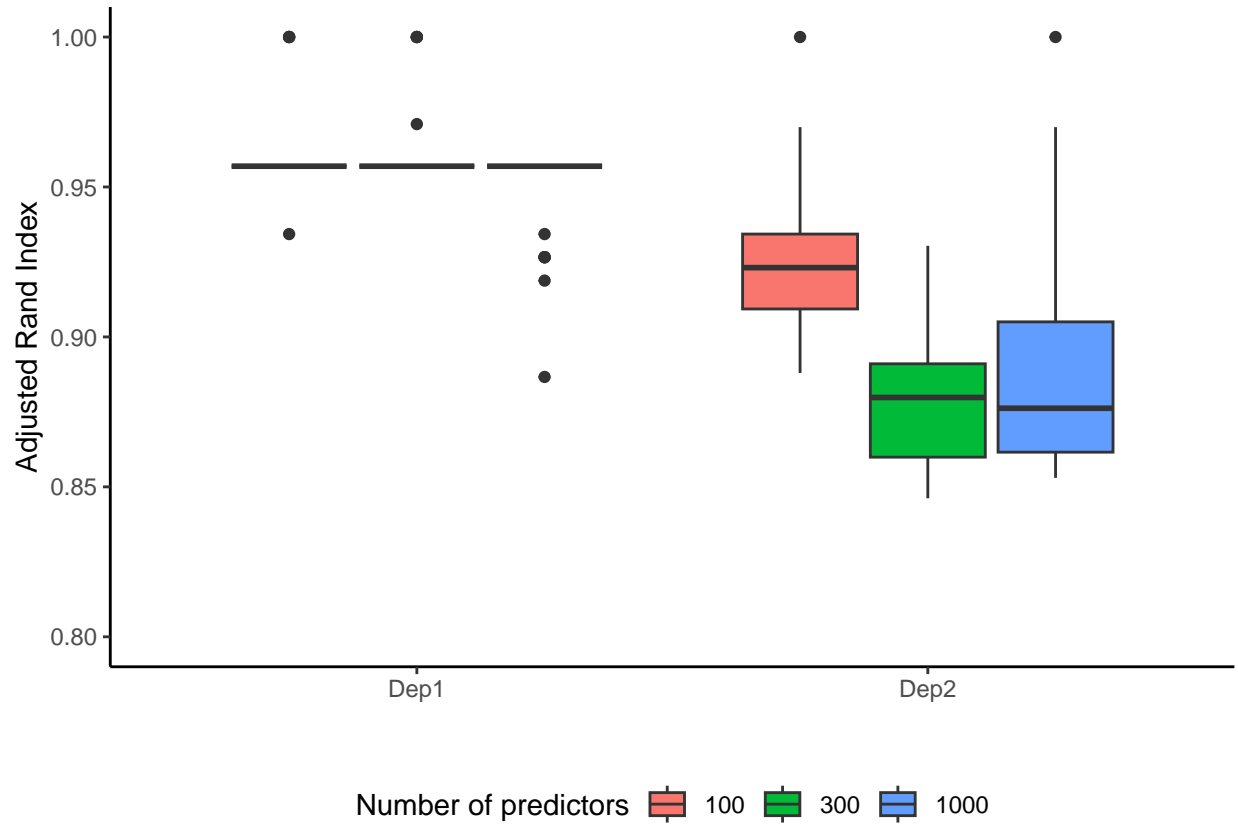

BRACE achieves high values of the ARI for all cases, indicating its ability to recover the true cluster labels with high accuracy.